The development of a Persian reading span test for the measure of L1 Persian EFL learners’ working memory capacity

Mohammadtaghi Shahnazari-Dorcheh
(Assistant Professor, University of Isfahan, Isfahan, Islamic Republic of Iran)
m_t_shah@yahoo.com

(Received: 01.09.2012, Accepted: 01.06.2013)

Abstract
This study describes working memory and developing and validating of an L1 Persian reading span test for the measurement of working memory of L1 Persian EFL learners. The test, which included 64 Persian sentences, was developed based on Daneman and Carpenter’s (1980) reading span test. The shortcomings of the test were identified and removed over three pilot studies on 74 participants. The final test was used in a study with 140 participants at three different proficiency levels. The results of an item analysis, as indicated by Cronbach’s Alpha, displayed an internal reliability of .844 and .790 for the RST processing and recall scores respectively. This suggests that the newly developed test is reliable enough and could be used to measure working memory capacity for future L2 studies. This study also provides a clear procedure for the development of a reading span test for speakers of other languages.

Keywords: Working memory capacity, reading span test, processing, storage

What is working memory?
Working memory (WM) can be defined as a cognitive workspace with a limited capacity pool of attentional resources for the temporary storage of information while performing higher order cognitive tasks such as reasoning, learning and comprehension (Baddeley & Hitch, 1974; Baddeley & Logie, 1999). Baddeley and his colleagues view WM as that which simultaneously maintains and processes the input it receives through different channels of communications (e.g., touch, long-term memory, sight, and hearing) (Baddeley, 1986, 1996, 2003, 2007; Baddeley & Hitch, 1974; Baddeley & Logie, 1999; Gathercole & Baddeley, 1993). A three-component model of WM was proposed by Baddeley and Hitch (1974). This model consists of a central executive and two “slave” components, the phonological loop and the visuo-spatial sketchpad. This model was in use until 2000, when Baddeley added a new component to it, the episodic buffer, to account for the studies on densely amnesiac patients with long-term memory deficits. This model, as shown in Figure 1, specifies a functional role of memory as well as an economical and coherent account of information on each memory component.

![Baddeley's Model](www.SID.ir)
Baddeley’s (2000) model of WM, revised to incorporate links with long-term memory (LTM) by way of both the subsystems and the newly proposed episodic buffer.

The most important component in this model is the central executive or supervisory attentional system, which is a limited capacity pool of general resources. According to N. Ellis, (2001), “It regulates information flow within WM, activates or inhibits the whole sequences of activities, and resolves potential conflicts between ongoing schema-controlled activities” (p., 33). The reading or listening span tests are usually used to measure central executive and give an index for WM.

The phonological loop is in charge of the temporary storage and processing of verbal information. It plays a role as a phonological store by holding phonological representations of auditory information for a brief period of time, and as an articulatory rehearsal system by enabling the reader to use inner speech to refresh the decaying representations in the phonological store (Baddeley, 2000, 2007; N. Ellis, 2001). Phonological loop is often measured by presenting spoken lists of words (word span), digits (digit span) or non-words (non-word span), and asking participants to recall the lists of words and/or digits in the order in which they are presented. The maximum number of items that the individual can correctly recall is considered to be their phonological memory score.

The visuo-spatial sketchpad is an interface between visual and spatial information received either through the senses or from long-term memory (Baddeley & Hitch, 1974, p., 854). It is also involved in generating visual images, temporarily maintaining them, and manipulating information with visual or spatial dimensions. Furthermore, it can be activated by spoken words by using long-term knowledge to convert the auditory presented words into visuo-spatial code (Baddeley, 2007; N. Ellis, 2001). To measure visual memory, Della Sala, Gray, Baddeley, Allaman & Wilson’s (1999) pattern span test is usually used by researchers. In this test, the individual is presented with 2 x 2 matrixes, with two of the cells filled. Then after 3 seconds, the individual is asked to indicate which cells were filled in the stimulus matrix, using an empty 2 x 2 matrix. The size of the matrix is increased by two cells every three trials, with half of the cells of each matrix being randomly filled. The individual’s pattern span is determined by the maximum number of the cells that the participant is able to recall correctly.

The Corsi Block task is typically used to measure spatial memory (Milner, 1971). In this test, the subject is presented with an array of nine cubes arranged at random locations on a board placed between the tester and the participant. The test starts with the tester initially tapping two of the blocks one after the other and then asking the subject to imitate the sequence. The sequence of taps gradually increases to a point at which performance breaks down.

