



Published in final edited form as:

*Neuropsychol Rehabil.* 2018 April ; 28(3): 309–351. doi:10.1080/09602011.2016.1174718.

## The use of standardised short-term and working memory tests in aphasia research: a systematic review

Laura Murray<sup>a</sup>, Christos Salis<sup>b</sup>, Nadine Martin<sup>c</sup>, and Jenny Dralle<sup>d</sup>

<sup>a</sup>Department of Speech & Hearing Sciences, Indiana University, Bloomington, IN, USA

<sup>b</sup>Speech & Language Sciences, Newcastle University, Newcastle upon Tyne, UK

<sup>c</sup>Department of Communication Sciences & Disorders, Temple University, Philadelphia, PA, USA

<sup>d</sup>Department of Neurology, Brandenburgklinik, Bernau bei Berlin, Germany

### Abstract

Impairments of short-term and working memory (STM, WM), both verbal and non-verbal, are ubiquitous in aphasia. Increasing interest in assessing STM and WM in aphasia research and clinical practice as well as a growing evidence base of STM/WM treatments for aphasia warrant an understanding of the range of standardised STM/WM measures that have been utilised in aphasia. To date, however, no previous systematic review has focused on aphasia. Accordingly, the goals of this systematic review were: (1) to identify standardised tests of STM and WM utilised in the aphasia literature, (2) to evaluate critically the psychometric strength of these tests, and (3) to appraise critically the quality of the investigations utilising these tests. Results revealed that a very limited number of standardised tests, in the verbal and non-verbal domains, had robust psychometric properties. Standardisation samples to elicit normative data were often small, and most measures exhibited poor validity and reliability properties. Studies using these tests inconsistently documented demographic and aphasia variables essential to interpreting STM/WM test outcomes. In light of these findings, recommendations are provided to foster, in the future, consistency across aphasia studies and confidence in STM/WM tests as assessment and treatment outcome measures.

### Keywords

Short-term memory; working memory; aphasia; standardised tests; assessment

### Introduction

The presence of short-term and working memory impairments in aphasia is ubiquitous (Martin & Gupta, 2004; Murray, 2012a; Schuell, Jenkins, & Jimenez-Pabon, 1964). Short-term memory (STM) involves storage of information for a brief period of time, usually a few

---

CONTACT: Laura Murray, lmurray@indiana.edu, Department of Speech & Hearing Sciences, Indiana University, 200 S. Jordan Ave, Bloomington, IN, 47405 USA.

#### Disclosure statement

No potential conflict of interest was reported by the authors.

seconds, in a relatively unprocessed state (Baddeley, 2012; Cowan, 2010). This information could be auditory or visual and, within each of these modalities, verbal or non-verbal. When information, while being temporarily stored, is mentally manipulated to achieve a particular goal or plan, the manipulation is attributed to working memory (WM). Both STM and WM are considered capacity-limited systems indicating that a limited amount of information can be retained for a finite period of time (Cowan, 2010; Logie, 2011). A distinctive feature of STM is that of recall or recognition of information (often serially) in a relatively unprocessed state, whereas the emphasis in WM is deliberate manipulation, which draws on processes related to attention and goal execution. Therefore, assessments designed to measure STM and WM share some features (e.g., temporary maintenance of information); WM tests, however, include additional task demands such as updating or manipulating the information while it is being briefly retained.

To determine whether or not the integrity of STM and WM following brain damage is within normal limits, there is a need to rely on measurement instruments (or tests) that would help ascertain the presence and severity of the impairment, be it STM and/or WM, for rehabilitation planning, advising patients and caregivers, as well as documenting treatment outcomes. However, a construct can be measured with a range of tests, each placing different demands on STM and WM and bringing its own perspective on the nature of the impairment and its behavioural manifestation, the so-called mono-method bias (Coolican, 2014). The related issue of task impurity is also relevant because each task that measures an allegedly specific construct would rely upon a range of related or unrelated corollaries (cf., Miyake & Friedman, 2012). For example, in the context of aphasia, WM tests are inherently complex in terms of understanding task demands, and rely on understanding verbal instructions and examples. Consequently, it is especially important that the validity and reliability of STM and WM tests be of the highest quality. Indeed, a test with a higher quality in terms of psychometric properties would be associated with greater clinical confidence for accurate evaluation.

