Investigating executive working memory and phonological short-term memory in relation to fluency and self-repair behavior in L2 speech

Effrosyni Georgiadou (Zayed University) & Karen Roehr-Brackin (University of Essex)

Abstract

This paper reports the findings of a study investigating the relationship of executive working memory (WM) and phonological short-term memory (PSTM) to fluency and self-repair behavior during an unrehearsed oral task performed by second language (L2) speakers of English at two levels of proficiency, elementary and lower intermediate. Correlational analyses revealed a negative relationship between executive WM and number of pauses in the lower intermediate L2 speakers. However, no reliable association was found in our sample between executive WM or PSTM and self-repair behavior in terms of either frequency or type of self-repair. Taken together, our findings suggest that while executive WM may enhance performance at the conceptualization and formulation stages of the speech production process, self-repair behavior in L2 speakers may depend on factors other than working memory.

Word count: 129

Keywords: working memory capacity, executive working memory, phonological short-term memory, L2 speech production, fluency, hesitation phenomena, self-repair behavior

INTRODUCTION

An individual difference factor that has attracted considerable interest in the field of second language (L2) learning in recent years has been working memory (WM) (for a review, see Juffs & Harrington, 2011). For many L2 learners, and especially for those at early stages of acquisition, the skill of speaking can be challenging because of the limited automaticity of the underlying linguistic processes and the resulting dependence on attentional resources for simultaneous content generation and message formulation (Levelt, 1989, 1993; Kormos, 2006). Accordingly, the role of executive WM and phonological short-term memory (PSTM) in L2 speech production has become a key area of investigation. Researchers have used a variety of WM measures and oral L2 tasks to explore the relationship between WM and L2 speakers' performance in different aspects of oral production, such as fluency, accuracy and complexity (Gilabert & Muñoz, 2010; Mota, 2003) with inconclusive and sometimes even contradictory results. What is more, oral fluency measures have mostly focused on speed of production, while other dimensions of fluency such as hesitation phenomena or self-repair behavior have not been included. Finally, the role of L2 proficiency, though acknowledged to be of potential importance in oral production (Kormos, 2006), has typically not been included as a variable either (for exceptions, see Kormos & Sáfár, 2008; Mizera, 2006). The present study sought to address these gaps in existing research by examining the relationship between WM and oral production in L2 speakers at different levels of proficiency, using a research design that incorporated measures of executive WM both in L1 and L2 as well as a measure of PSTM. Moreover, in addition to speech rate, oral fluency was investigated in terms of hesitation phenomena and self-repair behavior.

THEORETICAL AND EMPIRICAL BACKGROUND

Speech production and self-monitoring

Levelt's (1999) modular theory of speech production remains the most comprehensive and widely used theory of speech processing not only in L1, but also in L2 research (de Jong, Groenhout, Schoonen & Hulstijn, 2013; Kormos, 2006; Segalowitz, 2010). The model posits five speech processing components (labeled in capitals in Figure 1) which have specialized functions and operate autonomously. According to Levelt (1989), most of the components are highly automatic in L1, which allows for the modules to work in parallel, and it also explains the generation of fluent and effortless speech. While most processes involving L1 speech are automatic, Levelt postulates that attention is necessary in constructing (conceptualizing) and monitoring one's message, both conceptually and linguistically. According to Levelt's perpetual-loop theory of monitoring (Levelt, 1983, 1989, 1999), self-monitoring can be carried out prior to or after articulation (see Figure 1).

[INSERT FIGURE 1 HERE]

Self-monitoring cannot be performed efficiently unless the speaker is able to attend to the message. Although in L1 self-monitoring and error detection entail error correction, it has been suggested that L2 speakers make conscious decisions about whether to correct errors and/or which errors to correct (Mackey, 1992). In other words, self-monitoring and error-detection do not entail error correction in L2. Apart from not having the competence to correct a certain error, there are other potential reasons for this, among them a desire not to slow down one's speech (Kormos, 2006), wishing to sound native-like (Lennon, 1990), feeling embarrassed to self-correct because it may direct the listener's attention to the mistake (Krashen, 1981), being bothered by frequent mistakes (Seliger, 1980), and perceptions as to what extent one's errors may result in communication breakdown (Kormos, 2006).

The concept of attention is closely connected to the notion of automaticity. At earlier stages of acquisition when automatization is limited, speech production in L2 requires more attention than in L1 and is thus more effortful (Segalowitz & Hulstijn, 2005). For instance, it has been shown that processes such as grammatical and phonological encoding during formulation (Levelt, 1989) are not automatic in novice L2 learners and only partially automatized in more advanced L2 speakers (De Bot, 1992). Therefore, besides performing the same attention-dependent processes as an L1 speaker, i.e. conceptualizing and self-monitoring, an L2 speaker must additionally rely on his/her attentional resources in order to successfully formulate his/her message and intelligibly articulate it. Attentional resources are dependent on an individual's WM. Indeed, some researchers would regard attentional control and WM as essentially synonymous (e.g. Engle & Kane, 2004).

Working memory

Although various models of WM have been put forward in an attempt to explicate the exact nature and functions of the construct (e.g. Cowan, 1988; Engle, Kane & Tuholski, 1999; Kieras, Meyer, Mueller & Seymour, 1999), it is Baddeley and Hitch's (1974) multi-component model of WM (see Figure 2) that has arguably been most influential in L2 learning research (Juffs & Harrington, 2011).

[INSERT FIGURE 2 HERE]

According to Baddeley (1996, 2012), the three key functions of the central executive are to focus, divide and switch attention. It also constitutes the link between working and long-term memory (Baddeley, 1996). The phonological loop, one of the slave systems, temporarily

holds verbal information, and it is vital to vocabulary acquisition in both L1 and L2 (Gathercole & Adams, 1993; Gathercole, Hitch, Service & Martin, 1997; Gathercole, Willis, Emslie & Baddeley, 1992; Papagno & Vallar, 1992). The phonological loop consists of a phonological store and an auditory rehearsal mechanism that helps maintain decaying information through internal repetition of stored items (Baddeley, Gathercole & Papagno, 1998).

