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 (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, Allamano & 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
quantify 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**

- اگر فردی تصمیم به انجام کاری می‌گیرد که فوق فقرت و توانایی او است.
- احتمالاً ناتوان اثری از افراد بشر کمی است که تنواند کمی با دیگری دوست شود.
- بی نوشته‌های هنری که نمایش دیانتی از روز آن اجرا می‌شود نمایش‌املی می‌گویند.
- بر عهده‌ی هر انسانی در این دنیا است که استعدادهای را خدمتادی پاسند.
- خوب است که ما در یک ورودی خاص برای رشد استعدادهای خود ماهر شویم.
- در گرفتن‌های بی‌هیچ بستگی و در ناامنی‌ها تو پناه‌گاه را خویش‌یافته.
- اگر من تجربه‌ی این چند سال را داشتم شاید دیگری را راه بی‌می‌گزید.
- من خاضور برای بدست آوردن مدرک علمی بالاتر با همه مشکلات روبرو بجنگم.
- قرآن کتاب آسمانی ما می‌گوید ارزشمندترین مردم نزد خدا با توانائی‌ای است.
- گیاهان نور و دی اکسیدکربن را جذب می‌کنند، تا آن برای غذا بسازند.

**Test Session**

- اگر در آمد بپردازید در ماه‌های آینده داشته باشیم، شاید امسال یک کامپیوترکریفی به روز بخورم.
- گل نرگس زیبایی مانند شهپر اگر خودم جلوی می‌دهم بپردازد.
- تازه‌اصل سه‌باره‌ی اینکه به طریقی دوست دوستی از خود به‌خواست.
- تسمای ایرانی در زندگی باید پیروی کرده، اما هر کسانی که سمادها دوباره بپذیرد.
- با شش‌انگشتی یک ذهنی به‌پرسیند باید این که امید خدا و خاک‌یابان.
- من با آشنا شدن با فلسفه بسیاری از افکار قدیمی خود را به دوی انداختم.
- من در کودکی از تاریکی زیاد می‌ترسیدم هنوز هم گاه‌گاهی تاریکی مرا می‌ترسد.
- زنبور زیبایی در نازه‌ی این کورپارک دیدم که با طراحی هر چه تنام پر شهید گل‌ها را می‌نوشید.
- بدر می‌همیشه به ما می‌گفت: با بعضی امراض ویمارها باید این امر را ساخت.
- ای توانستی حکمی معرفی سرمایه علم و تمدن و هنر، قرن‌هاست که در آسمان علم جهان می‌درخشد.
- بعضی از ما را مشکل‌هایی زیاد می‌توانند به‌طور ملی به دشمن پا ناپاید.
- اگر مانند سالهای گذشته‌ی امکان هم بازندگی کم باشد داچ و می‌شوم.
- دکتر دست شکسته می‌گفت و با یک پارچه به آن گردید و را ایستاد.
- هوش و توان مانند هنر، در روز سیر و بست گاو را با دست راس می‌دوشند.
- پدرم همیشه می‌گفت به دلیل مرگ مادرت اینگونه از پرو با افتاده شده‌ام.
اگر بازیان وظیفه شناس مهریان، مهربانانه اشاره می‌کردند، کسی این گل‌های زیبا را نمی‌چید.

ای کاش هر کودک را در زبان شایسته محبت، در، دوستی آرام و واقعیت مادر خود آرام بگیرد.

دوستان خاص و قمز خود را در دیگر ارزیابی دقیق می‌کنند و آنها را با معنی‌های جدید می‌سنجم.

با آنکه روی اولویت آب را پوشاندم، باز هم زمستان امکان لوله ترکیب.

در مهمانی، به‌طور آرام مشتیه بود اما دلم و سرکه مثل بی‌سیر می‌چشید.

حسن پرچم را به دست گرفته بود و نشانه‌ای از پرچم را به آن می‌چرخاند.

این شانه در فکری است که با آن‌وی عیب‌هایمان را به عنوان هیجان‌آوری از دوست‌پذیریم.

فرهاد با افتخار گفت که از پارسال تاکنون دو هزار مترسیماً کشاورزی خردسال.