This review aims to identify and appraise standardised tests of STM and WM used in peer-reviewed studies of aphasia resulting from acquired and non-progressive neurological conditions affecting the language dominant hemisphere. We define aphasia as a range of impairments that affect a person's ability to produce and often understand linguistic units, that is, words, sentences, or discourse (Edwards, Salis, & Meteyard, 2015; Murray & Clark, 2015). In contrast to a circumscribed language problem related to a relatively isolated linguistic or perceptual issue (e.g., pure alexia; pure word deafness), aphasia is a complex disorder, with the majority of individuals with aphasia displaying a combination of spoken and written language production and comprehension symptoms. To our knowledge, this is the first systematic review of studies of aphasia involving standardised STM and WM tests.

Both verbal and non-verbal STM/WM deficits in auditory and visual modalities may co-occur in aphasia (e.g., De Renzi & Nichelli, 1975; Lang & Quitz, 2012). Such STM/WM deficits have been evoked as contributory and sometimes explanatory constructs in relation to several language abilities in aphasia. These range from broader language variables, such as aphasia severity (Crocket, Clark, Spreen, & Klonoff, 1981), potential for aphasia recovery (Seniów, Litwin, & Le niak, 2009), and prognosis for linguistic treatments (Harnish &

Lundine, 2015), to more discrete linguistic levels, such as lexical processing (Martin & Ayala, 2004), aspects of sentence processing, as well as spoken and written discourse comprehension (Caspari, Parkinson, LaPointe, & Katz, 1998; Leff et al., 2009; Lehman & Tompkins, 1998; Martin & Allen, 2008; Sung et al., 2009). Furthermore, Sulleman and Kim (2015) have recently argued that WM limitations may negatively affect the ability of people with aphasia to make well-informed decisions about aspects of their rehabilitation. The clinical implication suggested by these studies, albeit not always explicitly, is that STM/WM abilities, both verbal and non-verbal, need to be assessed and, consequently, incorporated into the clinical decision making process to understand a person's difficulties and strengths. Finally, a recent trend in the experimental rehabilitation literature has been the development and examination of STM and WM treatment protocols not only to remediate memory impairments but also concurrently to improve language and, in some cases, psychosocial functioning (see reviews by Murray, 2012a; Salis, Kelly, & Code, 2015). If such treatments are to be replicated, refined, and ultimately implemented in clinical practice, there would be a need for psychometrically sound STM/WM measurement instruments to establish a diagnosis, explicate the nature and severity of the impairments, implement and monitor treatment, and measure the outcome (Turkstra, Coelho, & Ylvisaker, 2005; de Vet, Terwee, Mokkink, & Knol, 2011).

Nonetheless, several issues augur investigation of the tests used to qualify and quantify STM/WM abilities in people with aphasia (Mayer & Murray, 2012; Wright & Fergadiotis, 2012). A plethora of STM/WM measures, both standardised and experimental, have been utilised in the empirical aphasia literature, in part a reflection of the different theoretical conceptualisations and the multidimensional nature of these memory constructs. However, such diversity in measures poses challenges. First, it confounds resolving discrepant findings regarding the presence and/or strength of relationship between these memory skills and specific linguistic processes (e.g., Martin, 2009 vs. Majerus, Attout, Artielle, & Van der Kaa, 2015). Second, it muddles the search for appropriate STM/WM assessment tools by both researchers and clinicians. Third, it remains challenging to find research documenting the extent to which standardised tests and experimental tasks represent valid and reliable measures of STM/WM in the aphasic population. Such research is essential when using STM/WM measures to prognosticate and/or evaluate aphasia treatment outcomes. Another challenge, particularly pertinent in evaluating auditory-verbal STM and WM in aphasia, is that the response modality of many STM and WM tests involves the very same modalities that are impaired in aphasia. For example, repetition and word retrieval difficulties are impaired in aphasia and may confound STM and WM measurement, which often draws upon repetition and word retrieval (cf., Howard & Franklin, 1990). Likewise, motor speech skills can also be impaired (i.e., apraxia of speech, dysarthria), even in cases of mild aphasia (Basilakos, Rorden, Bonilha, Moser, & Fridriksson, 2015; Bose & van Lieshout, 2008). It is also of interest to examine whether recent advances in cognitive testing (e.g., computerised test delivery) have been incorporated into the evaluation of STM and WM abilities in individuals with aphasia.

