Journal of Teaching Language Skills (JTLS)
36(2), Summer 2017, pp. 129-162- ISSN: 2008-8191
DOI: 10.22099/jtls.2017.24678.2214

The Development of Reading and Operation Span Tasks in Persian as Measures of Working Memory Capacity for Iranian EFL Learners

Rahim Najjari *
PhD Candidate
Urmia University
mirana32@yahoo.com

Mohammad Mohammadi
Assistant Professor
Urmia University
mohammdi680@urmia.ac.ir

Abstract

This paper presented two complex span tasks in Persian as measures of working memory capacity (WMC). Firstly, the construct of working memory (WM) and possible WM measures that could assess this construct efficiently were critically reviewed. Accordingly, as measures of domain-general components of WM, a reading span task and an operation span task which were documented to assess this construct efficiently were chosen and developed. Following this, the developed tasks were conducted on 151 teenage learners in a foreign language institute in West Azerbaijan Province. Then, the administration and scoring of the measures were described step by step. The tasks were validated against each other as well as against a digit span task which assessed the domain-specific aspect of WMC as well. The result showed a strong correlation between the reading span task and the operation span task, and a moderate correlation of either of the functions with the digit span task. Consequently, both of the developed span tasks can be valid indicators of WM capacity for Iranian individuals and can be used in a wide array of research domains in second language acquisition studies.

Keywords: working memory, reading span task, operation span task.

As a further development of short-term memory, working memory (WM) is concerned with the concurrent storage and manipulation of items, while short-term memory deals with just the storage of

---

Received: 01/03/2017 Accepted: 14/11/2017

* Corresponding author

information (Baddeley, 2012). Generally, WM is linked to language learning in that it takes care of demanding and effortful tasks (Harrington 1992; McLaughlin, 1995; Wen, 2015). As for second language learning, it is known that WM effect is likely to be higher than first language acquisition (Cowan, 2015) especially in situations where individuals struggle to deal with materials on the basis of their capabilities (e.g. Sanz, Lin, Lado, Stafford, & Bowden, 2014; Tagarelli, Mota, & Rebuschat, 2015) and their limited L2 competence (Skehan, 2015). However, since WM is an imported issue from cognitive psychology (Carruthers, 2013; Conway et al., 2005), it has received different translations and increasingly "misconception and misuses" across disciplines, and thereby reliability and validity problems (Conway et al., 2005, p. 770).

Additionally, research on WM and second language learning is intricate and at the same time in infancy (Tagarelli et al., 2015). One aspect of this intricacy is methodological, i.e., the way that WM tasks are theorized, operationalized, administered, and scored (Wen, Mota, & McNeill, 2015). Although some valuable endeavors have already been done in general (Conway et al., 2005) and in particular in ELT (Juffs & Harrington, 2011; Wen, 2014; Wen et al., 2015) to account for WM issues, it seems that further attempts for a straightforward characterization of WM and WM measures in terms of construct, development, administration, scoring, and validation (e.g., Mitchell, Jarvis, O'Malley & Konstantinova, 2015; Sanz et al., 2014) will be helpful for ELT researchers and especially for Iranian researchers. More importantly, the operationalization and development of WM measures by these issues will serve as the validity concerns in Persian.

With these accounts, in this study, a general description of WM and WM measures will be provided. Then, WM measures will be critically reviewed. Following that, WM measure/s which can assess working memory capacity (WMC) in Persian will be developed, administered, scored, and validated.

General Description of WM in SLA

It is known that in the first and second language learning, phonological working memory (PWM) and executive working memory (EWM) as two main components of WM affect language learning differently. PWM is domain-specific and deals with "*acquisitional and developmental* aspects of vocabulary, formula, and grammar" (Wen, 2015, p. 56). PWM is measured by short-term memory tasks, known as simple span tasks. On the other hand, EWM, as a domain-general component, is concerned with executive and attentional functions affecting "*the processing and performance-related areas* of L2 comprehension, L2 interaction and L2 production" (Wen, 2015, p.56), and is measured through complex span tasks. EWM is also known as central executive (Baddeley & Hitch, 1974), supervisory attentional system (Norman & Shallice, 1986), executive function (Miyake & Friedman, 2012), controlled/ executive attention (Engle, Kane, & Tuholski, 1999), and functional working memory (Montgomery, 2002).

Literature Review

Critical Review of WM Measures

Although simple and complex span tasks tap different components of WM, it is known that they assess the same construct, namely the *storage system* which accounts for their correlation with other cognitive tasks (e.g. Colom, Rebollo, Abad, & Shih, 2006; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Tehan, Hendry, & Kocinski, 2001; Unsworth & Engle, 2007). The only difference is that complex span tasks measure the shared construct more than simple span tasks (Unsworth & Engle, 2007).

Notwithstanding, not all short-term memory tasks assess this construct, storage capacity, efficiently. For example, a non-word repetition task serves many functions. First of all, it represents *mostly* psycholinguistic processing (Sasisekaran, Smith, Sadagopan, & Weber-Fox, 2010) and acts as a clinical marker to diagnose language impairment (Dollaghan & Campbell, 1998). Secondly, it is considered as a measure of language proficiency or language skill (Dufva & Vauras, 2002). More importantly, the assessment of a non-word repetition task is affected by language familiarity especially long-term phonological knowledge (Roodenrys & Hinton, 2002).

Regarding complex span tasks, which are used to tap EWM, they were initially measured by Daneman and Carpenter's (1980) classic reading span (Rspan) task and its derivatives (listening and speaking span tasks), and Turner and Engle' (1989) operation span (Ospan) task. These tasks assess individuals' ability to store and retrieve materials, which are accompanied by a secondary processing task like reading a statement or solving an equation. Indeed, WM measures have dual-tasks consisted of storage/recalling and processing components which are known to rely on "a common and limited pool of resources" (Lépine, Bernardin, & Barrouillet, 2005, p. 332). The point is how the processing components in WM tasks limit storage capacity so that individuals struggle to maintain and recall items.

However, some complex span tasks which are supposedly known to measure WM capacity also have flaws. Two of these tasks are backward digit span tasks and letter-number sequencing tasks. They are measures of the subscales of Wechsler Adult Intelligence Scale (WAIS) intelligence quotient (IQ) tests. Another task is an n-back task, which is mainly employed in the fields of neuroscience, clinical, and aging research (Schmiedek, Lövdén, & Lindenberger, 2014) and sheds light on the neural substrate of WM (Kane, Conway, Miura, & Colflesh, 2007).