The episodic buffer (Baddeley, 2000) is a limited capacity temporary storage system. According to Baddeley (2007), “It combines information from the loop, the sketchpad, long-term memory, or indeed from perceptual input into a coherent episode” (p., 148). Moreover, it plays a role in interfacing between WM and long-term memory through the central executive, interacting phonological loop and sketchpad. It is also proposed that retrieval from the episodic buffer is through conscious awareness. However, no method of measurement has
been proposed yet to assess the episodic buffer (Baddeley, 2007).

Rationale of the study
Since an important role for working memory has been found in the first language acquisition (e.g., Daneman, 1991; Daneman & Green, 1986; Waters & Caplan, 1996), research on the role of working memory is emerging as an area of concern for second language acquisition (e.g., Atkins & Baddeley, 1998; Miyake & Freidman, 1998; Robinson, 2002, 2005). Working memory is typically measured by a reading span test (RST) or listening span test in L1 or L2.

The Reading span tests were first introduced by Daneman & Carpenter (1980). They were used to measure and give an index for working memory capacity (WMC). In a reading span test (RST), participants are asked to read sets of sentences, report on the semantic acceptability of each sentence (processing assessment), and then recall the final word of each sentence when prompted (storage assessment). Since the introduction of the RST by Daneman and Carpenter (1980), many researchers have used either Daneman and Carpenter’s original RST or the modified versions of that in their studies (e.g., Alptekin & Erçetin, 2009; Chun & Payne, 2004; Daneman & Carpenter, 1980; Harrington & Sawyer, 1992; Lesser, 2007; Osaka & Osaka, 1992; Swanson, 1993; Walter, 2004). These studies measured WM either through an L1 RST (Chun & Payne, 2004; Lesser, 2007), an L2 RST (Alptekin & Erçetin, 2009), or both L1 and L2 RSTs (Harrington & Sawyer, 1992; Walter, 2004). As prior research indicated that WM is language independent (e.g., Miyake & Freidman, 1998; Osaka & Osaka, 1992; Osaka, Osaka & Groner, 1993), measuring WM in L1 was then became popular in cognitive psychology and studies in second language learning. This could also help to avoid conflating WM and L2 proficiency. However, while there may be considerable number of L1 RSTs for some languages, there are few L1 RSTs in some others. In Persian, there may be few reliable versions of RST, and if any, none of them has been published or accessible for the use in other L2 studies. This issue points to the need for the development of a RST in this language for the research with L1 Persian EFL learners. The present study focused on the process of development and validation of an L1 Persian RST for the use in second language learning studies. More specifically, this study describes the stages at which RST items were developed, piloted, revised, and finally employed in the research with L1 Persian participants.

Methodology
Subjects
74 L1 Persian EFL learners at three proficiency levels participated in three pilot studies. Then the newly developed test was administered to 140 L1 Persian EFL learners in an experimental study. They included both males and females, 16-35 years old, studying English as a foreign language in a private language school in Iran.

Material
A corpus of Persian sentences was constructed by an expert in the Persian language. The sentences contained general information, and lacked of any technical and scientific content. 64 sentences were selected from this corpus to form the RST. This test included 10 practice session sentences and 54 test sentences, all of which were in an active and affirmative form within a range of 13-16 words. Half of the sentences were constructed as ‘nonsense’ sentences. This was done by rearranging a few words in such a way that sentences were semantically anomalous (Chun & Payne, 2004; Harrington & Sawyer, Lesser, 2007,
Turner & Engle, 1989; Waters & Caplan, 1996). This was to make sure that the participants processed sentences for meaning without focusing only on the retention of recall items. This test was administered individually using a computer-based format. Because Persian sentences follow SOV syntax (the sentences initiate with a subject followed by an object and a verb respectively), each sentence ends in a verb, similar to the reading span tests in Japanese (Osaka & Osaka, 1992) and German (Osaka et al., 1993; Roehr & Ganem-Gutierrez, 2008). Each verb appeared only once in the test. Therefore, the final words in this test were 64 different verbs. The verbs in each set were not semantically related. The sentences in the test were arranged in three sets of 3, 4, 5, and 6 sentences. Half of the sentences in each set were nonsense.

