MEMO: Standardisation of a Novel Computerised Test of Auditory Verbal Short-Term Memory

Christos Salis1 (christos.salis@hcl.ac.uk), Faustina Hwang2(f.hwang@reading.ac.uk), Helen Kelly3(helen.kelly@ucc.ie), Niamh Coughlan3, Susan Swain1
1 Newcastle University, 2 University of Reading, 3 University College Cork

Background
Short-term memory (STM) impairments are prevalent in adults with acquired brain injuries. Assessing auditory-verbal short-term memory (STM) in people with aphasia and/or motor speech disorders using standard STM tests (e.g., forward digit span, Wechsler, 1981) can be challenging as word finding difficulties and high speech demands may confound the assessment accuracy of STM. In a systematic review, Snapka and De Leeuw (2007) noted the exclusion of people with severe strokes and people with aphasia from memory assessments because these groups could not complete evaluation with tests such as the forward digit span, a commonly used test in the reviewed studies.

A popular alternative to the forward digit span task in studies of STM deficits in aphasia is the matching listening span task (similar to the non-normed PALPA 13, Kay, Lesser & Coltheart, 1992). In this task, paired sequences of words (usually digits) are presented. The person hears the first sequence, followed by the second, and then decides whether the two sequences are the same or different in terms of word order. The matching span task (especially PALPA 13) is popular in studies of STM in aphasia (and possibly clinical practice) but crucial psychotic properties are not known. This current study provides normative data from healthy older adults and examines the test-retest reliability and concurrent validity of the matching listening span task, using a novel custom software application (MEMO). As STM is sensitive to lexical properties (Saffran & Martin, 1999), matching listening span for digits, words and non-words were investigated.

Method
Participants: 48 healthy participants (>60 years of age) were recruited from south-west Ireland, and the UK (north-east England and central Scotland). For the digit span all 30 participants were recruited from the UK and 12 of them also completed the word and non-words tasks (see Table 1). Participants lived independently in the community and did not have a history of neurological conditions. They achieved scores within normal limits in the Mini-Mental Status Examination and Boston Naming Test (short version). Hearing acuity in all participants (bar one) was 85% or above in the real word minimal pairs test (PALPA 2, Kay et al., 1992). The person who scored less than 85% was not excluded.

Table 1. Biographical characteristics of participants

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Gender</th>
<th>Education</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digits</td>
<td>72.2 (7.4)</td>
<td>15 men</td>
<td>15 higher</td>
<td>30 UK</td>
</tr>
<tr>
<td></td>
<td>range 60-85</td>
<td>15 women</td>
<td>15 lower</td>
<td></td>
</tr>
<tr>
<td>Words/Non-words</td>
<td>70.7 (7.0)</td>
<td>15 men</td>
<td>14 higher</td>
<td>18 Ireland 12 UK</td>
</tr>
<tr>
<td></td>
<td>range 60-88</td>
<td>15 women</td>
<td>16 lower</td>
<td></td>
</tr>
</tbody>
</table>

Method (cont.)

Lexical items: 60 word-list pairs for the three tasks – words (mono-syllabic nouns controlled for frequency and phonological and semantically dissimilar), non-words (mono-syllabic items controlled for phonotactic probability from Vluch & Luce (1999)), digits (1-9) were pre-recorded and assembled into word-lists using the MEMO software (Molero Martin, Laird, Hwang & Salis, 2013). The two word-lists presented contained the same words, with either two adjacent words transposed or an identical sequence. Across tasks there was an equal number of “same” and “different” sequences for each word-list length, with the order of transpositions evenly distributed between initial, medial and final positions to reduce recency and primacy affects.

Timing of presentation was specified by the researcher using the MEMO application and was controlled by the software. Each word was presented to participants auditorily through MEMO on a laptop computer, at an inter-word interval of 1 second and a delay of 2 seconds between the first and second word-lists in each pair. Each subset comprised span lengths progressively from 2 to 7 items (60 sequence pairs). There were 10 trials at each span length across subsets.

Scoring: Accuracy of performance in each task was calculated with A’ scores. A’ is a response bias measure which takes into account hits (i.e., responding “same” on “same” trials, responding “different” on “different” trials) and false-alarm rates. A’ scores typically range from 0.5, which indicates that a person is unable to distinguish “different” trials, to 1.0 which is perfect performance.

Test-retest reliability: For each subset 15 participants were tested again 3-4 weeks after the first testing session.

Results

Table 2. Normative performance on each task (expressed as A’ scores)

<table>
<thead>
<tr>
<th></th>
<th>Digits</th>
<th>Words</th>
<th>Non-words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>0.906 (0.045)</td>
<td>0.864 (0.056)</td>
<td>0.823 (0.065)</td>
</tr>
<tr>
<td>Range</td>
<td>0.819 - 0.985</td>
<td>0.731 – 0.974</td>
<td>0.686 – 0.924</td>
</tr>
<tr>
<td>SEM</td>
<td>0.021</td>
<td>0.037</td>
<td>0.029</td>
</tr>
</tbody>
</table>

Table 3. Test-retest reliability correlations (Spearman’s rho [N=15])

<table>
<thead>
<tr>
<th></th>
<th>Digits</th>
<th>Words</th>
<th>Non-words</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>0.786, p = 0.028</td>
<td>0.546, p = 0.035</td>
<td>0.801, p = 0.00</td>
</tr>
</tbody>
</table>

Discussion

- Findings from this study suggest the following cut-off scores (taking into account the SEM): for digits 0.797, for words, 0.693 and non-words 0.656. As sample size is small and non-random, use of these norms should be cautious.
- Age was associated with non-word task performance, whereas there was no association found between performance and gender, level of education or recruitment location (i.e. Ireland versus UK participants).
- Considering levels of test-retest reliability (Kline, 2000): Non-word matching span task - “good” level of reliability (0.8-0.9) suggesting that the non-word task is likely to produce a more reliable measurement than digits or words.
- Digit matching span task - “acceptable” level of reliability (0.7-0.8) indicating a more stable measure than the word tasks.
- Word matching span task - “unacceptable” level of reliability (<0.6) suggesting that it is not a stable measure.
- MEMO was found to be a flexible tool for collecting normative data which offered timing precision using different types of verbal stimuli with potential for treatment.
- Test availability: The computerised application (Molero Martin et al., 2013) and testing materials will be made freely available for clinical and research use (contacts: F. Hwang, C. Salis).

References

Acknowledgements
Project supported by a Dunhill Medical Trust Research Project Grant (R225/0512). We thank all participants and Laura McCann for their help with this study.