The main shortcoming of these tasks is that they are not by nature dual complex span tasks but single complex tasks which require second processing. In backward digit span tasks, participants hold a series of digits and transform their order. Moreover in letter-number sequencing tasks, they have to keep a mixture of digits and letters and sort them numerically and alphabetically. In n-back tasks, participants monitor a verbal/nonverbal string of stimuli and report whenever the stimulus presented goes with the stimulus which has already been displayed n-stimulus back. In fact, the retrieval of target items depends on recognition rather than recalling (Redick & Lindsey, 2013). Moreover, in all of these three tasks, the processing task depends initially on the storage capacity of individuals. What is missing in these tasks is that processing and storage components cannot be simultaneously tapped to tackle what is referred to as a common cognitive and limited pool of resources between processing and storage capacity (Lépine et al., 2005).

Maybe the inseparability of dual-tasks, storage and processing components, in backward digit span tasks, letter-number sequencing tasks, and n-back tasks has caused further shortcomings for these tasks as measures of WMC. For example, backward digit span tasks have a weak correlation with other WM measures (Hilbert, Nakagawa, Puci, Zech, & Bühner, 2014). They are also affected by strategy use especially cognitive strategies which affect the processing component so that individuals resort either to verbal or visual strategies to recall and sort to-be-remembered items (Dunn, Gaudia, Lowenherz, & Barnes, 1990). Consequently, those with visual strategies outperform in backward digit span tasks (Hilbert et al., 2014).

As for letter-number sequencing tasks, they share most of their variance with forward and backward digit tasks (68%) (Crowe, 2000). Arguably, the criticisms of backward digit tasks also hold true for letter-number sequencing tasks as they are documented to measure visuospatial aspects of tasks as well as processing speed (Crowe, 2000).

Regarding n-back tasks, they have a weak or nonsignificant-to-weak correlation with simple and complex span tasks (Jaeggi, Buschkuehl, Perrig, & Meier, 2010; Kane et al., 2007; Redick & Lindsey, 2013). According to Schmiedek et al. (2014), this low correlation is since (i) they measure distinct constructs, (ii) task-specific variance dominates individual differences, and (iii) measurement errors also restrict this correlation. Besides, a mixed reliability (low or high) is reported for n-back tasks (e.g. Jaeggi et al., 2010; Schmiedek, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009; Unsworth, 2010).

Another complex span task is math (computation) span tasks developed by Salthouse and Babcock (1991). The task is presented with

2-6 strings of arithmetic problems, and individuals have to do the issues and remember the last digits of each problem. Although this task shares similarities in processing with Ospan tasks, i.e., a nonverbal processing component, digits as to-be-remembered items are more prone to chunking strategies (Cowan, 2013) especially the range of span is also limited, i.e., from 2 to 6 strings.

Counting span tasks (Case, Kurland, & Goldberg, 1982) are also complex span tasks. In these tasks, participants are presented with the set of 2-6 screens of two color dots, say yellow and blue. They have to count yellow dots and remember their totals. This task involves both simultaneous processing and storage, and like Rspan and Ospan tasks, enjoy reliability and validity (Conway et al., 2005). Notwithstanding, this task is concerned with the linguistic and arithmetic learning difficulties (Hitch & McAuley, 1991) and is the complex version of digit span tasks. Danahy, Windsor, and Kohnert (2007) have used it for children with language impairment. They believe that it identifies this problem through online processing efficiency. In fact, the task fixes the deficiency of digit span and non- word repetition tasks.

As for the use of WM measures in Iran, especially for ELT purposes, to the best of our knowledge, these tasks are not well received in Persian. This is because Iranian researchers have used short-term memory tasks or one of the tasks as mentioned earlier which cannot represent WMC efficiently. Or in some cases, they have developed their Rspan or Ospan tasks without introducing them to other researchers. Moreover, the only Rspan task is that of Shahnazari (2013) which has followed the earliest version of Daneman and Carpenter's (1980) reading span task. This task suffers from some serious drawbacks. First of all, it is dependent on individuals' language proficiency both concerning processing and recalling. Secondly, the number of syllables, words, and the length of sentences are not controlled in the processing component; there are no punctuation marks so that sentences are received ambiguously; and sentences are processed for semantic anomaly or just for one or two syntactic judgments like addition of one element that was not necessary or wrong use of propositions. Finally, for recalling, the last word of each sentence has to be kept and recalled. Thus, some individuals with varying degrees of language proficiency are likely to coin words from the theme of each sentence to recall words (Conway et al., 2005). Moreover, all sentences in Persian conclude with verbs. In other words, in Shahnazari's Rspan task, individuals have to recall just verbs, and since verbs carry the central themes of each sentence, they can be easily recalled. Therefore, these deficiencies are needed to be accounted for.

Beyond giving a simple and straightforward understanding of WM construct and critical review of the existent WM measures, the main purpose of this study is to develop a Rspan task and an Ospan task in Persian while providing guidelines for the administration and scoring of WM measures. The rationale for the development of both tasks is that they enjoy high validity and reliability (Conway et al., 2005; Redick et al., 2012; Unsworth, Heitz, Schrock, & Engle, 2005) so that any significant relationship between these two measures will account for their concurrent validity. Additionally, since an Ospan task is nonverbal and a Rspan task is verbal, any considerable correlation will also account for the validity of each of the tasks in term of taking care of verbal and nonverbal accounts of WMC.

A simple span task, a forward digit span (Dspan) task is also included for the further concurrent validity of the developed Rspan and Ospan tasks. Beyond the fact that both simple and complex span tasks assess the same construct (e.g. Colom, Rebollo, et al., 2006; Conway et al., 2005; Kane et al., 2004), it is known that simple span tasks almost always have either strong or moderate correlation with complex span tasks (Colom, Shih, Flores-Mendoza, & Quiroga, 2006; Conway et al., 2002; Kane et al., 2004; Unsworth & Engle, 2007). Besides in comparison with other simple span tasks, a Dspan task is less sensitive to language familiarities (Thorn & Gathercole, 1999), and as a standard test (Jones & Macken, 2015), it presents a clear picture of WMC (Gathercole & Alloway, 2008). Besides, a digit span task represents simultaneously both verbal and nonverbal versions of simple span tasks (Dehaene, 2002).

Method

Participants

The participants were 151 (78 females and 73 males) English language learners at a private language institute in West Azerbaijan Province. They were selected through nonprobability convenience sampling design and were offered some extra lessons as compensation for their participation. Socially and economically, they belonged to the middle-class society. Their mother tongue was Turkish, but they could not read and write in Turkish. However, given the fact that Persian in Iran was the national language acquired from the early childhood, and that the medium of communication and instruction was officially in Persian, the participants could read and write in Persian. Additionally, they were high school students and had at least six years of experience in learning school subjects through Persian. As far as their English language proficiency was also concerned, they had more than four years of experience in learning English, and most of them were attending pre-intermediate courses. Their age ranged from 13 to17 (mean =14.52, SD=1.21), roughly with an equal distribution of females: 13(n=18), 14(n=23), 15(n=19), 16 (n=11), and 17(n=7), and males: 13(n=16), 14(n=25), 15(n=16), 16 (n=10), and 17(n=6).