Phonological loop capacity is typically measured by means of simple word or digit recall tests as well as non-word recall tests. The former type of test assesses both storage capacity and the efficiency of the rehearsal mechanism, whereas a non-word recall test assesses the duration of information retention in PSTM without rehearsal and the accuracy with which phonologically unfamiliar items can be reproduced (Williams, 2012). Executive WM, i.e. the combined capacity for both storage and processing including the central executive and its direction of attentional resources, is typically measured by tests that require simultaneous storage and processing of information or that require participants to perform two simultaneous but unrelated tasks tapping processing and storage respectively, such as verifying the result of strings of arithmetic operations (processing component) and recalling words or letters that are presented following each operation (storage component), as in the operation span test (Turner & Engle, 1989). The appropriateness of different measures of executive WM in the context of L2 research continues to be under debate, with researchers' positions depending on whether they regard WM as a relatively domain-general construct (Engle et al., 1999; Turner & Engle, 1989) or as essentially skill-specific (or task-specific) and thus featuring an exclusively linguistic component (Daneman & Carpenter, 1980; Daneman & Green, 1986; Just & Carpenter, 1992).

Empirical research on working memory and L2 speech production

As the attentional resources that are vital for L2 speech production depend on WM capacity, the relationship between executive WM in particular and different dimensions of L2 speech has been relatively widely investigated in recent years (e.g. Fortkamp, 1998, 1999, 2000; Gilabert & Muñoz, 2010; Kormos & Sáfár, 2008; Mizera, 2006; Mota, 2003; Trebits & Kormos, 2008). Cumulative findings suggest a positive relationship between executive WM capacity and speed of speech, operationalized as words or syllables spoken per minute. More specifically, Fortkamp (2000) found a significant correlation between performance on an L2 speaking span test and words per minute in the speech of advanced L2 speakers of English, while Mota (2003) reported significant correlations between learners' scores on an L2 speaking span test and several temporal measures of fluency, including mean length of run and pruned and unpruned speech rate. However, the use of a speaking span test in L2 is potentially problematic because the measure may not only assess executive WM capacity, but also L2 proficiency, given that the test requires participants to produce grammatically accurate and meaningful sentences.

In order to avoid this problem, some researchers have opted for language-independent executive WM measures and/or language-dependent measures administered in L1. Mizera (2006) investigated the relationship between a number of L2 fluency measures and executive WM, operationalized via an L1 math span test and an L1 speaking span test. A weak significant correlation was found between speech rate and speaking span. Along similar lines, Trebits and Kormos (2008) used a backward digit span test administered in L1. They identified a significant correlation with L2 speech rate. Finally, Gilabert and Muñoz (2010), using a reading span test in L1, also found a significant but weak correlation with L2 speech rate. The results from these studies are all indicative of a relationship between L2 speech rate

and executive WM capacity operationalized via language-independent or L1-based measures. It is worth bearing in mind, however, that speech rate is not the only measure of fluency, although it has certainly been widely used, with only two of the aforementioned studies (Mizera, 2006; Mota, 2003) including hesitation phenomena as additional measures. No relationship between hesitation phenomena and executive WM was found in either study.

Whereas most research has focused on executive WM, only two studies (Kormos & Sáfár, 2008; Mizera, 2006) have additionally examined the relationship between PSTM and fluency and have also taken into account learners' levels of L2 proficiency. In both studies, PSTM was measured by means of a non-word recall test. Kormos and Sáfár (2008) found a significant correlation in their pre-intermediate learner group between PSTM and scores for fluency and vocabulary range assessed according to the Cambridge First Certificate of English rubric (which relies on subjective rater judgements), but no such relationship was identified in the beginner group. At the same time, executive WM as measured by an L1 backward digit span test was found to correlate with beginners' overall scores on the Cambridge First Certificate speaking exam, but not pre-intermediate learners' scores. By contrast, Mizera (2006) reported a significant correlation between PSTM and fluency for the pre-intermediate participants in his study, that is, the higher-proficiency group. With regard to executive WM as measured by a speaking span test and fluency, a slightly stronger correlation was found for the lower-proficiency learners compared with the higher-proficiency learners. Taken together, these findings suggest that PSTM and executive WM may make distinct contributions to the speaking performance of learners at different levels of proficiency.

Thus far, only a single published study has investigated the relationship between WM and self-repair behavior in L2 learners. Mojavezi and Ahmadian (2014) examined speakers' self-repair behavior in L1 and L2 in relation to executive WM operationalized via an L1 listening span test. The study revealed significant correlations between executive WM capacity and two types of self-repair in L1 (appropriacy and error repairs; see Table 1 for self-repair types) and two types of self-repair in L2 (different-information and error repairs). L1 and L2 speakers with greater WM capacity performed more appropriacy but fewer different-information repairs, respectively. Interestingly, the relationship for error repairs was negative in L1, with greater WM capacity associated with fewer error repairs, and positive in L2, with greater WM capacity associated with more error repairs. The authors suggest that the negative relationship between WM and L1 error repairs indicates that L1 speakers can attend and monitor their speech more effectively due to the automaticity of their language production processes, which results in more accurate speech and thus the need for fewer grammatical and lexical repairs. In contrast, they argue that the positive relationship between WM and L2 error repairs indicates that L2 speakers with extra attentional resources will utilize them to attend to form (in addition to meaning) and, consequently, perform more error repairs. The authors postulate that this is in keeping with previous findings (Ahmadian, Abdolrezapour & Ketabi, 2012; Guará-Tavares, 2009) according to which L2 speakers with greater WM capacity produce more accurate language because they can monitor more effectively. Although the proposed explanations are plausible and in accordance with Levelt's (1983, 1999) theory, they are based on the assumption that monitoring equals self-correction. This may be the case for L1, but does not have to be the case for L2 (Mackay, 1982). In addition, the relationships observed were based on the number of error repairs performed, not on the rate of corrected errors out of all errors committed – an approach which

would have arguably provided stronger support for the authors' claims regarding the interconnection between monitoring, self-repair and linguistic accuracy.

