Supplementary Table S1. Full Electronic Search Strategy for Web of Science

#1 TI=(cognit* OR memory OR meta-memory OR metamemory) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article)

#2 TI=(abilit* OR checklist* OR complaint* OR everyday OR deficit OR difficult* OR failure* OR functional OR impairment* OR insight OR lapse*OR questionnaire OR rating scale OR self-efficacy OR self-knowledge OR self-perception OR self-rating scale OR subjective) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article)

#3 TI=(aphasia OR cerebro-vascular accident OR cerebrovascular accident OR cva OR dysphasia OR stroke) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article)

#4 #1 AND #2 AND #3
Supplementary Table S2. Subjective Memory Measure Psychometric Properties Quality Rating Tool

Note about scoring: For any variable/construct with items rated on excellent to poor scale (i.e., from COSMIN checklist), if even one item rated as poor, the score for that variable/construct is poor.

A. Validity:

1. **Content/Face validity** – looks like it should assess functional memory; degree to which scale is model-based; possible systematic sources of bias have been examined (e.g., cultural); experts designed the scale; theoretical/empirical foundation of scale and characteristics of items/subscales specified.

   a) Was there an assessment of whether all items refer to relevant aspects of the construct to be measured?

   Excellent = Assessed if all/relevant items refer to relevant aspects of the construct to be measured

   Fair = Aspects of the construct to be measured poorly described AND this was not taken into consideration

   Poor = NOT assessed if all/relevant items refer to relevant aspects of the construct to be measured

   b) Was there an assessment of whether all items are relevant for the study population? (e.g., age, gender, disease characteristics, country, setting)

   Excellent = Assessed if all items are relevant for the study population in adequate sample size (≥10)

   Good = Assessed if all items are relevant for the study population in moderate sample size (5-9)

   Fair = Assessed if all items are relevant for the study population in small sample size (<5)

   Poor = NOT assessed if all items are relevant for the study population OR target population not involved

2. **Concurrent validity** - scale outcomes have been found consistent with other valid measures of memory; have scores been found significantly related to other indices of functional memory ability?

   a) Was an adequate description provided of the comparator instrument(s)?

   Excellent = Adequate description of the constructs measured by the comparator instrument(s)

   Good = Adequate description of most of the constructs measured by the comparator instrument(s)

   Fair = Poor description of the constructs measured by the comparator instrument(s)

   Poor = NO description of the constructs measured by the comparator instrument(s)

   b) Were the measurement properties of the comparator instrument(s) adequately described?

   Excellent = Adequate measurement properties of the comparator instrument(s) in a population similar to the study population
Good = Adequate measurement properties of the comparator instrument(s) but not sure if these apply to the study population

Fair = Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population

Poor = No information on the measurement properties of the comparator instrument(s)

c) Were design and statistical methods adequate for the hypotheses to be tested?

Excellent = Statistical methods applied appropriate

Good = Assumable that statistical methods were appropriate, e.g. Pearson correlations applied, but distribution of scores or mean (SD) not presented

Fair = Statistical methods applied NOT optimal

Poor = Statistical methods applied NOT appropriate OR statistical methods not reported OR no correlation between two different measures of memory

3. Predictive validity - the scale should predict performance on other measures to which the results will be generalized; does the scale predict performance on other measures beyond functional memory?

Yes/no  

If yes, state other measures:

4. Discriminant validity - the scale has been shown to discriminate those with and without typical functional memory abilities

Yes/no

B. Responsiveness:

1. Were hypotheses about changes in scores formulated a priori (i.e., before data collection)?

Excellent = Hypotheses formulated a priori

Fair = Hypotheses vague OR not formulated but possible to deduce what was expected

Poor = Unclear what was expected

2. Was the expected direction of correlations or mean difference of the change scores of the scale included in these hypotheses?

Excellent = Expected direction of the correlations or difference stated

Good = Expected direction of the correlations or difference NOT stated

C. Cross-cultural validity:

1. Were both the original language in which the scale was developed, and the language in which the scale was translated described?