Materials: The development of Persian Rspan and Ospan tasks

A: Catering for processing content of the developed tasks

Note that complex span tasks were consisted of processing and recalling components. The processing component of the Ospan task was taken from Turner and Engle's original Ospan task (1989) as it was nonverbal and language independent, thereby presenting no problems in Persian. However, for the processing component of the Rspan task in Persian, some materials were developed, and some were taken from the processing section of Shahnazari's (2013) Rspan task, which contained general and nontechnical knowledge and seemed to be appropriate concerning content. As already mentioned, the sentences in Shahnazari's (2013) Rspan task had some technical problems. Firstly, they were controlled for the number of words and syllables as recommended (Van den Noort, Bosch, Haverkort, & Hugdahl, 2008) so that for each sentence, the number of words ranged from 12 to 17 words and the number of letters from 55 to 77. Although a range of 20 to 22 syllables is also advised (Van den Noort et al., 2008), it is not possible for Persian words as most of the syllables in Persian are made with short vowels, thereby making Persian words multi-syllables. In all, the average number of words, letters, and syllables of sentences was 15, 58, and 29, respectively.

Then, in line with Turner and Engle's (1989) version of Rspan task, syntactic criteria like wrong/inappropriate uses of tenses, wrong uses of pronouns in term of the agreement with plural and singular references, and so on were included in the sentences. After that, sentences were tested on a group of high school students who had similar characteristics like the participants of the present study. It was noticed that some sentences had more than one interpretation. Thus, with the help of an expert, a PhD student in Persian literature, the sentences were revised and punctuated in some cases. Then, forty-eight sentences for the Rspan task and an equal number of 48 mathematical equations were chosen for the Ospan task. Half of the sentences or equations were correct and the other half incorrect. Additionally, 12 sentences or equations were selected for the practice sessions for each task.

B: Selection of isolated letters for recalling components of Rspan and Ospan tasks

Given that the use of separate letters is strongly advised for the recalling parts of WM measures (Kane et al., 2004), and that both Rspan and Ospan tasks can jointly share isolated letters for their storage components, we focused on the use of isolated letters for the storage components of the tasks. To this end, we methodologically analyzed the choice and arrangement of English to-be-remembered letters presented by Millisecond Software (www.millisecond.com), a leading provider of software for psychological testing. In Millisecond Software models for Rspan and Ospan tasks, 12 English consonant letters were used: F, K, P, S, H, L, Q, T, J, N, R, Y. We tried to select their close counterparts in Persian. Since letters S and T could be represented with other letters in Persian as well, we included the ones that were more phonotactically incongruent with the other selected letters (e.g. ط instead of ت, and صinstead of س، ث) so that participants could not use any strategies to make up words, syllables, or consonants clusters. The final selected letters were:ص، پ، ع ، ه، ن، ط، ف، ش، ل، ک، غ ، ج The letters are pronounced at NOON for ن, TA for ط, FE for ف, SHEEN for ش, LAM for ل, KAF for ک, GHAIN for غ, JEEM for ج, SAD for ص, PE for پ, EAIN for ع, and HE for ه.

C: What do the Rspan and Ospan tasks look like in Persian?

After selecting isolated letters, we put sentences and equations along with the letters through span ranges. Ranges of two to five or six to-be-remembered items in complex span tasks are common, and a preferable range is from two to five (Conway et al., 2005). Accordingly, we observed the upper limit range and chose three to five letter-span ranges. For each span range, four-letter strings were developed, and each of 12 Persian letters was equally distributed four times in either of the tasks. Following this, letters were carefully arranged so that those with similar rhymes and beats did not follow each other within a set (span). For example پ، ج، غ، the letters read as GHAIN, JEEM, and PE respectively were arranged in a set of a three-letter string. The letters ش، ص، ط، ک read as SHEEN, SAD, TA, and KAF, respectively were put in a set of a four-letter string. And the letters ش، ل، ع، ک، ه read as SHEEN, LAM, EAIN, KAF, and HE, respectively were placed in a set of a five-letter string. This procedure for arrangement could prevent participants from using any strategies to make consonant clusters or to coin words with the letters.

Additionally, the names of letters were put beside each letter within parentheses like (م) میم. This was methodologically innovative and was done for many reasons. Firstly, the participants were familiar with languages of Persian, Arabic, and English so that they would probably mistake not only letters of one language for letters of another one but also sounds of one language for the sound of another one. Secondly, we were not sure that most of the participants could correctly utter the names of the letters of their language, though they knew what the letters sounded. Thirdly, the initial introduction of the names of letters would make participants pronounce them as words rather than as sounds. This would also make it difficult for them to make up words, to make sentences with names of letters, or to associate the names of letters with each other or other objects to remember them. Finally, reading aloud the sounds of letters takes a shorter time than reading the name of letters. Thus, putting the name of letters and then letters in parentheses would guarantee that all participants read them aloud as individual words. This would also improve the reliability of tasks as all participants had to follow the same procedure.

To sum up, the developed Rspan task had (1) four sets of three sentences followed by three isolated letters for each set, (2) four sets of four sentences followed by four isolated letters for each set, and (3) four sets of five sentences followed by five isolated letters for each set. The Ospan task also had the same 12 sets, but instead of sentences, equations were followed by letters. Additionally, for each task, a set of three, four, and the five-span range was used for practice sessions. Sample sets of three, four, and five-span ranges representing the general content and format of the tasks were given in Appendix A and B.

For the Dspan task, a 3 to 9-digit span range was selected. Each span range had four trails of numbers, and in all, there were 35-digit strings: 28 digit-strings for the test and seven digit-strings for the practice session (Appendix C).

Pilot Study for the Rspan Task

Indentation of the first paragraph the pilot study, 10 participants individually took the Rspan task on a computer screen and were immediately interviewed. Slides were set on 7-second length by the previous Rspan task developed by Shahnazari (2013). When we got participants' feedback on this way of measurement and reading times, most participants said they did not feel comfortable in front of a computer since they had never taken a test in such a condition. They also complained about the inadequacy of time for each slide in that they were psychologically and emotionally under time pressure. They said that they were overwhelmed by maintaining and monitoring the passage of time so that they had to make a kind of compromise between processing and recalling. They preferred to attend in one set for processing and in another one for recalling. Therefore, we had to give the tasks through paper and pencil formats; each set of the Rspan and Ospan tasks was developed in a booklet in which every sentence and equation was presented via a sheet followed by an isolated letter. Finally, at the end of the booklets, there was an instruction requiring the participants to recall and write letters in the order presented in each set.