Test procedure

After the initial form of the RST was developed, three pilot studies were administered to three groups of L1 Persian EFL learners. This was to identify the potential problems with the test. Then the newly developed test was used in an experimental study for the measurement of working memory capacity.

The test was in a PowerPoint format and was taken individually. It assessed two WM components, processing and storage (e.g., Chun & Payne, 2004; Daneman & Carpenter, 1980; Harrington & Sawyer, 1992; Lesser, 2007; Waters & Caplan, 1996). The participants had to read each sentence, judge whether or not it made sense and say their judgment aloud while their answer was recorded. This was the measure of WM processing. They also had to remember the last word of each sentence up to the end of the set until a visual prompt (three hash keys) along with a two-second auditory prompt appeared on the computer screen. The pilot study results suggested that these two simultaneous prompts could well put a clear boundary between the sets and help the participants not to miss the recall time. At this time, the participants had to recall the sentence-final words and say them out loud while their answers were recorded by the researcher. This was the measure of the WM storage component. To control the recency effect, the participants were required to recall the words in the order in which they appeared (Baddeley & Hitch, 1993; Waters & Caplan, 1996).

A test instruction guide followed by an oral explanation which included an example set of three sentences was given to the participants prior to the test. Then they were given a practice session consisting of 10 sentences in two sets of three and a set of four sentences. Then the test began with a set of 3 sentences, and as the test progressed, the number of sentences presented on each trial increased successively from three to six, with three trials being presented at each series length. The prompt slide transitions increased accordingly from 12 to 18 seconds based on the length of each set.

Pilot studies

To identify the potential problems with the RST, three pilot studies were administered to three different groups of L1 Persian EFL learners. In the first pilot, a group of 12 L2 participants completed the RST, followed by a retrospective report. In their retrospective report, they all reported that the transition time, 6 seconds, for each slide was too short to read through the sentence. They also wrote that a few sentences were too vague for them to determine whether they made sense or not. The results of an item analysis indicated that there were some poor test items in the test. They were identified as being too difficult. These results indicated
that the participants had performed poorly on both the processing and recall components. The sentences which the students had identified as too vague were located among the ones which had been identified as too difficult by the item analysis. In consultation with the Persian language expert, these sentences were either revised or replaced with new sentences. Then the transition time for each slide was increased to 8 seconds as well.

In the second pilot study, similar to the procedure in the first pilot study, a group of 18 L1 Persian EFL learners completed the revised RST followed by a retrospective report. In their retrospective report, they wrote that they had had sufficient time to read through the sentence on each slide and even rehearse the sentence final words (target). They also reported a case where two sentence final words were semantically related, and they had been able to make an association between them for better recall. The results of this study supported the participants’ claims. Their performance on the RST was better than the prior group’s. Most of them were also able to obtain the scores for the two semantically related targets. Since the participants’ rehearsing could have inflated the recall scores, the transition time for each slide was decreased to 7 seconds. Furthermore, one of the sentences including a semantically related word was replaced with a new sentence including a different target word. The new sentence was developed and proposed by the same Persian language expert.

In the third pilot study, the revised reading span test was administered to a group of 44 participants. They reported that the transition time for each slide was just enough to read the sentence through and decide whether it made sense or not. No one reported any opportunity for rehearsing the targets. Moreover, they believed that all sentences throughout the test had been neither too easy nor too difficult for them. The results of the item analysis also indicated that each item made a good contribution to the test. The internal reliability for this test, as indicated by Cronbach’s Alpha, was .834 & .737 for the RST recall and processing respectively.

Application of the newly developed reading span test in L2 research
The final test was used in an experimental study conducted by the researcher. This study investigated the relationship between WM and L2 reading ability on 140 L1 Persian EFL learners at three proficiency levels. The sentences in the test were arranged in three sets of 3, 4, 5, and 6 sentences. Half of the sentences in each set were nonsense. Each sentence appeared on screen for 7 seconds, when the computer transitioned to the next slide. After each set, a slide with 3 hash keys and a two-second auditory prompt appeared. This was to signal to the participants to recall the final word of each sentence in the set.