Consequently, the applied goal of the present systematic review is to put forward recommendations for clinicians, researchers, and other stakeholders regarding the suitability of tests when identifying or monitoring STM/WM, both verbal and non-verbal, in

individuals with aphasia. To our knowledge, such a comprehensive analysis of tests has not been attempted previously. Specific aims of the present review are as follows:

1. To identify standardised tests of STM and WM (verbal and non-verbal) utilised in the adult, acquired, non-progressive aphasia literature from 2000 to 2015. We focused on standardised tests as opposed to experimental tasks because the former category is likely to have more robust psychometric properties and wider availability, and thus more suitable appeal to evidence-based clinical practice. The time period reflects our goal to offer the most current assessment recommendations, and thus identify tests based on contemporary conceptions of STM/WM with recently documented normative data and a recently established evidence base.
2. To evaluate critically the psychometric strength of these tests. This aim is embedded in key principles of evidence-based practice in that tests with stronger psychometric profiles would be preferable to those with weaker profiles (Greenhalgh, 2014). This critical appraisal is also essential to identifying tests worthy of recommendation for future use in research and clinical practice.
3. To evaluate critically the quality of the investigations utilising these tests. In a similar vein to evidence-based practice, *ceteris paribus*, if a study has utilised tests with stronger psychometric properties, its findings would be more robust. Likewise, if an investigation in which a test was developed and/or utilised has a strong study design and reporting features, its findings would be more robust. In contrast, when an investigation is poorly designed and lacks methodological detail, it cannot be replicated and such procedural issues confound confident interpretation and future application of the test results and study outcomes. Recommendations regarding STM and WM tests suitable for individuals with aphasia, therefore, should be developed in consideration of not only what tests have been used in the aphasia literature, but also the quality of studies using such tests.

## Method

Procedures adhered to previously established methods for performing and describing systematic reviews (Khan, Kunz, Kleijnen, & Antes, 2003; Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009; Schlosser, Wendt, & Sigafos, 2007). This included developing beforehand our systematic review protocol for the literature search, including eligibility criteria and methods to gather and assess the quality of the data of interest.

## Search strategy

A comprehensive list of previously established search terms was developed and operationalised into three subcategories: construct related, population related and topic related (see Table 1). Using the terms in Table 1, the following electronic databases were searched: Cumulative Index to Nursing and Allied Health (CINAHL), Linguistics and Language Behaviour Abstracts (LLBA), Medline, and PsychINFO. Search terms within a subcategory were combined with the operator “OR” and across subcategories with the

operator “AND” to derive a final list of citations. In addition, the on-line search functions of Science Direct and Taylor & Francis were also searched through the advanced search option using a simpler, two-step search strategy with the following terms: “short-term memory” AND “aphasia”, “working memory” AND “aphasia”. The final list of citations from all databases and Science Direct was exported into EndNote™ reference management software, which removed duplicate citations. A subsequent hand search of the eligible citations removed further duplicate papers that EndNote™ did not identify. These digital searches were supplemented by searching other sources. These were as follows: (1) a hand search of all papers published in the journal *Aphasiology* was also carried out to identify relevant papers; (2) a search of reference lists in STM/WM review papers that appeared in a special issue of *Aphasiology* on short-term memory and aphasia (Murray, 2012a); (3) for commercial tests the websites of Pearson and Psychology Press were reviewed; and, (4) contacting authors for difficult to obtain studies (i.e., Rey Complex Figure Test, version by Meyers & Meyers, 1995) or additional information about tests (i.e., Friedmann & Gvion, 2002). Duplicate citations that had been generated from the electronic searches were noted and excluded. In all searches, the timescale was from January 2000 until 15 April 2015.