Data Collection Procedures

The participants firstly took the complex span tasks and then the simple span task. And for the administration of the complex span tasks, the participants were divided into two groups of A and B. Group A was requested to take first the Rspan task and then the Ospan task, whereas group B was given first the Ospan task and then the Rspan task. This was due to the fact both the duties had jointly isolated letters for recalling. Thus, if participants took one of them before the other one, it could have practice effect for the different task, thereby affecting the performance of participants.

Following Conway et al.'s (2005) recommendation, individual administration was considered. As already mentioned, each task had 12 sets (booklets). The experimenter had to give booklets one by one. In the Rspan task, for instance, the experimenter gave a booklet. As soon as the participant got it and was about to take, the experimenter started to measure test time in milliseconds by pressing the button of a stopwatch.

Following this, the participant had to read each sentence aloud and tick whether it was correct or not. Then, the participant turned a page to the to-be-remembered letter, read it aloud. S/he had to keep on reading and processing sentences, and reading aloud letters to finish doing the booklet. At this time, the experimenter immediately got the booklet and the participant wrote the isolated letters on a sheet in the order presented. As soon as the participant was finished recalling and writing isolated letters, the experimenter gave the next set randomly. In the random presentations of sets, participants do not know which span range they are going to take, so they cannot trust on using strategies for recalling longer span ranges (Engle, Cantor, & Carullo, 1992). Meanwhile, in line with Conway and Engle' (1996) procedure, the experimenter encouraged the participant during the administration to take both the processing and recalling of the tasks as quickly as possible. Besides, when the participant was observed to read sentences silently to process, s/he was immediately reminded to keep reading them aloud. This was due to the fact that in silent processing of sentences, the participant might devote more time to letters to rehearse them accumulated and to make up a kind of word, pattern, rhyme, and consonant cluster with letters. The participant was also monitored so as not to turn a page or pages to refresh decaying letters. All of these considerations were taken to prevent the overestimation of participants' WMC. Meanwhile, the test went on and the experimenter kept on giving and collecting the booklets (sets) until the participant took the last set of the task and finished writing isolated letters. At this time, the experimenter pushed the button of the stopwatch to stop time, and then recorded the test time. Finally, the participants were given a sheet of paper and requested to jot down any strategy they might have employed for recalling letters immediately after they took the whole sets of the tasks.

As for the Dspan task, it was also individually administered, and the presentation of one digit per second was preferred to the quick presentation (Phye & Pickering, 2006). The experimenter read each trial steadily and monotonously and dropped his voice slightly to signal the last number. The participants had to listen to the numbers and then wrote them in the order presented on a sheet. Like the administration of the Rspan and Ospan tasks, the participants had to take the whole task.

Data Analysis Procedure

The scoring of WM tasks is complicated (Juffs & Harrington, 2011), and its interpretation is a matter of controversy (Friedman & Miyake 2004; Waters & Caplan, 2003). In all, the final output of simple and complex span tasks is participants' 'span or capacity' score, which generally can be reported through (i) absolute scoring; and (ii) partial-credit unit (PCU) scoring.

In the absolute scoring, the span/capacity score is represented either as the sum of all correctly recalled sets of the to-be-remembered items (Unsworth & Engle, 2007) or "the last item size recalled with a specific probability (say, four out of five items)" (Conway et al., 2005; p. 774). In the former way, if a participant can recall the sets of 3 to 4 to-be-remembered items (e.g., isolated letters for the Rspan and Ospan tasks, and digits for the Dspan task in the study), his or her score will be equal to 7, while for the latter way of absolute scoring, it will be 4. This way of scoring is commonly used when WM tasks are partially administered. That is, participants are stopped from continuing the tasks whenever they cannot recall a specific span range. Participants' scores were reported through the second way of this procedure in the study. However, due to its shortcomings explained later, we focused on the PCU scoring.

In PCU scoring, the average proportion of to-be-remembered items (e.g., recalled letters or digits in the study) is calculated. When WM measures are fully administered, PCU scoring is used. It is more advantageous as it presents better psychometric properties and is also sensitive to individual differences (Conway et al., 2005; Friedman & Miyake, 2005; Unsworth & Engle, 2007). In this way of scoring for complex span tasks, each set receives one mark. For example, the number of isolated letters in the study differed within each set so that there were four sets of three letters, four sets of four letters, and four sets of five letters. Therefore, in a set of three, four, and five isolated letters, the correct recalling of each letter would receive .33 .25, and .20 of the score 1, respectively. If someone could recall all the three letters in a set of three, s/he would receive score 1 one, while for the correct recalling of two letters, .67 of the score 1. Finally, the scores of all sets were summed up and divided by the number 12-the total sets. That score was considered as a participant's WMC score.

For PCU scoring of the Dspan task, the correct recalling of each digit in a string of 3, 4, 5, 6, 7, 8, and nine digits received .33, .25, .20, .17, .14, .13, and .11 of the score 1, respectively. Then, the average recalling of four strings within each set was calculated. Finally, since there were7 sets of digit strings, the average of 7 sets was reported as a participant's final score.

Additionally, Waters and Caplan's (1996) scoring procedure for complex span tasks was applied. Interestingly, a review recently done by Sagarra (2017) reveals that most of WM studies which have reported WM effect on SLA domains have followed this procedure for scoring. From this way of scoring, the z-scores of recalling isolated letters, processing, and (reaction) time multiplied by -1 were calculated, averaged, and reported as a participant's WMC. Remembering scores were computed through PCU scoring. Processing scores were reported by the proportion of correct answers to all the questions. That is, the proportion of correctly answered sentences to all sentences in the Rspan task, and of correctly verified equations to all equations in the Ospan task. Finally, since the whole time on doing the complex span tasks was measured, to get the reaction time, the time spent on the whole test was subtracted from the mean time of all the participants.

Results and Discussions

The overall responses on using strategies during the administration of the complex span tasks suggested that the participants were unable to use effective strategies to recall isolated letters. The only strategy that a few participants mentioned was relating the letters with the initials of their names. However, since letters were evenly distributed across the span ranges with the four times of distribution in the Rspan or Ospan tasks, the strategy use could not make a significant difference. Additionally, the test time of each task was correlated with the processing and recalling scores of that task. The results demonstrated that the correlation of test time with the processing and recalling scores of the Rspan task was -.30 and -.10 and of the Ospan task was -.12 and -.14, respectively. Statistically, the negative correlation shows that participants have taken these tasks without any effective strategy use (St Clair-Thompson, 2007). In other words, the participants' WMC was not overestimated. Besides the successful administration of Rspan and Ospan tasks regarding preventing the participants from strategy use, all of these would seem to suggest that the selection, arrangement, and presentation of Persian isolated letters had worked efficiently for the recalling components of the developed Rspan and Ospan tasks. However, if it was noticed that participants had used any strategy or there had been no negative correlation between test time and the other two components of complex span tasks, the 80 percent of correct processing components of the developed Rspan and Ospan tasks had to be considered as a criterion for reporting participants’ WMC.