To score the test, one mark was allocated to the participants’ correct judgment and one mark to their correct recall of the test session items, with the total of 54 each. Thus, since there were 54 sentences across all the trial sets, the range of the participants’ processing and recall scores was between 0 and 54 for each participant. No marks were given to the practice session items. This was consistent with the scoring method in recent studies (e.g., Alptekin & Erçetin, 2009). Then a composite WM score was used as an indicator of the participants’ WMC (e.g., Lesser, 2007; Waters & Caplan, 1996). The composite WM was obtained by adding the processing and recall z-scores. This is a more reliable scoring method of WMC compared to the traditional span scores that
The development of a Persian quantifying the highest set size completed or the number of words in correct sets (Freidman & Miyake, 2005). An item analysis was conducted on this measure. The internal reliability for this measure, as indicated by Cronbach’s Alpha, was .844 and .790 for the RST processing and recall respectively. This suggests that the newly developed RST is reliable enough and could be used for the measurement of WM in future studies.

Conclusion
This study described developing an L1 Persian reading span test for the measurement of L1 Persian EFL learners’ working memory capacity. The Persian reading span test was developed, piloted and successfully used in a study with 140 participants. As the internal reliability of this measure was quite high, the test can be used to measure working memory capacity in future second language learning studies. The same procedure could also be used to develop a reading span test for speakers of other languages.

References


Appendix

The List of Reading Span Test Items

This list includes both the sense and nonsense Persian sentences as follows:

Practice Session

گاه فردى تصميم به انجام كاری مي گيرد كه وقت قدرت و توانايی او است.
احتمالا ناتوان ترين افراد بشر كسي است كه تنوان كسي با ديدگي روست شود.
به نوشته‌هاي هنری كه تعييم دفینا از روي آن اجرا مي شود نمايشنامه مي گوند.
بر عهده هر انسان در آين دنيا است كه استعدادهای را خدا براي شناسد.
خوب است كه ما در چك ووزش خاص برای رشد استعدادهای خود ماهر شويم.
در گرفتاري ها به تو امید بسته و در نا امیتها تو پنجه را خوشی باقتام.
اگر من تجربه من چنین حسن سال را لانه ديدگي را راه بر مي گريم.
من حاضر بدم چنین مدرك علمي بالاتر با هميه مشکلات روبرو بجنگ.
قرآن كتاب اساسي ما مي گويد ارزشمندترین مردم دوست داشته باشند اما انسان.
گياهان نور و دی اكسيد کربن را جذب مي كنند تا آن یک حواشم بسازند.

Test Session

اگر در آمد بهتري در ماه هاي آينده داشته باشم، شاید مسائل یک كامپیوتری که روز پچم.
گل ترگس زيبرپ مين بر اثر سهولت انگاری خودم جلوی مي خوند.
تازه وارده آن مقاومت به تصمیم را ثابت كرديم از جا خوید برخاست.
تمي ايران لانه زیادی برای پرورش گردن، اما حريق مناسبات دويره به باخت.
با شيلنگ آب زيابي زمين باشيدم به اين كه کم اميز گر و خاک بخوابد.
من با آتشان شدن به فلسه بسياري از افزاید قدمي خود را به دور انداختم.
من در كودكی از تاريكي زياد مي ميژرسيم هنوز هم گاهي از تاريكي مرا مي ترسانم.
زنبور زيبرپ از نفت خوراک ديدم كه با نامنی مرد چه جه تماز تعهد گلا را مي نوشيد.
پدر همينه به ما مي گفت: با بعضي امراض ویمارها یا دارد آنها هملي.
ابوعلی سينا حکيم معروف سرزمين علم و تمدن وحشت فرهنگسته كه در آسمان علم جهان مي درخش.
بعضي از مارا به مي توانند زهر خود را تا دو متر طرف به دشمن پاشند.
اگر ماندي سال را گذشته انسان هيما بازندگي کم باشد دچار کم امي مي شويم.
دکتر دست شکسته مي را گرفت و با یک پارچه به آن گرميم را اوپخت.
هوشنگ هنوز مانند قديم یا روز شير بست گاو را با دست راس مي دوشي.
پدرم همينه به مي گفت به دليل هر مارد اينگونه از پرو یا افتاده شده ام.

www.SID.ir
اگر بازیابی وزن ناسازی مهربانی، مهربانی اشتهاره می کرد، کسی این غلبه زیبا را نمی چید.

ای کاش چه در کودکی در زمان تابیه محبوب، در آغوش پدر مه و واقع مادر خود آرام بگیرد.

دوستان خاص و ا Tâm خود را بار دیگر ارزیابی دیق می کنند و آنها را با میکرومایی جدید می سنجم.