Research issues and questions

In summary, a significant positive relationship between executive WM and L2 fluency in terms of speech rate has been corroborated by a number of existing studies. This suggests that speakers with greater WM capacity, as assessed in the majority of cases exclusively by L2-based measures, produce more words per minute than speakers with less WM capacity. However, hesitation phenomena have been explored in only a limited number of studies, with findings showing no relationship with executive WM. It is not immediately obvious why this aspect of fluency should be less sensitive to WM capacity than speech rate, at least in L2 learners at early stages, so further examination of the relationship seems desirable. Similarly, investigations into the relationship between PSTM and aspects of fluency are scarce, and to date, they have produced contradictory findings. As greater phonological loop capacity should facilitate storage of more verbal material, including material that is being processed prior to articulation, one would expect a positive relationship between PSTM and fluency. Moreover, with only one study examining self-repair behavior in connection with executive WM in L1, nothing is known about the relationship between self-repair behavior and executive WM in L2 or PSTM. Last but not least, very few studies to date have included speakers' level of L2 proficiency as a variable, although different degrees of automaticity in L2 speech production, which can be assumed to be a function of level of L2 proficiency, should play a role in the association between WM capacity and fluency.

The present study was aimed at addressing these issues by (a) including not only speech rate as a measure of fluency, but also hesitation phenomena as a measure of disfluency, (b)

including a measure of self-repair behavior, (c) incorporating both language-independent and L2 skill-specific measures of executive WM and examining separately the contributions of executive WM and PSTM, and (d) investigating the role of executive WM and PSTM in speakers at different levels of L2 proficiency. The following research questions were posed:

1. What is the relationship of executive WM and PSTM with speech rate and hesitation phenomena in the oral speech produced by elementary and lower intermediate L2 speakers?

2. What is the relationship of executive WM and PSTM with elementary and lower intermediate L2 speakers' self-repair behavior?

METHODOLOGY

This section provides information about the participants of the present study, the instruments used and the procedure followed for data collection and analysis.

Participants

Seventy-seven learners of L2 English (mean age = 18, range = 17-20 years) participated in the study, all of them female and L1 speakers of Arabic. The participants were enrolled in an intensive English language course at pre-baccalaureate level at the female-only campus of the state university in Abu Dhabi where one of the authors worked at the time. There were no instances of sensory or mental impairment officially recorded in the participants' student files. The learners had been exposed to English instruction at school for 9-12 years. None of the participants had lived or studied in an English-speaking country for a period of more than six months (mean = 1 month). Based on their scores on the Oxford Quick Placement Test (UCLES, 2004; see below for details), 42 participants were at an elementary level (mean = 20.1/40, SD = 2.088, range = 16-23) and 35 at a lower intermediate level (mean = 25.7/40,

SD = 1.487, range = 24-30) of L2 proficiency. In general, language students in the UAE, especially in the two largest emirates, Dubai and Abu Dhabi, have considerable exposure to English outside the classroom. Due to the diversity of the population residing in the country, English is widely used as a lingua franca by international speakers in everyday communication outside the students' homes.

Instruments

Background questionnaire

The participants completed an 11-item background questionnaire which gathered basic biographical information about their language learning history and exposure to English, as reported above.

L2 English proficiency

In order to arrive at an estimate of participants' level of L2 English proficiency, the paper-and-pen version (Part 1) of the Oxford Quick Placement Test (UCLES, 2004) was administered. The 30-minute test assesses vocabulary and grammar knowledge as well as reading ability.

Fluency (speech rate, hesitation phenomena) and self-repair in L2 speech

Samples of participants' L2 speech were obtained by means of an oral interview. The structure of the interview was based on a component of the standard IELTS (International English Language Testing System) speaking test. Participants were given a task card with a topic and questions to which they had to respond for two consecutive minutes. In the present study, participants were required to talk about their best friend. They were allowed one

minute of preparation time, during which they could take notes if they wished. None of the participants in the present study took advantage of this option.

Executive working memory

Participants' executive WM was measured by means of a backward digit span test (BDST) (Wechsler, 1997) administered in L1 Arabic and a tailor-made listening span test (LST) administered in L2 English. Prior to use in the present study, each test was piloted twice with smaller samples representative of the targeted population in two exploratory studies.

The BDST requires participants to listen to sequences of single-digit numbers and then recite the sequences in reverse order. The test consists of seven sets of four sequences each that gradually increase in length. Thus, the sequences in the first set comprise 3 digits each, the sequences in the second set comprise 4 digits each, and so forth, up to the final set, where sequences comprise 9 digits each. The administration and scoring of the BDST was based on Blackburn and Benton's (1957) BBII procedure according to which there were two trials for each sequence within a set, and the test was terminated after two consecutive failures at repeating a sequence. In addition, credit was awarded for every sequence of a given set repeated correctly. In other words, participants were given .25 for each correctly recalled sequence within a set after the first sequence which attracted 3 points as correctly recalling 3 digits in reverse was considered the baseline for the test, indicating a WM span of 3. Based on this scoring procedure, the maximum possible score was 9.75. The test's internal reliability was high (Cronbach's alpha = .91).

The LST was developed for the present study following the guidelines and format of listening span tests used in previous studies (Daneman & Carpenter, 1980; Mackey, Philp, Fujii &

Tatsumi, 2002). It was tailored to the world knowledge and language proficiency of the participating learners. The LST required participants to listen to sets of English sentences and perform two tasks, that is, (1) to orally judge each sentence as either true or false, and (2) to memorize and recall the final word of each sentence. The test comprised a total of 36 English sentences in nine sets, that is, three sets each of 3, 4 and 5 sentences. All sentences were declarative statements in the active voice and were 6-9 words in length. Sentences of different lengths were distributed as evenly as possible throughout the sets. Each sentence ended in a singular or plural noun, which ranged from 1-3 syllables in length. None of the final nouns in a single set were semantically related or could potentially form a compound noun with another final noun. Truth judgements were made after each sentence, while final-word recall was performed at the end of a set of sentences. Participants were permitted to recall sentence-final words in a given set in any order, and both the singular and the plural form of the sentence-final nouns were accepted. There were 18 true and 18 false sentences. Truth judgements could be made by drawing on basic general world knowledge and/or the meaning of the words in the sentences. Example sentences from the test are: *Humans have two arms and four legs. A day has twenty-four hours. People celebrate their birthday twice a year.* The purpose of the truth judgement task was to ensure semantic processing of the sentences and thus prevent strategic rehearsal of the sentence-final words for subsequent recall.