Ospan and Rspan tasks

Tables 1 and 2 display the descriptive statistics of the developed tasks. The first column in each table shows the average of test time which is longer in the Rspan task in comparison to the Ospan task (760 vs. 573). Computing this average time by the number of sentences or equations shows that reading aloud each sentence or equation and determining its accuracy, and reading aloud each letter and writing it on order presented take15 seconds for the Rspan task and 11 seconds for the Ospan task. This average time can be considered as a criterion for computerized versions of these tasks.

Table 1.

Descriptive Statistics of the Rspan Task

	Test time in seconds	Processing score	PCU scoring of 3 letters	PCU scoring of 4 letters	PCU scoring of 5 letters	Total PCU scoring	z score (processing, recalling, time)
Mean	760.33	.77	.80	.62	.57	.66	33.26
Std. Error of Mean	9.07	.01	.01	.01	.01	.01	.47
Std. Deviation	110.36	.13	.14	.16	.16	.13	5.73

Table 2.

Descriptive Statistics of the Ospan Task

	Test time in seconds	Processing score	PCU scoring of 3 letters	PCU scoring of 4 letters	PCU scoring of 5 letters	Total PCU scoring	z score (processing, recalling, time)
Mean	573.31	.91	.81	.71	.59	.70	33.34
Std. Error of Mean	8.02	.00	.01	.01	.01	.01	.55
Std. Deviation	97.628	.10	.16	.18	.16	.15	6.80

The second column in each table deals with the processing scores. The average processing score for the Rspan task was .77 and for the Ospan task was .91. The foremost cause of the difference is probably due to the fact that processing is "highly complex and demanding" (Lépine, Bernardian & Barrouillet, 2005, p. 336) in Rspan tasks, while it is automatic in Ospan tasks (Barrouillet & Fayol, 1998) as it involves direct retrieval from memory and needs a restricted pool of equation knowledge (Lépine et al., 2005).

The processing score of complex span tasks is not usually taken into account in reporting individuals' WMC (Daneman & Carpenter, 1980) as its main function is to prevent participants for using any strategy and to make recalling of the to-be-remembered stimuli challenging. Even in the later versions of WM span tasks (Lépine et al., 2005), the sentence judgment of Rspan and the verification of equations in Ospan tasks were relegated to simple tasks of adding or subtracting. However, sometimes to make sure that participants have taken the processing section seriously and have not used any strategy for recalling, the average of .80 or .85 percent of accurate processing is considered as a criterion (Conway et al., 2005; Turner & Engle, 1989) for reporting recall scores as participants' WMC. Accordingly, the average of .77 percent accurate processing in the Rspan task and .91 percent in the Ospan task indicate that the processing tasks were successful in involving participants and preventing them from seizing any effective strategy to recall isolated letters. Additionally, as already discussed the tasks were almost taken free of strategy use.

As for absolute scoring computed through the second way of this procedure (see Data analysis procedure), the number of participants who could successfully recall three isolated letters was 35, for four isolated letters were 12 and for five isolated letters was 1 in the Ospan task. For the Rspan task, and the number of participants who managed to retain and recall three isolated letters was 25, while no one could recall four isolated letters and only two participants could remember five separate letters. A possible explanation for the odd performance of participants is due to the shortcomings of absolute scoring. Indeed, this score was reported for all the four sets of three to five letter span ranges. If only one set, two sets or even three sets of each span range were considered, the number of participants for each span range would rise. To support this, they were participants who had recalled three sets of four sets entirely, but they had missed recalling just one letter in the fourth set. For this reason, the four-span score could not be reported for them. Or considering the Rspan task, only two of out of 151 participants were able to recall five isolated letters. This number is insignificant and additionally may be because as already mentioned the recalled letters maybe were similar to initials of participants' names. Moreover, only 35 out of 151 had remembered three letters entirely in the Ospan task. This means that 116 out of 151 participants had a span range below three. So, it cannot account for the individual difference of 116 participants in

WMC. In fact, absolute scoring fails to render an extensive range of WMC/span scores (Juffs & Harrington, 2011) and is not sensitive to individual differences (Conway et al., 2005; Oberauer & Süß, 2000).

On the other hand, as a strength of PCU scoring is that recalling each letter receives a mark depending on the range of spans. Thus, in PCU scoring, there is an insignificant difference between a participant who has missed only one letter in four letter-span ranges and a participant who has recalled all the five letter-span ranges. Since individuals' mean score in all sets is reported as their WMC, recalling one letter even by chance can not make a significant difference especially in high span ranges in which remembering a letter receives less mark than a recalling a letter in low span ranges.

Regarding columns of 3, 4, and 5 of each table, they give information on PCU scoring of 3 to 5 to-be-recalled letters. For the Ospan task, the average recalling scores for 3 to 5 isolated letters show a logical descending order from .81 percent to .71 and then to .59. On the other hand, the average recalling scores for 3 to 5 isolated letters for the Rspan task show a radical decline from 3 to 4 isolate letters (.80 vs. 62) and a slight decline to 5-letter span ranges, .57. What is remarkably interesting is that for both the Rspan and Ospan, the average of recalled letters for spans of 3 and 5 is closely the same (Tables 1 & 2). This appears to imply that both of the tasks have reliably tapped the lower and higher span ranges.

The sixth column of each table displays total PCU scoring, which is the average of PCU scoring of three-to-five recalled letters, reported as participants' WMC. Participants slightly outperformed in the Ospan task compared with the Rspan task (70 vs. 66). Finally, the last columns show the scoring of the tasks via Waters and Caplan's (1996) scoring procedure. Interestingly, the scores are the same for both Rspan (33.26) and Ospan task (33.34) and are highly correlated with scores of the tasks reported through (total) PCU scoring. The correlation is .86 for the Rspan task and .71 for the Ospan task (Table 4). This would seem to indicate that both tasks have likely tapped the construct of WM appropriately so that participants' performance tends to be the same even with the inclusion of time and processing scores.