با اینکه روی لوله آب را پوشانده، باز هم زمستان امسال لوله ترک.

در مهمانی، به ظاهر آرام نشسته بود معده دلم و سرکه مثل سارم می چاشید.

حسن پدرم را به دست گرفته بود و نشانه را به برنزه آن می خردانید.

اشتباه لفظی و غیرقابل پیش بینی در کلاس باعث شدن سال گذشته همه همه ای به خندها.

پس لوله بخاری را کشید و با این کار از لوله در جای خود رفت.

حمیده یکی از جواهرات خود را پارسال کم کرد بود، پشت کمد لباس هایی یافت.

مادر با لبخند گفت، بسیار خوب امر ماه که یکی از این مادر از باش خود ترک.

هنر بسیاری از دختران خرسان روستایی با زحمت و مشقت زیاد هزار سال یک فرش دستی می بافند.

در کتابخانه آن شهر کتابخانه بسیاری بودند که احساس زیبایی در آرایش سوزانده.

پیرمرده که نرسیده در لازم شد به طول و زمان تابیه تا باسی از نسبت و فشار سنگهای نیلی را در آرامی.

این شانه را رشد فکری است که با تیرهای آینه هایی به سو برده های ای گرفتم تا درست بپذیریم.

فرهاد با افتخار گفت که از پارسال تاکنون دو هزار مترزمین کشاورزی خریده.

دیشب در مارا انتش بود، به همین دلیل پشتو نداشتن، نصف مرا گزیندند.

پرستار مهربان و خوشرو، هر روز بصورت اول وقت قطره چشم پدر را می چکاند.

بیشتر مهربان و فراموش کرده، هر روز مریم به قصه سوزن و مهربان خود دو مردم را می دوانند.

بعضی که از دوست یک است که با تاریخ و همراه یک دوست دوستی دوستی از راه آسمانی.

به هر دو دوست دوستی می گوید، تا بتوانند با یکدیگر از راه آسمانی.

درک‌بیشتر که از دوست دوستی و همراه به دوست دوستی از راه آسمانی.

با اینکه راحتی همه لاله خون را کرد، باز هم سرخی به ما نرسید.

دکتر به محمد گفت به خوبی باش، به خوبی باش، به خوبی باش.

پس از اینکه راحتی همه لاله خون را کرد، باز هم سرخی به ما نرسید.

خلاص زدها برای اسیر می کرد، کمی استراحت کن، تا زاده دور آمدید.

پیک روی تفاویل نامه با نامی نامه این است که تمام حکایات برای پایگاه را فیلم نامه در می نویسند.

زمستان که می دهد، چند تا سحر یکپزیرگ باربیاکی شناخته می خوانند.

از پنجره انتظار او را در حیات خانه‌ای می دیدم که آب قدم می زد و بامگی را می داد.

پیش از آن که بار دیگر تحمل کنند، کسانی می توانند باشد که خوب برکلمه بیرتسلیت دارند.

اخترا فیزیک محاسبه از علم مهیج یک است که به مطالعه اجرام آسمانی دور، و بررسی.

من برای یک پست الکترونیک در آزمایش بي روح و باز شده ما هر روز مخفی قابل توجهی راه بپیمایم.

به محض اینکه آب چش در ریان رخت، ریان به دسته بلندی شکست.

منهج تمام سمشتی به طرف گذاشته و طرف پایین گذا شده از لاس‌افتد.

دریوز هر چه تلاش کردی نتوانستی خوب توب سکتونی را به سردر بی‌پایان.

هرچه بی‌پایین مریم می زدیم، به ما زنگ می کرد و هم دنبال‌های را به ما می می‌رسید.

همیشه شوق دیدن کشوری که خواهد اوی شکفته در رگ من و پی دن می آورند.

گره کوچکی خانگی ما، گاهی با همان پنجاهها مرا به زیبایی و شکنندش صورت می خاوشد.

رنگین گیلاس پس از بارش سیل اسای باران در آسمان ظاهر می گردد و بعد از انکی ناپدید می شود.
مردم محله‌ی ما با وجود بدن کوچک، حاج مرتضی را پهلوان مرتضی می‌نامند.
هنوز صداي سرود ملی بلند نشده بود كه پرچم را سربازان در میدان حاضر برافراشتند.
من براساس عادت گذشته، هنگام غذا به خوردن روي نمک آن مي‌پاشم.