Participants were given composite scores for the LST (Mackey et al., 2002). These were calculated on the basis of the number of correctly recalled sentence-final words in addition to the number of correctly judged sentences. Thus, a participant could earn one point for recalling the final word of each of the 36 sentences and one point for accurately judging the

sentence as true or false. Therefore, the maximum possible composite score for the LST was 72. Inter-item reliability was acceptable (Spearman-Brown split-half = .62).

Phonological short-term memory

Participants' PSTM was measured by means of a serial word recall test (WRT) in L2 English that was developed for the present study and piloted prior to use. The test comprised five sets of words, with each set consisting of three lists. The words were presented orally and recall took place immediately after each list. The number of words per list increased from set to set, so that the first set consisted of three-word lists and the final set of seven-word lists. The words had been selected from a list of the most frequent words in American English (Kucera & Francis, 1967) and were 1-2 syllables in length. Words within the same list were not semantically related, and no word could be part of a potential compound noun formed with the preceding or subsequent words in a set. Scores for the WRT were calculated on the basis of the number of words recalled throughout the test. Participants were awarded one point per word recalled in the order of presentation. Thus, the maximum possible score was 75. Inter-item reliability was good (Spearman-Brown split-half = .80).

Data collection and analysis

The measures were administered in a fixed order. First, participants completed the background questionnaire and the L2 proficiency test in groups, supervised by the first author. Then, the LST, the BDST and the WRT were administered individually and on separate days by the first author and scored in situ. All tests were accompanied by instructions in L1 and L2 as well as practice items to ensure that participants were clear about what they were required to do. In a final session, participants were interviewed individually in L2 English by the first author. All oral interviews were audio-recorded and transcribed in

CHAT format for subsequent analysis via the CLAN program of the CHILDES database (MacWhinney, 2000).

The interviews were analyzed for fluency, disfluency and self-repair behavior. Fluency was operationalized by means of speech rate and disfluency by means of hesitation phenomena. Speech rate was measured in words per minute (Ellis & Barkhuizen, 2005; Fortkamp, 2000). An unpruned speech rate was used, so filled pauses, repeated words and phrases and self-corrections were included in the count (Gilabert & Muñoz, 2010). Hesitation phenomena (Skehan & Foster, 1999) included number of pauses and number of repetitions, calculated for the two-minute sample of speech available for each participant. The measure of number of pauses comprised filled and unfilled pauses. A filled pause was operationalized as a) non-lexical fillers when an elongated single vowel (i.e. uh, um) or portion of a syllable was produced and b) lexical fillers which are words that do not add to the content of the message (i.e. I mean, you know, etc.) (Fletcher, 2010). Although there is no consensus as per the cut-off point in determining unfilled pauses in L2 oral production, in the present study this was operationalized as a silence of more than two seconds (Lennon, 1990; Fortkamp, 1998).

The measure of number of repetitions included words, phrases or clauses that were spoken twice (or more often) by a participant without modification and immediately following each other.

Self-repair behavior was analyzed in terms of the number and types of overt self-corrections participants made during their two-minute turn. Overt self-corrections were coded in accordance with Kormos' (1998) taxonomy for L2 self-repairs shown in Table 1. Inter-coder agreement based on a sample of 10% of the transcripts was 90%, with cases of disagreement between the two coders resolved through discussion.

[INSERT TABLE 1 HERE]

In order to address the research questions, descriptive statistics and bivariate correlations were calculated. As BDST scores, speech rate, number of repetitions and number of self-repairs were not normally distributed, non-parametric statistics were used throughout.

RESULTS

The first research question asked about the relationship of executive WM and PSTM with speech rate and hesitation phenomena in L2 oral productions by speakers of elementary and lower intermediate L2 proficiency. To begin with, descriptive statistics were calculated for all the variables. The descriptive statistics for speech rate and hesitation phenomena are shown in Table 2.

[INSERT TABLE 2 HERE]

The descriptive statistics in Table 2 indicate that the elementary and lower intermediate participants performed similarly in terms of hesitation phenomena, but the lower intermediate group produced a greater mean number of words per minute than the elementary group. A Mann-Whitney U test revealed that the difference between the two groups was significant (p = .044).

Correlational analyses conducted on the fluency measures for the whole sample showed that number of pauses correlated weakly with number of repetitions (rho = .25, p = .028). There was also a weak negative correlation between number of pauses and words per minute (rho = -.24, p = .036). No significant correlations between the fluency measures were found for the

lower intermediate group. On the other hand, there was a significant but weak correlation between number of pauses and words per minute for the elementary group (rho = -.31, p = .049), suggesting that the relationship observed for the sample as a whole was attributable to the elementary participants. Thus, in learners of lower proficiency, speed of speech and the hesitation phenomenon of pausing are associated, with more words per minute translating into fewer pauses and vice versa. This relationship disappears in learners who are slightly more advanced in terms of L2 proficiency.

The descriptive statistics for the working memory measures are displayed in Table 3.

[INSERT TABLE 3 HERE]

The descriptive statistics in Table 3 suggest that the lower intermediate group seemingly performed better than the elementary group on all measures. A Mann-Whitney U analysis revealed that the differences between the two proficiency groups were significant for the LST (p = .013) and WRT (p < .001), but not for the BDST (p = .23), showing that the more proficient speakers performed better on the L2-based working memory measures. Correlational analyses confirmed this in part. In the sample as a whole, proficiency test scores were significantly and positively correlated with all working memory measures, including the BDST administered in L1 (rho = .27, p = .014). The coefficients for the LST (rho = .36, p = .001) and the WRT (rho = .51, p < .001) were slightly stronger. While the latter results may suggest a potential confound in that LST and WRT scores might have depended to some extent on test takers' overall level of English, the association between BDST performance and L2 proficiency allows for another possible interpretation, namely that their greater WM capacity had enabled some learners to achieve a higher level of L2

proficiency. Unfortunately, it is not possible to resolve the question of cause and effect on the basis of the present data set. [1]

In order to address the first research question, correlations between the working memory measures, speech rate and hesitation phenomena were calculated. The results for the sample as a whole are shown in Table 4.

[INSERT TABLE 4 HERE]

As multiple correlations were run, a Bonferroni correction was applied. With an adjusted alpha level, no statistically reliable correlations were in evidence. As the results reported above suggest that elementary and lower intermediate learners may behave differently, the analysis was repeated for each proficiency group separately, as shown in Table 5.

