English Language Teaching; Vol. 5, No. 11; 2012
ISSN 1916-4742 E-ISSN 1916-4750
Published by Canadian Center of Science and Education

Developing a Non-language Related Span Test for the Use in Language-specific and Cross-language Studies

Mohammadtaghi Shahnazari-Dorcheh[1] & Saeed Roshan[2]

[1] Department of English Language and Literature, Faculty of Foreign Languages, University of Isfahan, Isfahan, Iran

[2] Department of Applied Language Studies and Linguistics, University of Auckland, Auckland, New Zealand

Correspondence: Mohammadtaghi Shahnazari-Dorcheh, Department of English Language and Literature, Faculty of Foreign Languages, University of Isfahan, Hezar Jerib Street, Isfahan, Iran, Tel: 98-793-2107. E-mail: m_t_shah@yahoo.com

Received: July 25, 2012 Accepted: August 21, 2012 Online Published: September 12, 2012

doi:10.5539/elt.v5n11p104 URL: http://dx.doi.org/10.5539/elt.v5n11p104

Abstract

Due to the lack of span test for the use in language-specific and cross-language studies, this study provides L1 and L2 researchers with a reliable language-independent span test (math span test) for the measurement of working memory capacity. It also describes the development, validation, and scoring method of this test. This test included 70 simple math problems, and was developed based on Salthouse and Babcock's (1991) and Robert and Gibson's (2002) math span tests. The shortcomings of the test were identified and removed over five pilot studies on 48 participants. The final test was used in an experimental study with a group of L1 Persian EFL learners. Results of an item analysis, as indicated by Cronbach's Alpha, indicated an internal reliability of .850 and .863 for the math span test processing and recall respectively. This suggests that the newly developed test is reliable enough and could be used to measure working memory capacity in L1 and L2 studies. This study also provides a clear procedure for the development and scoring of a math span test for the use in L1 and L2 studies.

Keywords: working memory capacity, math span test, processing, storage, composite score

1. Introduction

Working memory is a cognitive workspace (e.g., Baddeley & Hitch, 1974; Baddeley, 2007) with a limited pool of attentional resources for temporary storage and processing information while performing higher order cognitive tasks such as comprehension, learning and reasoning (Baddeley & Logie, 1999). A good body of research suggests that working memory plays a very important role in the acquisition of L1 (e.g., Daneman, 1991; Daneman & Carpenter, 1980; Daneman & Green, 1986; Waters & Caplan, 1996) and L2 (e.g., Ando, Fukunaga, Kurahachi, Stuto, Nakano, & Kage, 1992; Atkins & Baddeley, 1998; Mackey, Philp, Egi, Fujii, & Tatsumi, 2002; Mackey, Adams, Stafford, & Winke, 2010; Shahnazari-Dorcheh & Adams, in press). Working memory is typically measured by a language-related or a non-language-related span task. In the language-related span task such as a reading or listening span test, the participants read or listen to a set of unrelated sentences and judge whether they make sense or are nonsense (processing assessment), and then try to recall the final word of each sentence at the end of the set (storage assessment). In the non-language related span task such as an operation span test, the participants view some simple arithmetic equations and verify whether the stated solution is correct or incorrect (processing assessment), and at the end of the set, they have to recall the stated solutions from each equation in the set (storage assessment). In both language and non-language related span tasks, an index of working memory is calculated with the composite score of these two assessments (e.g., Freidman & Miyake, 2004; Waters & Caplan, 1996).

However, the language-related span tasks are language-specific and differ from one language into the other. Then the prior L1 and L2 reading (e.g., Daneman & Carpenter, 1980; Harrington & Sawyer, 1992) or listening (e.g., Mackey, Adams, Stafford, & Winke, 2010; Mackey & Sachs, 2011) span tests can be used by the researchers for the measurement of working memory capacity of speakers of just those languages. Furthermore, if working memory is measured through L2 reading or listening span test, a reliable index for working memory capacity may not be obtained because memory performance may be confounded with L2 proficiency (e.g., Juffs & Harrington, 2011). It appears that a reliable non-language related span task such as a math span test is a requirement. This math span test can be taken in English as an international language provided that those L2 participants to whom the test is administered are familiar enough with English digits. This means that they should be at least at the beginning level. The test can be taken in their L1 or other languages if the English digits are converted into that specific language. To meet this requirement, this study was designed to describe the development and validation of an English math span test for the use with L1 and L2 researchers in measuring working memory capacity. Using such a test may help these researchers to control task-specific factors available in reading or listening span test and provide a more reliable index of working memory capacity. It may also yield more reliable results once it is used to measure working memory capacity in cross-language studies.