Dspan Task

As for the absolute span score for the forward Dspan task, nearly most of the participants recalled the sets of three and four-digit strings, but for the sets of five-digit strings 90 (60%) participants, six-digit strings 24(16.2%) participants, seven-digit strings 9(6%) participants, and eight-digit strings, only one participant could recall them correctly. No one was able to recall nine-digit strings. These results are in agreement with Cowan (2001) in that the real number of items that can be consciously attended and worked with is four plus or minus one. However, as mentioned earlier, this way of scoring has shortcomings.

The average of PCU scoring for the sets of 3-digit strings to the sets of 7,8 and 9-digit strings was .87, .81 and .75, respectively (Table 3). The last column was reported as the mean score of participants in the Dspan task. Interestingly enough, a pattern of .6 percent differences existed from sets of 7-digit strings to 9-digit strings.

Table 3.
Descriptive Statistics of the Dspan Task

	PCU scoring of 7 digit-strings	PCU scoring of 8 digit-strings	PCU scoring of 9 digit-string
Mean	.87	.81	.75
Std. Error of Mean	.0070	.0078	.0078
Std. Deviation	.08	.09	.09

Validity of the Developed Tasks

As can be seen from Table 4, from the total PCU scoring for the three WM measures in the study, there was a strong correlation between the Rspan task and the Ospan task (r=.66, n=151. P<000 (two-tailed), and a moderate correlation of the Rspan task and the Ospan task with the Dspan task, i.e. (r=.41, n=151. P<000 (two-tailed) and (r=.43, n=151.

P<000 (two-tailed), respectively. Remarkably enough, when Waters and Caplan's (1996) scoring was also considered, scores of both tasks had a highly strong correlation with the scores reported through the total PCU scoring procedure (Table 4). Besides, nearly the same magnitude of the strong correlation between the Rspan task and the Ospan task (r=.65, n=151. P<000 (two-tailed) was observed. The Dspan task was also moderately correlated with the Ospan task (r=.43, n=151. P<000 (two-tailed), and the Rspan task (r=.53, n=151. P<000 (two-tailed).

Table 4.
Pearson's Product Moment Correlations of the Rspan, Ospan, and Dspan tasks

	Rspan task Total PCU	Rspan task Average Z scores	Ospan task Total PCU	Ospan task Average Z scores	Dspan task Total PCU
Rspan task Total PCU	-	.859**	.66**	.60**	.41**
Rspan task Average Z scores	-	-	.606**	.651**	. 53**
Ospan task Total PCU	-	-	-	.712**	.431**
Ospan task Average Z scores	-	.651**	-	-	.439**

**. Correlation is significant at the 0.01 level (2-tailed).

Note that in *total PCU*, WMC was reported just by recall scores of measures while in *average Z scores*, i.e., Waters and Caplan's (1996) scoring procedure, WMC was reported from the average z scores of recalling, processing, and time multiplied by -1.

The strong correlation between the Rspan and Ospan tasks, and the moderate correlation between either of these two complex span tasks with the Dspan task in either way of scoring substantiate previous findings in that both simple and complex span tasks assess the same construct (e.g., Colom, Shih et al., 2006; Conway et al., 2002; Kane et al., 2004; Unsworth & Engle, 2007). According to Conway et al. (2005), this implies the tasks are probably influenced by something stable. In other words, the study seems to have been successful in tapping a constant construct of WM. Furthermore, the correlation between the Rspan task and Ospan task can serve as an index for their concurrent validity and convergent validity (Conway et al., 2005). This is of a particular significance in that Ospan tasks are a non-verbal representation of participants' WMC, while Rspan tasks account for the verbal manifestation of WMC.

Conclusion

WM studies have carried positive sentiments in ELT during recent years. One of the main requirements to conduct such studies is to develop WM tasks in participants' native language or the language in which they are competent (Ardila, 2003). To this end, this study was an attempt to give a straightforward characterization of WM and WM construct, review WM tasks, and accordingly develop and validate Persian versions of Ospan and Rspan tasks- the two widely used WM tasks in L2 studies.

Taken all together, both the developed Persian Rspan and Ospan tasks seem to represent WMC of Iranian foreign language learners. The Ospan task is the first version of these kinds of tasks in Persian, while the Rspan task is the second attempt compensating for the shortcomings of the first version developed by Shahnazari (2013) and thereby improving it by the standards of recent WM tasks. These tasks may be used for investigating the relationship between WMC and learning school subjects as well. Furthermore, the study has rendered a methodological contribution through a fairly rigorous methodology in the development, administration, and scoring of the developed tasks. This can cater for the methodological concerns (e.g., Gass & Lee, 2011; Linck, Osthus, Koeth, & Bunting, 2014; Wen, 2014) and standardization of WM tasks (Mitchell et al., 2015) which are frequently echoed. Additionally, in order to get a more valid and comprehensive assessment accounting for verbal and nonverbal representation of WMC of participants, it will be really efficient to conduct both of these tasks and average their scores as recommended (Conway et al., 2005; Waters & Caplan, 2003) or merge these two WM tasks into one task. However, if it is not possible to conduct both of the tasks or merge them into one job, conducting and correlating a Dspan task along with either of them can methodologically determine that researchers are assessing WM.

Regarding limitations of the study, some motivations were provided for administering the developed tasks through a paper and pencil format. However, it could methodologically be sound if the developed tasks were computerized or run through computers. The study has also investigated a narrow age range. Thus, the findings might not be transferable to situations where a relatively broad age range of participants is involved. However, the results of the study represent WMC of a suitable age range on the ground that WMC starts to develop when individuals are five years old, and that WMC is fully developed when individuals are around 16 years old (Wright, 2015). Furthermore, most of the foreign language learners in Iran are school students who also take English classes in foreign language institutes. Notwithstanding, further research is also required to validate the developed tasks with broad age ranges for individuals who are over 17 years old and under 45 years old since WMC "starts to decline beyond the mid-adult" (Wright, 2015, p. 290). Besides, although the new Rspan task is validated against the Ospan task as a measure of nonverbal WM tasks, the validity of findings will be further strengthened if the new Rspan task is conducted on participants who are competent only in Persian or on other bilinguals (e.g., like Kurds, Arabs, etc.) who live in Iran. Even, it would be interesting to conduct the Rspan task on different bilinguals as well as on Persian monolinguals to make a cross-linguistic comparison. Two other areas for further research which we are currently working are 1) the validation of a Persian listening span task through changing the new Rspan task into an oral format in which participants listen to the recorded sentences and then determine the accuracy of sentences and finally list the letters in orders heard; and 2) computerized versions of the Ospan task, Rspan task, and listening span task in which different ways of scoring can be reported as well.

References

Ardila, A. (2003). Language representation and working memory with bilinguals. *Journal of Communication Disorders*, 36, 233-240.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1-29.