[INSERT TABLE 5 HERE]

Table 5 demonstrates that there are no significant relationships between the WM measures and fluency for the elementary participants. In the lower intermediate group, a significant negative correlation (Bonferroni corrected) was found between the BDST and number of pauses (rho = -.45, p = .006). Thus, lower intermediate participants with greater executive WM capacity as measured by the BDST paused less.

The second research question asked about the relationship between executive WM, PSTM and self-repair behavior in speakers of elementary and lower intermediate L2 proficiency. The descriptive statistics for self-repair behavior overall are shown in Table 6.

[INSERT TABLE 6 HERE]

The results in Table 6 suggest that participants in the two proficiency groups behaved similarly in terms of overall self-repair. Table 7 displays the number of self-repairs broken down by type.

[INSERT TABLE 7 HERE]

Table 7 indicates that the majority of self-repairs across the sample as a whole were different-information repairs and error repairs, i.e. repairs of accidental lapses. The two proficiency groups considered separately seemed to behave slightly differently with regard to specific self-repair types. The elementary participants made more error repairs, while the lower intermediate participants made more appropriacy and rephrasing repairs. Mann-Whitney U tests revealed that the difference between the two groups was only significant for the rephrasing repairs (p = .030), though. This might suggest that the slightly more advanced participants were able to use a wider range of lexical items and grammatical structures in order to reformulate their message in a more precise manner, although it should be borne in mind that overall relatively few rephrasing repairs were carried out.

In order to address RQ2, correlations were run for the sample as a whole and the two proficiency groups separately. No significant relationships were found between any of the WM measures and the overall number of self-repairs made by the participants for either the sample as a whole or the individual proficiency groups. The results of a correlational analysis by type of self-repair for the sample as a whole are shown in Table 8. Again, no statistically significant correlations are in evidence if the alpha level is adjusted for multiple comparisons.

Similarly, correlational analyses by proficiency group revealed no significant relationships either.

[INSERT TABLE 8 HERE]

DISCUSSION

Working memory and fluency

The first research issue addressed in the present study concerned the relationship of executive WM and PSTM to speech rate and hesitation phenomena in L2 speech production. We found that executive WM as measured by the BDST correlated negatively with number of pauses in the lower intermediate group, indicating that those slightly more advanced speakers in our sample who had greater executive WM capacity paused less. This finding is in keeping with Daneman's (1991) claim of processing efficiency or processing speed (Richardson, 1996) in the conceptualization of a message and the execution of a preverbal plan. Daneman (1991) considers hesitations as evidence of problems occurring at either of these stages of production, that is, conceptualization or formulation, in Levelt's (1989, 1999) terms. Indeed, generating relevant and original content may well constitute a problem for L2 speakers in tasks such as the one used in the present study. The same applies to formulating a message once appropriate content has been generated. The efficiency with which a learner can select appropriate words and access the necessary syntactic and morphophonological information in order to produce accurate and meaningful speech strings at the formulation stage should be evident in (the absence of) hesitation phenomena. Frequent pauses provide speakers with additional time to perform what other speakers can perform faster. In this sense, greater capacity in executive WM, i.e. simultaneous storage and processing capacity as well as better attentional control, should constitute an advantage, resulting in less need to pause in order to create extra time for conceptualization and formulation. Our analyses yielded a significant correlation between executive WM as measured by the BDST and number of pauses for the

lower intermediate group, but not for the elementary group. Whilst an explanation for this pattern of results is not immediately obvious, it is possible to tentatively speculate that greater executive WM capacity may be particularly advantageous for slightly more advanced L2 speakers who have acquired and thus need to sort through a greater amount of linguistic material at the formulation stage than beginning learners.

The present study did not reveal any significant correlations between executive WM as measured by either the BDST or the LST and speech rate. This is in contrast to a number of previous L2 studies which reported significant albeit mostly weak to moderate correlations between executive WM and speech rate (Fortkamp, 2000; Gilabert & Muñoz, 2010; Mota, 2003; Trebits & Kormos, 2008). However, our results are in keeping with the findings reported by Fortkamp (1998) and Kormos and Sáfár (2008). Fortkamp (1998) found no association between speech rate and executive WM as measured by an operation span test and an L2 speaking span test. Kormos and Sáfár (2008) likewise found no relationship between participants' BDST scores and fluency, though it is worth bearing in mind that in contrast to other studies, fluency was evaluated in terms of the criteria of the Cambridge First Certificate oral exam rubric rather than via objective measures.

Taken together, our results suggest that greater executive WM capacity may contribute to enhanced processing speed and efficiency at the stages of conceptualization and formulation. This manifests itself in fewer observable pauses, since L2 speakers with greater executive WM have less need to 'play for time' while conceptualizing and formulating their message.

Working memory and self-repair behavior

The second research issue addressed in the present study referred to the relationship between executive WM, PSTM and self-repair behavior as measured by the frequency and types of self-repair in participants' speech. No significant correlations were identified for either the sample as a whole or for the two proficiency groups considered separately. This is in contrast with the findings reported in the only published study investigating working memory and self-repair behavior (Mojavezi & Ahmadian, 2014). The absence of a relationship between the overall number of self-repairs participants made and either executive WM or PSTM could mean that self-correction as an indication of successful external monitoring and self-repair depends on factors other than working memory, especially in L2 speakers at relatively low levels of proficiency for whom error detection may not necessarily entail error correction. It has indeed been suggested that L2 speakers may make conscious decisions about correcting or not correcting a detected error (Mackey, 1992). As was previously mentioned, feeling embarrassed to self-correct because it may direct the listener's attention to the mistake (Krashen, 1981), wishing to sound native-like (Lennon, 1990), being bothered by frequent mistakes (Seliger, 1980), not wanting to slow down one's speech (Lennon, 1990) and determining whether a specific error may lead to a communication breakdown (Kormos, 2006) are among the reasons that have been put forward. Apart from the perceptions and preferences of individual speakers, however, task characteristics may also be relevant. For instance, a task that requires an L2 learner to focus on grammatical accuracy may yield more grammatical error repairs than a task which is more meaning-oriented. The task used in the present study did not direct participants towards either end of the form-meaning continuum.