### Inclusion and exclusion criteria

To be included in the review, a study had to meet the following inclusion criteria:

1. Study participants included adults (i.e., 18 years or older) with non-progressive, acquired aphasia due to any aetiology (e.g., stroke, traumatic brain injury, tumour, infection); we did not apply restrictions of aetiology although we were mindful that in some aetiologies, particularly traumatic brain injury and communication disorders associated with right hemisphere damage, the term aphasia per se may be used (e.g., Myers, 2001; Sarno, 1980). Unless the term aphasia was used to identify participants, such studies were not included.
2. When mixed participant groups were utilised (e.g., participants with and without aphasia within an acquired brain injury group), it was possible to identify the STM/WM assessment outcomes for the participants with aphasia (separate from those participants without aphasia).
3. STM and/or WM were assessed via a standardised test with norms clearly identified and/or referenced in the study; in this review, a standardised test was defined as a test with clearly defined procedures for administration and scoring that includes norms with reference to scores from a normative sample (Anastasi & Urbina, 2009; Turkstra et al., 2005). In addition, the duration of the information (either auditory or visual) that had to be retained or manipulated by participants following exposure of stimuli should not exceed 30 seconds (Peterson & Peterson, 1959).
4. The study was peer-reviewed or was a non-peer reviewed standardised test manual.
5. The study or test manual was published in English.

Studies were excluded if they did not meet one or more of the above criteria, did not include original data (e.g., meta-analysis, review paper), and/or were unpublished dissertations or conference presentations. Studies were also excluded if they used STM or WM experimental tasks but failed to provide the stimuli and/or a description of the standardisation process, either within the same study and/or a citation for such information. Finally, we excluded studies in which it was clear that the participants with aphasia had been duplicated (i.e., the same participants with aphasia were included in more than one study). In such cases, studies that included the largest number of participants with relevant measures were included, whereas studies that reported subsets of such participants were excluded. This decision is reflected in Figure 1. The purpose of this final exclusionary procedure was to maintain accuracy about the number and breadth of participants with aphasia involved in the literature base pertaining to STM/WM assessment.

### Screening and eligibility

After removing duplicates, study titles and abstracts from the searches were screened against the eligibility criteria. In cases in which neither the title nor abstract indicated eligibility, the full text was screened, recording the reasons if these studies were subsequently excluded. Although all authors participated in screening studies for inclusion, the involvement of each author varied at different points in the screening and eligibility process. To ensure inclusion of studies, any queries regarding the eligibility of individual papers were addressed by consulting at least one additional independent rater from the research team. Any disagreements were discussed and resolved jointly. Additionally, two authors (CS and LM) independently screened a randomly selected sample of 100 studies; inter-rater agreement was 93%, with discrepancies resolved via discussion.

### Data extraction

For each study meeting all eligibility criteria, data pertaining to the following were extracted: (1) study aims/objectives; (2) participant sample information including sample size, presence and type of comorbid conditions (e.g., hearing/vision screening; hemiparesis), age, education, gender, native language, aetiology, and aphasia type and severity profiles; (3) assessment setting (e.g., location at which testing took place, qualifications of assessor); and (4) STM/WM test(s) information, including the test name, which aspects of STM/WM were assessed (e.g., visual STM), type of test scores recorded (e.g., raw, scaled), and psychometric characteristics (e.g., inter- and intra-rater reliability, test construct validity). Data relating to other cognitive deficits (i.e., beyond aphasia and STM/WM abilities) were also gathered, but because of the inconsistency of these data across studies, this information is not reported. As in the screening and eligibility stage, all authors participated in data extraction of included studies, although the amount of involvement of each author varied at different points.

### Quality appraisal

Each study that underwent data extraction was evaluated for quality using an assessment tool adapted from the Guidance for Undertaking Reviews in Healthcare (Centre for Reviews and Dissemination, 2008), systematic review guidelines proposed by Khan et al. (2003), and checklists from the STARD (Bossuyt et al., 2003) and COSMIN (Mokkink et al., 2009,



2010) (see Appendix 1). An adapted rating tool was necessary given that existing quality appraisal scales were not suitable for the variety of study designs and/or participant sample characteristics and issues encountered in the aphasia literature. Our adapted tool appraised study quality in terms of five categories: study design, control for confounding factors, specification of aphasia and assessment variables, and STM/WM test score(s) interpretation. Ratings of high, moderate, or low were assigned for each quality category as well as the study as a whole. For a given study to receive an overall high quality rating, four of the five categories had to achieve a high rating with no category receiving a low rating; a study with an overall moderate rating could also not have any category receiving a low rating. Two authors (LM and JD) completed the study quality ratings. Inter-rater agreement was examined for 31 papers (out of 73 extracted papers; see Figure 1), and yielded 83% agreement across all items, with 99% agreement for each paper's overall quality rating. All discrepant ratings were resolved via discussion.