2. Methodology

This type of test was first developed by Turner and Engle (1989) to measure working memory capacity. They called it an operation span test. In this test, a set of simple arithmetic equations such as $(6/2) + 5 = 8$, $(3 \times 4) - 5 = 7$, and $(4/2) + 2 = 6$ are presented to the participants. For each equation, the participants' task is to verify whether the stated solution is correct or incorrect (processing assessment), and at the end of the set, they have to recall the stated solutions from each equation in the set (here, 8, 7, and 6) (storage assessment). The number of arithmetic problems on each set is successively increased from one to seven, with three sets being presented at each series length. The total number of stated solutions recalled from the perfectly recalled set is regarded as the participant's math span. This test has been used as a measure of working memory capacity in several prior studies (e.g., Daneman & Merikle, 1996; Mizera, 2006; Hambrick & Engle, 2002; Robert & Gibson, 2002; Salthouse & Babcock, 1991; Turner & Engle, 1989). Further support for the use of operation span test, as a reliable measure of working memory capacity is provided by a recent study in cognitive psychology (Sanchez, Wiley, Miura, Golflesh, Rioks, Jensen & Conway, 2010). This study suggests that an operation span test can be used to effectively assess working memory capacity and could be a predictor of a fluid intelligence test like RAPM (Raven's Advanced Progressive Material).

The math span test for the current study was based on Salthouse and Babcock's (1991) and Robert and Gibson's (2002) math span tests. This test was comprised of some simple arithmetic problems in the form of $X + Y = ?$ or $X – Y =?$ type. X and Y can be single digit numbers between 1 and 9, and none of the answers to the problems were negative. There were no identical (repetitive) arithmetic problems across the test or any repeated target digits for two consecutive problems. However, whereas Salthouse and Babcock (1991) provided three possible answers and asked their subjects to check off one, the participants here had to take the test individually and provide the answers orally, like in Roberts and Gibson's version of this task (2002), and their production was recorded by the researcher. This format of the math span test was used to make sure that the participants' correct answers would not be subjected to guessing as well as to control for the recency effect. Furthermore, it would ensure that the participants had recalled the target digits at the end of the set and not earlier during the processing time.

Thus, in this test, the participants viewed simple addition and subtraction problems (i.e., $4 + 2 =?$ or $9 – 6 =?$) on a computer screen. Each problem appeared on the screen for 2.5 seconds. The participants were required to state the answer to the problem aloud immediately (processing) and remember the second digit in each problem for later recall (storage). To control the speed of processing, and consequently possible rehearsal of the targets, each participant was required to view each math problem, and does it in his or her mind within a very constrained given time.

This test was developed by the researcher and piloted with different groups of L2 participants (L1 Persian EFL learners) (overall 48 participants) at three levels over five pilots. On each occasion, a different combination of participants completed the test, followed by a retrospective report. Based on their reports and results on each occasion, the shortcomings of the test, which were mostly related to the slide transition times, were removed until no further shortcomings were reported by the participants.

During the first pilot, the test was administered to a group of 10 L2 participants. The slide transition for each math problem was set on 5 seconds. However, the participants reported that they had some extra time to rehearse the targets (second digit at each math problem). Furthermore, they claimed that there had been two consecutive math problems within one set with the same target digits. To remove this problem, the positions of the digits were reversed in one of the math problems. Furthermore, the slide transition was decreased to 4 seconds. Then the revised test was piloted with another group of participants during the second pilot to see whether it worked well or not.

During the second pilot, the test was administered to a group of 9 L2 participants. They reported that they had no problems with the test. However, they said that the slide transition for each math problem had been too long, so they had time to rehearse the targets. The results of study also indicated that the participants' scores were very high. Then it was concluded that it might be the extra time that had led to inflated scores here. This problem was removed by decreasing the slide transition for each math problem to 3 seconds. Then the revised test was piloted with a new group of participants to see whether there was still extra time for the participants to rehearse the targets or not.