The results obtained from the analysis by type of self-repair seem to offer some support for the argument that self-repair is carried out following a potentially conscious decision, and specifically that a speaker may self-correct when s/he feels that successful communication

may be impeded. In the present study, 61% of the total number of self-repairs were different-information and appropriacy repairs, a pattern resembling the self-repair behavior of L1 speakers (Kormos, 1999, 2000; Van Hest, 1996). Such self-repairs respectively focus on encoding different content and encoding the same content in a less ambiguous, more detailed or more specific way. This suggests that the participants in the present study, though still far from a near-native level of L2 proficiency, were primarily concerned with communicating their ideas successfully, rather than with using accurate language in doing so.

CONCLUSION

The present study investigated the role of executive WM and PSTM in speakers at two different levels of L2 proficiency in relation to (a) fluency in oral speech in terms of both speech rate and hesitation phenomena, and (b) self-repair behavior. Our findings suggest that greater executive WM, especially in slightly more advanced learners, may enhance conceptualizing and formulating processes, thus reducing speakers' need to pause. In the absence of any significant association between either executive WM or PSTM and self-repair, we suggested that self-repair behavior may be subject to conscious control and therefore (additionally) depend on factors other than WM capacity. Although the present study has shown that it is worth taking into account participants' level of L2 proficiency, it has to be acknowledged that our two proficiency groups did not differ widely in terms of overall L2 ability. Groups that are more clearly distinct, such as beginners and upper intermediate or even advanced speakers, could usefully be compared to allow for firmer conclusions with regard to the role of working memory in the context of relatively more or less automatized L2 speech production processes. Additionally, it would be of interest to examine working memory, fluency and self-repair behavior in the context of different task types with different requirements for focus on form vs. focus on meaning. Such an approach

would yield insights into the extent to which speech rate, hesitation phenomena and in particular self-correction are influenced by task characteristics and, potentially, learners' conscious preferences and decisions.

NOTES

[1] The three working memory measures themselves were significantly and positively correlated both in the sample as a whole and in the two proficiency groups separately, with coefficients ranging from .36 to .56, indicating shared variance, but not complete overlap. This suggests that the measures tapped a common underlying ability in learners of both L2 proficiency levels.

Conflict of interest: The authors declare that they have no conflict of interest.

REFERENCES

Ahmadian, M. J., Abdolrezapour, P., & Ketabi, S. (2012). Task difficulty and self-repair behavior in second language oral production. *International Journal of Applied Linguistics*. 22(3), 310-330.

Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology* 49A-1, 5-28.

Baddeley, A. D. (2012). Working memory: Theories, models and controversies. *Annual Review of Psychology,* 63, 1-30.

Baddeley, A. D. & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (pp. 47-89). New York: Academic Press.

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

Blackburn, H. L., & Benton, A. L. (1957). Revised administration and scoring of the digit span test. *Journal of Consulting Psychology*, 21(2), 139-143.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin,* 104, 163-191.

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

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

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language,* 25, 1-18.

De Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics*, 13, 1-24.

De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (published online March 2013).Second language fluency: speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics.*

Ellis, R., & Barkhuizen, G. (2005). *Analyzing learner language*. Oxford: Oxford University Press.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp.145-199). New York: Elsevier.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (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 (*pp.102-134). Cambridge, U.K.; Cambridge University Press.

Fletcher, J. (2010). The prosody of speech: timing and rhythm. In W.J. Hardcastle, J. Laver & F.E. Gibbon (Eds.), The Handbook of Phonetic Sciences. UK: John Wiley and Sons Ltd.

Fortkamp, M. B. (1998). Measures of working memory capacity and L2 oral fluency. *Ihla do Desterro*, 35, 201-238.

Fortkamp, M. B. (1999). Working memory capacity and aspects of L2 speech production. *Communication & Cognition,* 32, 259-296.

Fortkamp, M. B. (2000). *Working memory capacity and L2 speech production: An exploratory study.* Unpublished doctoral thesis, Universidade Federal de Santa Catarina.

Gathercole, S. E. & Adams, A. M. (1993). Phonological working memory in very young children. *Developmental Psychology*, 29, 770-778.

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Short-term memory and long-term learning in children. *Developmental Psychology,* 33, 966-979.

Gathercole, S. E., Willis, C.S., Emslie, H. & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology,* 28, 887-898.

Gilabert, R., & Muñoz, C. (2010). Differences in attainment and performance in a foreign language: the role of working memory capacity. *International Journal of English Studies*, 10(1), 19-42.

Guará-Tavares, M. G. (2009). The relationship among pre-task planning, working memory capacity, and L2 performance: A pilot study. *Linguagem & Ensino,* 12(1), 165-194.

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

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

Kieras, D. E., Meyer D. E., Mueller, S., & Seymour, T. (1999). Insights into working memory from the perspective of the EPIC architecture for modeling skilled perceptual-motor and cognitive human performance. In A. Miyake & P. Shah (Eds.), *Models of working memory mechanisms of active maintenance and executive control (*pp.183-223). Cambridge, U.K.; Cambridge University Press.

Kormos, J. (1998). A new psycholinguistic taxonomy of self-repair in L2: A qualitative analysis with retrospection. *Even Yearbook, ELTE SEAS Working Papers in Linguistics*, 3, 43-68.

Kormos, J. (1999). The effect of speaker variables on the self-correction behavior of L2 learners. *System*, 27, 207-221.

Kormos, J. (2000). The role of attention in monitoring second language speech production. *Language Learning*, 50, 343-384.

Kormos, J. (2006). *Speech production and second language acquisition*. New York: Lawrence Erlbaum Associates.

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

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence: Brown University Press.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning,* 40(3), 387-417.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 33, 41-103.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1993). The architecture of normal spoken language use. In G. Blanken, J. Dittman, H. Grimm, J. C. Marshall, & C.- W. Wallesch (Eds.), *Linguistic disorders and pathologies: An international handbook* (pp. 1-15). Berlin: Walter de Gruyter.

Levelt, W. J. M. (1999). Language production: A blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), *Neurocognition of Language* (pp. 83-122). Oxford, U.K.: Oxford University Press.

Mackay, D. G. (1982). The problems of flexibility, fluency and speed-accuracy trade-off in skilled behavior. *Psychological Review,* 89, 483-506.