In studies in which a standardised STM/WM test(s) was only used to characterise the aphasia participant sample (versus examine the STM/WM test for use with the aphasia population), the test manual or reference paper cited within the given study was reviewed to identify the test's psychometric properties. Only the provided reference was analysed to describe and appraise psychometric strengths and weaknesses of the test as this was seen as the original source by the study authors. Different data extraction forms were developed for these test papers and manuals which included items on the following: (1) normative sample variables including whether or not adults with acquired, non-progressive aphasia were included in the test's standardisation process and the appropriateness of the standardisation sample (e.g., age and education appropriate) given the aphasic participant(s) characteristics in the eligible paper which cited the test; (2) test administration characteristics (e.g., information on the assessment environment); (3) validity (i.e., construct, content/face, criterion-related, and discriminant validity); (4) reliability (i.e., test-retest, split-half/internal consistency, and inter-rater); and (5) measurement error.

Each STM/WM test utilised in the set of included studies was also appraised to rate the quality of its psychometric properties. Currently, however, there is no widely used existing "gold standard" for assessing STM/WM in aphasia (cf., DeDe et al., 2014). Accordingly, an appraisal tool (see Appendix 2) was developed by adapting the COSMIN checklist, which has empirical support of its reliability (Mokkink et al., 2010) and validity (Mokkink et al., 2009), in concert with the criteria for test reliability and validity established by the Agency for Healthcare Research and Quality Evidence-Based Practice Program (Biddle et al., 2002). These criteria have been previously used by the Academy of Neurological Communication Disorders and Sciences to develop practice guidelines (e.g., Turkstra et al., 2005). As an example of how the COSMIN checklist was adapted, given that most STM/WM measures are subtests within a test battery (e.g., Digit Span of the Wechsler Memory Scale-Revised; Wechsler, 1987), COSMIN checklist items pertaining to internal consistency across test (sub)scales were not applicable and, thus, not included in our appraisal tool. Two authors (CS and JD) completed the quality ratings of the STM/WM tests. Inter-rater agreement was examined for 57.5% of the tests (i.e., 19 of 33 STM/WM tests). There was 94% agreement across all rated items, with 100% agreement for each test's overall quality rating. Discussion was utilised to resolve any discrepant ratings.

## Final study selection

Given the large number of studies that underwent data extraction, a post-hoc decision was made to categorise eligible studies into either (1) those in which the study purpose directly related to describing, assessing, or treating STM/WM abilities in aphasia, or (2) those in which STM/WM assessment was ancillary to the study purpose (e.g., the study focus was a word retrieval intervention and STM/WM assessment was completed as part of a comprehensive assessment of aphasic participants).

Primary reasons for study exclusion were: (1) the study listed aphasia as an exclusionary criterion; (2) there was no specification that acquired brain injury participants had aphasia (this was a particularly common basis for excluding traumatic brain injury sequelae or treatment studies); (3) when individuals with aphasia were included, their STM/WM test results were not separated from those of the individuals without aphasia (this was a particularly common basis for excluding stroke sequelae or treatment studies); (4) no standardised STM/WM test was used; and (5) the citations provided for standardised tests were wrong.

## Results

The search results across databases and other searches are shown in Figure 1, together with the results from the screening and eligibility processes as well as the post-hoc final study selection. Of the 7299 studies screened, only 73 were deemed eligible. The 36 studies that became the main focus of the review are shown in Table 2. On the basis of these studies, the STM/WM tests that were used within them were critically appraised and are shown in Tables 3–8. The studies that used STM/WM tests but the main purpose of these studies was not STM/WM are shown in Table 3. These studies will not be discussed further.