Baddeley, A.D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning & motivation* (pp. 47-90). New York: Academic Press.

Barrouillet, P., & Fayol, M. (1998). From algorithmic computing to direct retrieval: Evidence from number and alphabetic arithmetic in children and adults. *Memory & Cognition*, *26* (2), 355-368.

Carruthers, P. (2013). Evolution of working memory. *Proceedings of the National Academy of Sciences of the United States of America 11 (Supplement 2),* pp. 10371-10378.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*(3), 386-404.

Colom, R., Rebollo, I., Abad, F. J., & Shih, P. C. (2006). Complex span tasks, simple span tasks, and cognitive abilities: A reanalysis of key studies. *Memory & Cognition*, *34* (1), 158-171.

Colom, R., Shih, P. C., Flores-Mendoza, C., & Quiroga, M. A. (2006). The real relationship between short-term memory and working memory. *Memory*, *14*, 804-813.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163-183.

Conway, A.R.A., & Engle, R.W. (1996). Individual differences in WM capacity: More evidence for a general capacity theory. *Memory, 4*, 577-590.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D., Wilhelm, O. & Engle, R. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12* (5), 769-786.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *The Behavioral & Brain Sciences*, *24*(1), 87-114.

Cowan, N. (2013). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review*, *26*(2), 197-223.

Cowan, N. (2015). Second language use, theories of working memory and the Venetian mind. In Z. E. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp.29-40). Bristol: Multilingual Matters.

Crowe, S. F. (2000). Does the letter-number sequencing task measure anything more than digit span? *Assessment*, *7*(2), 113-117.

Danahy, K., Windsor, J., & Kohnert, K. (2007). Counting span and the identification of primary language impairment. *International Journal of Language & Communication Disorders*, *42*(3), 349-365.

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

Dehaene, S. (2002). Verbal and nonverbal representations of numbers in the human brain. In M. Galaburda, S. M. Kosslyn, & Y. Christen (eds), *The Languages of the brain* (pp. 179-190), Cambridge, MA: Harvard University Press.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, & Hearing Research, 41, 1136-1146.*

Dufva, M., & Vauras, M. (2002). Promoting at-risk pupils' foreign language literacy learning. *Precursors of Functional Literacy. Amsterdam: John Benjamins*, 317-337.

Dunn, G., Gaudia, L., Lowenherz, J., & Barnes, M. (1990). Effects of reversing digits forward and digits backward and strategy use on digit span performance. *Journal of Psychoeducational Assessment, 8*(1), 22-33.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18*, 972-992

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* (pp. 102-134). New York: Cambridge University Press.

Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory & Language, 51*(1), 136-158

Friedman, N.P., & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods, 37*(4), 581-590.

Gass, S., & Lee, J. (2011) Working memory capacity, inhibitory control, and proficiency in a second language. In M. Schmid. & W. Lowie (Eds.), From *structure to chaos: Twenty years of modeling bilingualism: In honor of Kees de Bot* (pp. 59–84). Amsterdam: John Benjamins.

Gathercole, S. E., & Alloway, T. P. (2008). Working memory and classroom learning. In K. Thurman, & K. Fiorello (Eds.), *Cognitive development in K-3 classroom learning: Research applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Harrington, M. (1992). Working memory capacity as a constraint on L2 development. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 123-135). Amsterdam: North Holland.

Hilbert, S., Nakagawa, T. T., Puci, P., Zech, A., & Bühner, M. (2014). The digit span backward task. Verbal and visual cognitive strategies in working memory assessment. *European Journal of Psychological Assessment*.

Hitch, G. J., & McAuley, E. (1991). Working memory in children with specific arithmetical learning difficulties. *British Journal of Psychology*, *82*(3), 375-386.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the n-back task as a working memory measure. *Memory*, *18*, 394-412.

Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition*, *144*, 1-13.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, *44*(2), 137-166.

Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the N-back task: a question of construct validity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *33*(3), 615-622.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189-217.

Lépine, R., Bernardin, S., & Barrouillet, P. (2005). Attention switching and working memory spans. *European Journal of Cognitive Psychology*, *17*(3), 329-345.

Linck, J. A., Osthus, P., Koeth, J. T. & Bunting, M.F. (2014). Working memory and second language comprehension and production: *A meta-analysis. Psychonomic Bulletin & Review, 21(*4), 861-883.

McLaughlin, B. (1995). Aptitude from an information processing perspective. *Language Testing*, *11*, 364−381.

Mitchell, A. E., Jarvis, S., O’Malley, M. & Konstantinova, I. (2015). Working memory measures and L2 proficiency. In Z. E. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second* *language acquisition and processing* (pp. 270-284). Bristol: Multilingual Matters.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*(1), 8-14.

Montgomery, J. W. (2002). Understanding the language difficulties of children with specific language impairments: Does verbal working memory matter? *American Journal of Speech-Language Pathology, 11*(1), 77-91.

Norman, D., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R.J. Davidson, G.E. Schwarts, & D. Shapiro (Eds.), *Consciousness and self-regulation* (Vol. 4; pp. 1-18). New York: Plenum Press.

Oberauer, K., & Süß, H. M. (2000). Working memory and interference: A comment on Jenkins, Myerson, Hale, and Fry (1999). *Psychonomic Bulletin & Review, 7,* 727-733.

Phye, G. D., & Pickering, S. J. (2006). *Working memory and education*. Academic Press.

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28*(3), 164-171

Redick, T. S., & Lindsey, D. R. (2013). Complex span and n-back measures of working memory: A meta-analysis. *Psychonomic Bulletin & Review, 20*(6), 1102-1113.

Roodenrys, S., & Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords? *Journal of Experimental Psychology: Learning, Memory & Cognition, 28,* 29-33.

Sagarra, N. (2017). Longitudinal effects of working memory on L2 grammar and reading abilities. *Second Language Research*, 1-23.

Salthouse, T. A., & Babcock, R.L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27*(5), 763-776.

Sanz, C., Lin, H. J., Lado, B., Stafford, C. A., & Bowden, H. W. (2014). One size fits all? Learning conditions and working memory capacity *in Ab initio* language development. *Applied Linguistics, 37*, 1-26.

Sasisekaran, J., Smith, A., Sadagopan, N., & Weber-Fox, C. (2010). Nonword repetition in children and adults: Effects on movement coordination. *Developmental Science*, *13*(3), 521-532.

Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 35*, 1089-1096.

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2014). A task is a task is a task: putting complex span, n-back, and other working memory indicators in psychometric context. *Frontiers in psychology*, *5*,1475.

Shahnazari, M. (2013). The development of a Persian reading span test for the measure of L1 Persian EFL learners' working memory capacity. *Applied Research on the English Language, 2*(2), 107-116.