Mackay, D.C. (1992). Awareness and error detection: New theories and research paradigms. *Consciousness and Cognition, 1,* 199-225.

Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 181-209). Amsterdam: John Benjamins.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk. 3rd Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Mizera, G.J. (2006). *Working memory and L2 oral fluency*. Unpublished Ph.D. thesis. University of Pittsburgh.

Mohavezi, A., & Ahmadian, M. J. (2014). Working memory capacity and self-repair behavior in first and second language oral production. *Journal of Psycholinguistic Research*, 43, 289-298.

Mota, M. B. (2003). Working memory capacity and fluency, accuracy, complexity, and lexical density in L2 speech production. *Fragmentos*, 24, 69-104.

Papagno, C., & Vallar, G. (1992). Phonological short-term memory and the learning of novel words: The effects of phonological similarity and item length. *Quarterly Journal of Experimental Psychology,* 44A, 47-67.

Richardson, J. T. E. (1996). Evolving issues in working memory. In J. T. E. Richardson, R. W. Engle, L. Hasher, R. H. Logie, E. R. Stoltzfus, R. T. Zacks, (Eds.), *Working memory and human cognition* (pp.120-154). New York: Oxford University Press.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in second language learning. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). Oxford, U.K.: Oxford University Press.

Seliger, H. W. (1980). Utterance planning and correction behavior: Its function in the grammar construction process for second language learners. In H. W. Dechert, & M. Raupach (Eds.), *Towards a cross-linguistic assessment of speech production* (pp. 87-99). Frankfurt: Peter Lang.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49, 93-120.

Trebits, A., & Kormos, J. (2008). Working memory capacity and narrative task performance. *Proceedings form the 33rd International LAUD Symposium*, Landau/Pfalz, Germany.

Turner, M. L, & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language,* 28*,* 127-154.

Van Hest, E. (1996). *Self-repair in L1 and L2 production*. Tilburg, Netherlands: Tilburg University Press.

UCLES (2004). *Quick Placement Test* (University of Cambridge ESOL Examinations). Oxford, U.K.: Oxford University Press.

Wechsler, D. (1997). *Wechsler memory scale – revised (manual).* San Antonio: The Psychological Corporation-Harcourt Brace Jovanovich.

Williams, J.N. (2012). Working memory and SLA. In S. M. Gass, & A. Mackey (Eds.) *The Routledge handbook of second language acquisition* (pp. 427-441). New York: Routledge.

Table 1 Taxonomy of L2 self-repairs (Kormos, 1998)

Type of self-repair	Operationalization	Examples from the present study
Different-information repair (D-repair)	Message replacement; different information is encoded.	"um my friend [//] I know my friend since grade one…" "um she like [//] we have a strong relationship.."
Appropriacy repair (A-repair)	Intended message is encoded in a modified way to provide more detailed, more specific or less ambiguous information.	"I give her a hug [//] big hug..." "and not I am in the university [//] in Z university…"
Error repair (E-repair)	Corrections of accidental grammatical, lexical or phonological lapses.	"and sometimes I went [//] I go with…" (grammatical) "my friend he [//] she is…" (lexical) "if I am ubsent [//] upset…" (phonological)
Rephrasing repair (R-repair)	Revision of form but not content of message due to uncertainty about correctness.	"we go together shopping [//] for shopping…" "I'll talk about my friend she's name [//] her name is Amani…"

Kormos's taxonomy of self-repairs with definitions and examples from the study's data set.

Table 2 Descriptive statistics: Speech rate and hesitation phenomena

		N	Min	Max	Mean	SD	Median
Speech rate (words/minute)	Whole sample	77	47	169	92	24	90
	Elementary	42	47	152	86	22	88
	Lower Intermediate	35	60	169	98	25	97
No. of pauses	Whole sample	77	2	18	10	3.5	10
	Elementary	42	2	18	9.6	3.6	10
	Lower Intermediate	35	5	17	10.3	9	10
No. of repetitions	Whole sample	77	0	13	3	2.6	3
	Elementary	42	0	9	3	2.1	3
	Lower Intermediate	35	0	13	3.2	3.1	2

Descriptive statistics for speech rate, number of pauses and repetitions for the whole sample as well as the elementary and lower-intermediate groups.

Table 3 Descriptive statistics: Working memory

		N	Min	Max	Max possible	Mean %	Mean	SD	Median
LST	Whole sample	77	44	66	72	77%	55.4	4.6	56
	Elementary	42	48	60	72	75%	54	3	54.5
	Lower Intermediate	35	44	66	72	79%	57	5	57
BDST	Whole sample	77	3.0	7.75	9.75	46%	4.5	.90	4.2
	Elementary	42	3.0	6.5	9.75	44%	4.3	.80	4.0
	Lower Intermediate	35	3.0	7.75	9.75	48%	4.7	1	4.75
WRT	Whole sample	77	37	73	75	73%	55	7	55
	Elementary	42	37	62	75	69%	52	3	52
	Lower Intermediate	35	45	73	75	77%	58	3	57

Descriptive statistics for the two executive WM tests (LST and BDST) as well as for the phonological STM test (WRT) for the whole sample as well as the elementary and lower-intermediate groups.

Table 4 Correlations: Working memory and fluency

	Listening span	Backward digit span	Word recall
Speech rate (words/minute)	.20 p=.078	.08 p=.616	.24 p=.038
No. of pauses	-.10 p=.392	-. 26 p=.023	-.07 p=.533
No. of repetitions	-.10 p=.396	-.20 p=.086	-.19 p=.097

Correlations between the WM measures (LST, BDST and WR) and the fluency/disfluency measures (speech rate, number of pauses and number of repetitions) for the whole sample (N=77).