For the third pilot, the test was administered to a group of 11 L2 participants. The results of the study indicated two ranges of scores, with some participants scoring quite highly and others quite low. Participants with low scores reported that the updated slide transition times had been just enough for them, while those with high scores said that they had had a little extra time for rehearsing the targets. The range of scores was wider than before.

During the fourth pilot, the slide transition was decreased to 2 seconds and the revised test was administered to a group of 8 L2 participants. However, the participants all reported that the slide transition had been too fast for them to do the math problem. Thus, they had had to skip some math problems and focus just on the targets for better recall. The results of the study also indicated that the participants had not obtained consistent scores for the processing and recall components, like those in the prior pilot studies.

During the fifth pilot, the slide transition was set to 2.5 seconds and the test was administered to a group of 10 L2 participants. The participants' scores demonstrated the widest spread of the pilot tests (31-60 & 16-58 for processing and recall capacities respectively). They also reported that they had had sufficient time to process each math problem but had had no more time to rehearse the second digits. Thus, the duration of 2.5 seconds was established as an appropriate slide transition for the final test. The results of this pilot were consistent with the findings of an experimental study where this test was used, which showed a wide spread of participant scores (33-60 & 13-59 for processing and recall capacities respectively). A satisfactory internal reliability, as indicated by Cronbach's Alpha, was found for this measure in the experimental study. The reliability was .850 and .863 for the MST processing and recall respectively.

The final version of the test was comprised of 60 simple addition and subtraction problems, 30 each, distributed equally in 3 sets of 2, 3, 4, 5, and 6 math problems. There was also a practice test including 10 math problems at the beginning of the test session. This was to familiarize the L2 participants with the test procedure. The participants were told that they would receive no points for the practice test items. Following this, they went through increasingly longer sets of math problems. At the end of each set, a prompt (three hash keys) appeared on the computer screen. This was to signal to the participants to recall the target digits aloud while their production was recorded. To control for the recency effect, the participants were instructed not to say the last target digit first.

To score the participants' math span test, each participant's score for the processing and storage components of working memory was calculated. The processing score was the total number of correct answers given to the math problems. The storage score included the total number of target digits recalled correctly across the test (Friedman & Miyake, 2005). Thus, since there were 60 math problems in this test, and one mark was allocated to each correct answer, the range of each participant's processing and recall score was between 0 and 60. A composite working memory score was obtained (Turner & Engle, 1989; Waters & Caplan, 1996). The composite working memory score was calculated by adding up the z-scores of the working memory components. This was an index for each participant's working memory capacity.

The final test was used in an experimental study conducted by the researcher. This study investigated the relationship between working memory and L2 reading ability on 140 L1 Persian EFL learners at beginning (56), intermediate(43) and advanced (41) levels. The final test included 70 simple math problems, 10 practice session math problems and 60 test session math problems. This test was administered individually using a computer-based format. Each math problem appeared on screen for 2.5 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 target digit of each math problem in the set.

The test was in PowerPoint format and was taken individually. It assessed two working memory components, processing and storage (e.g., Chun & Payne, 2004; Daneman & Carpenter, 1980; Harrington & Sawyer, 1992; Lesser, 2007; Waters & Caplan, 1996). The participants had to view each math problem, calculate the simple addition or subtraction problem and say their answer aloud while their answer was recorded. This was the measure of working memory processing. They also had to remember the second digit of each math problem 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 second digits and say them out loud while their answers were recorded by the researcher. This was the measure of the working memory storage component. To control the recency effect, the participants were required to recall the digits 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 math problems was given to the participants prior to the test. Then they were given a practice session consisting of 10 math problems in two sets of three and one set of four math problems. Then the test began with a set of 2 math problems, and as the test progressed, the number of math problems presented on each trial increased successively from two to six, with three trials being presented at each series length. The prompt slide transitions increased accordingly from 4 to 12 seconds based on the length of each set.

To score the test, one mark was allocated to the participants' correct answer and one mark to their correct recall of the test session items, with the total of 60 each. Thus, since there were 60 math problems across all the trial sets, the range of the participants' processing and recall scores was between 0 and 60 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 working memory score was used as an indicator of the participants' working memory capacity (e.g., Lesser, 2007; Waters & Caplan, 1996). The composite working memory was obtained by adding the processing and recall z-scores. This is a more reliable scoring method of working memory capacity 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 .850 and .863 for the math span test processing and recall respectively. This suggests that the newly developed math span test is reliable enough and could be used for the measurement of working memory in future studies.