Skehan, P. (2015). Working memory and second language performance: A commentary. In Z. E. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 189-204). Bristol: Multilingual Matters.

St Clair-Thompson, H. L. (2007). The influence of strategies on relationships between working memory and cognitive skills. *Memory*, *15*(4), 353-365.

Tagarelli, K. M., Mota, M. B. & Rebuscha, P. (2015). Working memory, learning conditions and the acquisition of L2 syntax. In Z. E. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second* *language acquisition and processing* (pp.224-247). Bristol: Multilingual Matters.

Tehan, G., Hendry, L., & Kocinski, D. (2001). Word length and phonological similarity effects in simple, complex, and delayed serial recall tasks: Implications for working memory. *Memory, 9*, 333-348.

Thorn, A. S., & Gathercole, S. E. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *The Quarterly Journal of Experimental Psychology: Section A, 52*(2), 303-324

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

Unsworth, N. (2010). On the division of working memory and long-term memory and their relation to intelligence: A latent variable approach. *Acta Psychologica, 134*, 16-28.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*(1), 104-132.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498-505.

Van den Noort, M., Bosch, P., Haverkort, M., & Hugdahl, K. (2008). A standard computerized version of the reading span test in different languages. *European Journal of Psychological Assessment, 24*(1), 35-42.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology Human Experimental Psychology, 49* (1), 51–79.

Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments & Computers, 35*, 550-564.

Wen, Z. E. (2014). Theorizing and measuring working memory in first and second language research. *Language Teaching, 47*(2), 173-190.

Wen, Z. E. (2015). Working memory in second language acquisition and processing: The phonological/executive model. In Z. E. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 41-62). Bristol: Multilingual Matters.

Wen, Z. E., Mota, M. B., & McNeill, A. (Eds.). (2015). *Working memory in second language acquisition and processing*. Bristol: Multilingual Matters.

Wright, C. (2015). Working memory and L2 development across the lifespan: A commentary. In Z. E. Wen, M. B. Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp.285-298). Bristol: Multilingual Matters.

Appendix A. Sample sets of the developed Rspan task

A three-span set

۱. اگر در ماه‌های آینده درآمد بیشتری داشته باشم، شاید امسال یک کامپیوتر جدید بخرم.

درست ☐ نادرست ☐

جیم (ج)

۲. متأسفانه، گل نرگس زیبای من بر اثر سهل‌انگاری خودم جلوی من چشمان پژمرد.

درست ☐ نادرست ☐

غین (غ)

۳. تازه وارد اتاق نشمین شده بودم که پسردایی دوستم سراسیمه از جای خود برخاست.

درست ☐ نادرست ☐

طا (ط)

A four-span set

۴. دوستان نزدیک و صمیمی خود را ارزیابی دقیق می‌کنم تا مطمئن بشوم که بهترین دوستانم هستند.

درست ☐ نادرست ☐

ف٫ (ف)

۵. امسال بااینکه روی لوله ای آب را پوشانیدم، از شدت سرما هم باز لوله ترکید

درست ☐ نادرست ☐

کاف (ک)

۶. در مراسم قرعه‌کشی، به‌ظاهر آرام نشسته بودم اما دلم و سرکه مثل سیر می‌جوشید.

درست ☐ نادرست ☐

پ٫ (پ)

۷. سرباز پرچم را به دست گرفته بود و باافتخار نشانه را به پیروزی آن می‌چرخاند.

درست ☐ نادرست ☐

نون (ن)

A five-span set

۸. بااینکه دوستم همه تلاش خود را کرد، بازهم معمول طبق به جلسه هفتگی ما نرسید

درست ☐ نادرست ☐

شین (ش)

۹. پزشک به محمد توصیه کرد که ماهی دو بار، خود سر با شامپوی گیاهی را بشوید.

درست ☐ نادرست ☐

لام (ل)

۱۰. بدون اینکه قصد بدی داشته باشم مادرم را با سخنان نسنجیده خود رنجاندم.

درست ☐ نادرست ☐

عین (ع)

۱۱. خاله زهرا اصرار می‌کرد و می‌گفت: « کمی استراحت کن، تو از راه دور آمده ای».

درست □ نادرست □

کاف (ک)

۱۲. گیاهان نور و دی‌اکسید کربن را جذب می‌کنند تا با آن برای غذا خود بسازند.

درست □ نادرست □

ه (ه)

Appendix B. Sample sets of the developed Ospan task

A three-span set

A three-span set

۱. (۷+۱) + ۵= ۹

درست □ نادرست □

عین (ع)

۲. (۵+۱) −۲= ۳

درست □ نادرست □

لام (ل)

۳. (۵×۱) −۱=۴

درست □ نادرست □

پ (پ)

A four-span set

۴. (۷×۴)+۲=۳۲

درست □ نادرست □

غین (غ)

۵. (۷+۱)+۲= ۹

درست □ نادرست □

ه (ه)

۶. (۹+۱) −۲=۳

درست □ نادرست □

جیم (ج)

۷. (۸×۳)+۲= ۳۴

درست □ نادرست □

شین (ش)

A five-span set

۸. (۶+۲)−۱=۳

درست □ نادرست □

شین (ش)

۹. (۲×۳)+۳=۹

درست □ نادرست □

لام (ل)

۱۰. (۴+۱)+۳=۶

درست □ نادرست □

عین (ع)

۱۱. (۴×۴)−۲=۷

درست □ نادرست □

کاف (ک)

۱۲. (۴+۲)+۱=۶

درست □ نادرست □

ه (ه)

Appendix C. Dspan task

Set of three-digit strings

۸۲۹

۱۳۲

۶۸۷

۳۵۶

Set of four-digit strings

۶۲۴۱

۲۳۵۹

۷۱۳۲

۷۳۹۲

Set of five-digit strings

۸۴۱۳۲

۶۲۱۴۳

۹۷۴۳۸

۸۵۲۹۳

Set of six-digit strings

۵۸۷۲۶۱

۶۳۲۱۴۷

۲۴۷۶۸۱

۴۲۹۷۳۵

Set of seven-digit string

۲۹۴۱۳۷۸

۱۲۸۳۹۴۸

۸۶۹۳۷۳۵

۶۲۹۷۸۶۵

Set of eight-digit strings

۶۵۱۴۷۲۷۹

۱۸۴۷۲۹۱۳

۴۲۷۸۵۹۲

۲۸۶۸۳۱۹۷

Set of nine-digit strings

۶۷۹۱۷۴۳۸۲

۷۴۶۲۳۱۹۵۸

۳۹۸۷۲۴۶۱۵

۲۳۹۸۷۴۲۶۵