Table 5 Correlations: Working memory and fluency by proficiency group

	Elementary (N = 42)			Lower intermediate (N = 35)		
	Listening span	**Backward digit span**	**Word recall**	**Listening span**	**Backward digit span**	**Word recall**
Speech rate (words/minute)	.20 p=.203	.08 p=.616	.11 p=.509	.13 p=.445	.19 p=.272	.19 p=.274
No. of pauses	-.17 p=.616	-.11 p=.483	-.16 p=.324	-.06 p=.710	-.45* p=.006	-.09 p=.593
No. of repetitions	-.13 p=.404	-.07 p=.664	-.13 p=.422	-.05 p=.789	-.34 p=.045	-.24 p=.165

Correlations between the WM measures (LST, BDST and WR) and the fluency/disfluency measures (speech rate, number of pauses and number of repetitions) by proficiency group. Note: * = significant at the .006 level (two-tailed, Bonferroni correction)

Table 6 Descriptive statistics: Overall number of self-repairs

	N	Min	Max	M	Median	SD	Total
Whole sample	77	1	15	5.6	5.0	3.1	448
Elementary	42	1	14	5.3	5.0	2.9	223
Lower Int.	35	1	15	6.4	6.0	3.2	225

Descriptive statistics for the measure of self-repairs for the whole sample as well as for the elementary and lower intermediate groups.

Table 7 Number of self-repairs by type

	Whole sample	Elementary	Lower intermediate
Different-information (D) repair	198	98 (44%)	100 (44.4%)
Appropriacy (A) repair	75	34 (15.2%)	41 (18.2%)
Rephrasing (R) repair	30	10 (4.5%)	20 (8.9%)
Error repair (E) overall	145	81 (36.3%)	64 (28.4%)
Error repair: grammatical	54	27 (12.1%)	27 (12%)
Error repair: lexical	70	41 (18.4%)	29 (12.9%)
Error repair: phonological	21	13 (5.8%)	8 (3.5%)
Total	448	223 (100%)	225 (100%)

Number and percentages of the different types of self-repairs performed by the elementary and lower intermediate participants.

Table 8 Correlations: Working memory and type of self-repair

	Listening span	Backward digit span	Word recall
D-repairs	.06 p=.570	.12 p=.293	.02 p=.860
A-repairs	.08 p=.495	.15 p=.204	.14 p=.223
R-repairs	.13 p=.262	.04 p=.759	.13 p=.247
E-repairs	-.01 p=.950	.06 p=.579	-.02 p=.871
Gram. E-repairs	.02 p=.841	.07 p=.568	.11 p=.339
Lex. E-repairs	.01 p=.900	.08 p=.509	-.04 p=.728
Phon. E-repairs	-.13 p=.247	-.23 p=.047	-.25 p=.029

Correlations between the WM tests (LST, BDST and WR) and the different types of self-repairs for the whole sample (N=77).
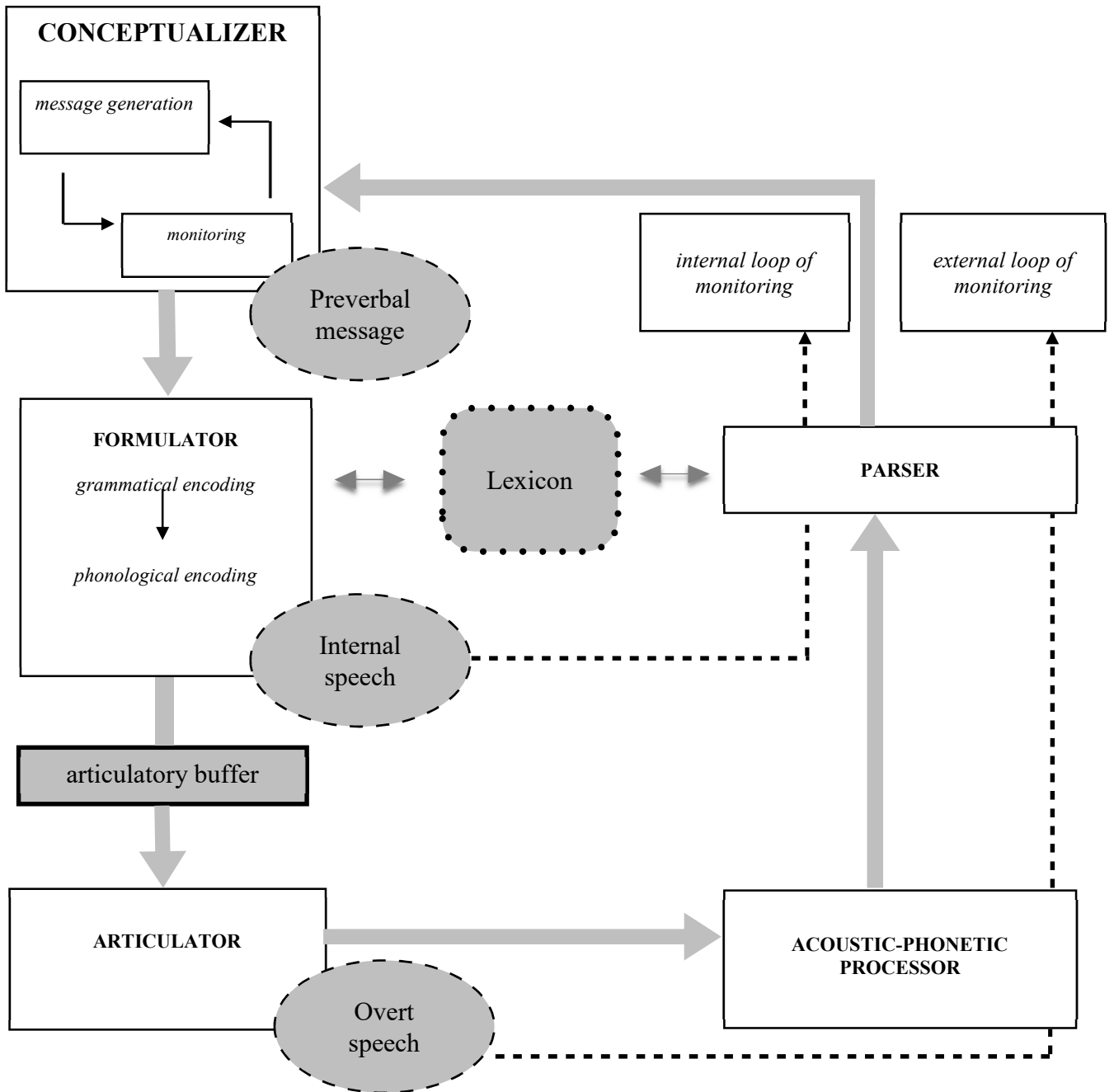
Figure 1. Levelt's speech production model (adapted from Levelt, 1989).

Figure 2. Baddeley and Hitch's multi-component WM model (adapted from Baddeley, 2012).