3. General Discussion and Conclusion

This study was designed to develop a math span test for the measurement of L1 and L2 learners' working memory capacity. The math span test was developed and piloted on five groups of L1 Persian EFL learners. The potential problems in the test were identified and removed. Then the test was successfully used in an experimental study with 140 participants. The math problems in this test included digits of 1-9. These digits appeared on the computer screen in English. This suggests that the test could be used for the speakers of other languages provided that they are familiar with English digits. As the internal reliability of this measure was quite high, the test can be used to measure working memory capacity in future L1or L2 studies. The same procedure could also be used to develop and score further math span tests for the measurement of working memory capacity.

Following Friedman and Miyake (2005), this study employed the total number of targets (second digit in each math problem) recalled as it was a more reliable method for scoring the storage capacity of working memory. In this method, the sum of the correctly recalled elements from all sets, regardless of whether the elements in each set are all recalled or not, is counted for the storage capacity score. In Conway, Kane, Bunting, Hambrick, Wilhelm and Engle's (2005) term, this is "partial-credit scoring" which is used to obtain recall scores for individuals whose processing scores meet the requirement (85% or above). In the current research, one point was allocated to each recalled item. This method of scoring is supported by the most recent research (Juffs & Harrington, 2011) where it is argued to provide "a finer discrimination between individuals and be more reliable" (p. 144). To control any recency effect (Baddeley & Hitch, 1993), no points were given to the targets in math problems appearing in final positions in sets if they were recalled first.

The same method was also used in the scoring of processing capacity in this study. The total number of correct answers to the math problems, regardless of whether the target in each of them had been recalled correctly or not, was regarded as the processing capacity score. The advantage of this scoring procedure for processing and storage capacities, other than being more reliable (Friedman & Miyake, 2005), is that it may involve a wider range of scores, better discrimination between high and low capacity participants as well as all correct responses in the total scores of storage and processing respectively.

Overall, these findings imply that a math span test, as a complex span task, is a reliable cognitive task tapping and measuring both the processing and storage components of working memory. This adds further support to prior studies where working memory was operationalized as the performance on the complex span tests such as the reading span test, operation span test or counting span test both in the L1 (e.g., Daneman & Carpenter, 1980; Turner & Engle, 1989; Waters & Caplan, 1996) and the L2 (e.g., Alptekin & Erçetin, 2010; Harrington & Sawyer, 1992; Lesser, 2007; Walter, 2004). Finally, the results of this study suggest that the newly developed test is reliable enough to be used in language-specific and cross language studies for the measurement of working memoy capacity. As mentioned before, since the digits are in English, the participants in these studies need to be able to read and calculate the digits in English.

References

Alptekin C., & Erçetin G. (2009). Assessing the relationship of working memory to L2 reading: Does the nature of the comprehension process and reading span task make a difference? *System, 37*, 627-639. http://dx.doi.org/10.1016/j.system.2009.09.007

Alptekin C., & Erçetin G. (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *Journal of Research in Reading, 33*(2), 206-219. http://dx.doi.org/10.1111/j.1467-9817.2009.01412.x

Ando, J., Fukunaga, N., Kurahashi, J., Stuto, T., Nakano, T., & Kage, M. (1992). A comparative study on two EFL teaching methods: The communicative and the grammatical approach. *Japanese Journal of Educational Psychology*, *40*, 247-256.

Atkins, W. B., & Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics*, *19*, 537-552. http://dx.doi.org/10.1017/S0142716400010353

Baddeley, A. D. (2007). *Working Memory, Thought and Action.* Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory revised. *American Psychologist*, November (2001), 851-864.

Baddeley, A. D., & Hitch, G. (1993). The recency effect: Implicit learning with explicit retrieval? *Memory and Cognition, 21*(2), 146-155.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple component model. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive central* (pp. 28-61). Cambridge: Cambridge University Press.