دیشب در اتاق نشسته بود به همین دلیل پخته‌ها بنده تا صحرا گرفته بود.

پرسرامه مهریان و خوشبخت، هر زمانی سعی بر اول وقت قطعه بزنم بر من را چکاند.

بسیاری از اشخاص رسمی یا از ابعادی هستند که برای امضا یک پرونده ساده به‌طور ساده کار می‌کنند.

ما اینکه راهنده هم نشان خود را کرد، باید به طریقی از سر کم‌مدت قرنطینه کنیم.

پس امشب در روز گذشته در اتاق، به این دلیل پرگذشت و پرگذشت.

هر کسی که خوب نویسند، دوستی می‌توانند باشند که خوب بر علم منابع حرام ان دست به هر کاری می‌زنند.

فرهاد با افتخار گفت که از پارسال تاکنون دو هزار مترسیماً کشاورزی خردسال.

دیشب در اتاق نشسته بود به همین دلیل پخته‌ها بنده تا صحرا گرفته بود.

پرسرامه مهریان و خوشبخت، هر زمانی سعی بر اول وقت قطعه بزنم بر من را چکاند.

بسیاری از اشخاص رسمی یا از ابعادی هستند که برای امضا یک پرونده ساده به‌طور ساده کار می‌کنند.

ما اینکه راهنده هم نشان خود را کرد، باید به طریقی از سر کم‌مدت قرنطینه کنیم.

پس امشب در روز گذشته در اتاق، به این دلیل پرگذشت و پرگذشت.

هر کسی که خوب نویسند، دوستی می‌توانند باشند که خوب بر علم منابع حرام ان دست به هر کاری می‌زنند.

هر کسی که خوب نویسند، دوستی می‌توانند باشند که خوب بر علم منابع حرام ان دست به هر کاری می‌زنند.

هر کسی که خوب نویسند، دوستی می‌توانند باشند که خوب بر علم منابع حرام ان دست به هر کاری می‌زنند.

هر کسی که خوب نویسند، دوستی می‌توانند باشند که خوب بر علم منابع حرام ان دست به هر کاری می‌زنند.

هر کسی که خوب نویسند، دوستی می‌توانند باشند که خوب بر علم منابع حرام ان دست به هر کاری می‌زنند.

هر کسی که خوب نویسند، دوستی می‌توانند باشند که خوب بر علم منابع حرام ان دست به هر کاری می‌زنند.

هر کسی که خوب نویسند، دوستی می‌توانند باشند که خوب بر علم منابع حرام ان دست به هر کاری می‌زنند.

هر کسی که خوب نویسند، دوستی می‌توانند باشند که خوب بر علم منابع حرام ان دست به هر کاری می‌زنند.

هر کسی که خوب نویسند، دوستی می‌توانند باشند که خوب بر علم منابع حرام ان دست به هر کاری می‌زنند.

هر کسی که خوب نویسند، دوستی می‌توانند باشند که خوب بر علم منابع حرام ان دست به هر کاری می‌زنند.

هر کسی که خوب نویسند، دوستی می‌توانند باشند که خوب بر علم منابع حرام ان دست به هر کاری می‌زنند.

هر کسی که خوب نویسند، دوستی می‌توانند باشند که خوب بر علم منابع حرام ان دست به هر کاری می‌زنند.

هر کسی که خوب نویسند، دوستی می‌توانند باشند که خوب بر علم منابع حرام ان دست به هر کاری می‌زنند.

هر کسی که خوب نویسند، دوستی می‌توانند باشند که خوب بر علم منابع حرام ان دست به هر کاری می‌زنند.

هر کسی که خوب نویسند، دوستی می‌توانند باشند که خوب بر علم منابع حرام ان دست به هر کاری می‌زنند.

هر کسی که خوب نویسند، دوستی می‌توانند باشند که خوب بر علم منابع حرام ان دست به هر کاری می‌زنند.
مردم محله ی ما با وجود بدن کوچک، حاج مرتضی را پهلوان مرتضی می نامیدند.
هنوز صدای سرود ملی بلند نشده بود که پرچم را سربازان در میدان حاضر پرافراشتند.
من براساس عادت گذشته، هنگام غذا به خوردن روی نمک آن می‌پاشم.