Yes/no
2. **Was the expertise of the people involved in the translation process adequately described?**

   - **Excellent** = Expertise of the translators described with respect to disease, construct, and language
   - **Good** = Expertise of the translators with respect to disease or construct poor or not described, but expertise of translators with respect to language described
   - **Fair** = Expertise of the translators with respect to language not described
   - **Poor** = No description of translators’ expertise

3. **Were translation procedures described (e.g., forward and backward translation)?**
   - Yes/no

**D. Reliability:**

1. **Test-retest reliability information**

   a) **was the sample size included in the internal consistency analysis adequate?**

      - **Excellent** = Adequate sample size (≥100)
      - **Good** = Good sample size (50-99)
      - **Fair** = Moderate sample size (30-49)
      - **Poor** = Small sample size (<30) OR sample size not reported OR test-retest reliability not reported

   b) **were at least two measurements available?**

      - **Excellent** = At least two measurements
      - **Poor** = Only one measurement

   c) **Were the administrations independent?**

      - **Excellent** = Independent measurements
      - **Good** = Assumable that the measurements were independent
      - **Fair** = Doubtful whether the measurements were independent
      - **Poor** = Measurements NOT independent OR not reported

   d) **Was the time interval stated?**

      - **Excellent** = Time interval stated
      - **Fair** = Time interval NOT stated

   e) **Were patients stable in the interim period on the construct to be measured?**

      - **Excellent** = Patients were stable (evidence provided)
      - **Good** = Assumable that patients were stable
Fair = Unclear if patients were stable
Poor = Patients were NOT stable OR patient status not reported

f) Was the time interval appropriate?
Excellent = Time interval appropriate
Fair = Doubtful whether time interval was appropriate
Poor = Time interval NOT appropriate

g) Were the scale conditions similar for both measurements?
Excellent = Scale conditions were similar (evidence provided)
Good = Assumable that scale conditions were similar
Fair = Unclear if scale conditions were similar
Poor = Scale conditions were NOT similar OR no reported

h) Was a correlation or other statistical analysis calculated?
Excellent = Kappa calculated and reported
Fair = Other statistical analysis calculated and reported
Poor = Only percentage agreement calculated OR method not reported

2. Internal consistency

a) Was the sample size included in the internal consistency analysis adequate?
Excellent = Adequate sample size (≥100)
Good = Good sample size (50-99)
Fair = Moderate sample size (30-49)
Poor = Small sample size (<30) OR not reported

b) Was Cronbach’s alpha calculated?
Excellent = Yes
Fair = Only item-total correlations calculated
Poor = No Cronbach’s OR item-total correlations OR method not reported

c) If Cronbach’s was reported, does it meet criterion (strict Cronbach alpha = or > .90; relaxed .80)? Report criterion.

E. Measurement Error:

1. Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?
Excellent = SEM, SDC, or LoA calculated

Good = Possible to calculate LoA from the data presented

Poor = SEM calculated based on Cronbach’s alpha, or on SD from another population OR not reported

F. Interpretability:

1. Was the sample size included in the analysis adequate?  
   Yes/no

2. Was the distribution of the (total) scores in the study sample described?  
   Yes/no

3. Was the percentage of the respondents who had the lowest possible (total) score described?  
   Yes/no

4. Was the percentage of the respondents who had the highest possible (total) score described?  
   Yes/no

5. Were scores and change scores (i.e. means and SD) presented for relevant (sub)groups? (e.g., for normative groups, subgroups of patients, or the general population)  
   Yes/no

6. Was the minimal important change (MIC) or the minimal important difference (MID) determined?  
   Yes/no
## Supplementary Table S3. Adapted Study Quality Rating Tool

<table>
<thead>
<tr>
<th>Appraisal Categories</th>
<th>Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High</strong></td>
<td><strong>Moderate</strong></td>
</tr>
<tr>
<td><strong>Study design</strong></td>
<td>case-series experimental design (experimental = control behaviour); large group (&gt; 10) with appropriate control group</td>
</tr>
<tr>
<td><strong>Demographic variables</strong></td>
<td>must report FIVE of the following: age; gender; education; marital status; language; socioeconomic status; ethnicity; co-morbid condition(s)</td>
</tr>
<tr>
<td><strong>Stroke variables</strong></td>
<td>must report FOUR of the following: stroke type (e.g., infarct); hemisphere affected/location; severity; time post-onset; stroke-related co-morbid conditions (e.g., hemiplegia, epilepsy); risk factors (e.g., smoking, blood pressure)</td>
</tr>
<tr>
<td><strong>Cognitive variables</strong></td>
<td>must report ALL THREE of the following: memory skills (e.g., working memory; attention or executive skills; language skills)</td>
</tr>
</tbody>
</table>

This Study Quality Rating Tool is based on information in National Institute of Health Research York University Guidelines and Criteria for Appraising Diagnostic Test Studies; Khan et al. [66]), STARD and COSMIN checklists. A study must score high in 3 out of 4 categories for an overall high rating (with no low rating); an overall moderate rating for a study cannot include any low rating.