Chun, D. M., & Payne, J. S. (2004). What makes students click: Working memory and look-up behavior. *System*, *32*, 481-503. http://dx.doi.org/10.1016/j.system.2004.09.008

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

Daneman, M. (1991). Working memory as a predictor of verbal fluency. *Journal of Psycholinguistic Research, 20*, 445-464. http://dx.doi.org/10.1007/BF01067637

Daneman, M., & Carpenter, P. A. (1980). Individual Differences in Working Memory and Reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

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

Daneman, M., & Merikle, P. M. (1996). Working memory and Language comprehension: A Meta-analysis. *Psycholinguistic Bulletin and Review, 3*, 422- 433.

Davis, F. B. (1994). Fundamental factors of comprehension in reading. *Psychometrica, 9*, 185-197.

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

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

Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is – power hypothesis. *Cognitive Psychology, 44*, 339-387. http://dx.doi.org/10.1006/cogp.2001.0769

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*, 25-38.

Juffs, A., & Harrington, M. W. (2011). Aspects of working memory in L2 learning. *Language Teaching: Reviews* *and Studies*, *42*(2), 137-166.

Lesser, J. M. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning*, *57*(2), 229-270.

Mackey, A., Adams, R., Stafford, C., & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning*, *60*(3), 501-533.

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

Mackey, A., & Sachs, R. (2011). Older learners in SLA research: A first look at working memory, feedback, and L2 development. *Language Learning*, 1-37.

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

Roberts, R., & Gibson E. (2002). Individual differences in sentence memory. *Journal of Psycholinguistic Research, 31*, 573-598. http://dx.doi.org/10.1023/A:1021213004302

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*, 763-776. http://dx.doi.org/10.1037/0012-1649.27.5.763

Sanchez, C. A., Wiley, J., Miura, T. K., Colflesh, G. J. H., Ricks, T. R., Jensen, M. S., & Conway, A. R. A. (2010). Assessing working memory capacity in anon-native language. *Learning and Individual Differences*, *20*, 488-493.

Shahnazari, M., & Adams, R. (forthcoming). The relationship between modified output and working memory capacity. *Journal of Asian TEFL*.

Turner, M., & Engle, R.W. (1989). Is working memory task dependent? *Journal of Memory and Language, 28*, 127-154. http://dx.doi.org/10.1016/0749-596X(89)90040-5

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

Appendix: Math Span Test items

Simple math span test problems are as follow:

Practice Session

Set One
7 + 9 = ?
8 – 1 = ?

Set Two
9 + 4 = ?
5 – 2 = ?

Set Three
3 + 7 = ?
6 – 1 = ?
3 + 9 = ?

Set Four
4 – 2 = ?
7 + 8 = ?
9 – 3 = ?

Test session

Set One
2 + 1 = ?
9 – 6 = ?

Set Two
8 – 5 = ?
3 + 2 = ?

Set Three
1 + 3 = ?
7 – 4 = ?

Set Four
4 + 3 = ?
9 – 5 = ?
6 + 2 = ?

Set Five
8 – 4 = ?
2 + 5 = ?
7 – 6 = ?

Set Six
4 + 8 = ?
5 + 3 = ?
6 – 5 = ?

Set Seven
9 – 4 = ?
5 + 1 = ?
6 – 2 = ?
5 + 7 = ?

Set Eight
1 + 9 = ?
2 + 4 = ?
3 – 1 = ?
8 – 6 = ?

Set Nine
5 – 3 = ?
8 – 2 = ?
2 + 7 = ?
6 + 9 = ?

Set Ten
5 + 4 = ?
7 – 1 = ?
3 + 6 = ?
9 – 8 = ?
6 + 5 = ?

Set Eleven
8 – 7 = ?
1 + 4 = ?
9 – 3 = ?
2 + 8 = ?
5 – 1 = ?

Set Twelve
6 – 3 = ?
9 – 7 = ?
8 + 1 = ?
7 – 5 = ?
1 + 6 = ?

Set Thirteen
9 – 1 = ?
6 + 4 = ?
3 – 2 = ?
7 + 6 = ?
4 – 3 = ?
8 + 5 =?

Set Fourteen
7 – 2 = ?
4 – 1 = ?
8 – 3 =?
4 + 7 =?
9 + 5 =?
6 + 8 = ?

Set Fifteen
2 – 1 = ?
5 + 6 = ?
7 – 3 = ?
2 + 9 = ?
6 – 4 = ?
1 + 7 = ?