### Participants in the final set of eligible studies

There were 898 participants with aphasia (Table 2). In terms of aphasia characteristics, neither aphasia type nor severity was consistently reported (e.g., Allen, Martin, & Martin, 2012; Lang & Quitz, 2012). For instance, severity of aphasia was specified in only 16 (of 36, or 44%) studies. When aphasia type was noted, a variety of aphasia classification systems was used: Some studies more broadly only noted whether participants had fluent versus nonfluent aphasia (e.g., Carragher, Sage, & Conroy, 2013), whereas other studies used a more complex system such as the Boston classification system (e.g., DeDe et al., 2014). Participants with anomic aphasia (144) and/or a mild severity of aphasia (73) were the most common when authors reported these variables. In contrast, among studies specifying aphasia type and/or severity, individuals with Wernicke's (38) or severe aphasia (13) were under-represented in the participant samples compared to the other aphasia types and severities, respectively. Across studies, participants with aphasia representing a range of education levels and ages were included. In several studies, however, education level information was either not provided (e.g., Galling et al., 2014; Sinotte & Coelho, 2007) or described in general terms (e.g., Lang & Quitz, 2012 who described education level in terms of less or more than nine years of formal education). Additionally, across studies, there were



relatively few participants with aphasia over the age of 70 compared to those younger than age 70.

### Standardised STM/WM tests used in final set of eligible studies

The auditory-verbal STM tests and WM tests utilised in the final set of extracted studies are displayed in Tables 6 and 7, respectively; Table 8 provides a summary of the visuo-spatial STM and WM tests used. Quality appraisal ratings of these tests are displayed in Tables 9–11.

**Auditory-verbal STM tests**—Review of Table 6 indicates that across studies, serial recall was the most frequently used task to assess auditory-verbal STM. Digit Span<sup>1</sup>, albeit from several different standardised tests and administered in a number of different languages, appeared the most popular auditory-verbal STM task being used with 272 (out of 898) or 30.2% of the participants with aphasia. The second most popular test was the Immediate Free Recall condition of the Rey Auditory Verbal Learning Test (RAVLT; Rey, 1964), which had been used with 34 (out of 898) or 3.7% participants with aphasia, albeit in only one study. The least popular test, used with only one participant with aphasia, was the Revised Token Test (RTT; McNeil & Prescott, 1978). Among the 20 different tests (or subtests of larger batteries) listed in Table 6, 17 emphasise serial recall (either forward or backward) or recognition. In contrast, only two tests focus on free recall<sup>2</sup> (RAVLT; Word Span Probe test of Friedmann & Gvion, 2002). In terms of the response demands, the majority of auditory-verbal STM tests (12 out of 20) require spoken output; instead of a spoken response, in the remaining eight tests, examinees indicate recalled information via either a pointing response or a recognition judgement (e.g., yes/no response).

**Auditory-verbal and visual-verbal WM tests**—Compared to the number of auditory-verbal STM tests just reviewed, fewer auditory-verbal working memory tests (i.e., six) were used within the eligible studies (Table 7). Complex span measures (i.e., Listening Span tests, Eye Movement WM task), which place demands on the shifting component of WM (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000), were most common, being administered to 156 (out of 898, 17.3%) of the participants with aphasia. The other less frequently used tests (i.e., TEA subtests, *n*-back of DeDe et al., 2014) place greater demands on updating functions within WM (Morris & Jones, 1990); these tests had been administered to only 88 (out of 898; 9.7%) of the participants with aphasia. There was an even representation of auditory-verbal WM tests requiring a spoken response versus a response in another modality (i.e., pointing or eye gaze).

**Visuo-spatial STM and WM tests**—In contrast to the great variety of tests used to measure auditory-verbal and visual-verbal STM or WM abilities, only a limited number of visuo-spatial STM and WM tests were identified in the literature; accordingly, both visuo-

<sup>1</sup>The combination of the two versions of the Digit Span (forward and backward recall) in Table 6 and visuo-spatial correlates in Table 8 (e.g., forward and backward WMS-R visuo-spatial span subtests), does not imply that the two versions reflect similar processes. Although backward recall is often regarded as a WM (as opposed to an STM) task, it does differ from the other WM tasks we came across in terms of its complexity. Serial recall is not an explicit feature of WM tasks. Furthermore, we did not encounter separate exploration or description of the psychometric properties of standardised forward versus backward recall subtests.

<sup>2</sup>Only a limited number of RTT subtests are non-serial tasks.

spatial STM and WM test findings are described here and collapsed into Table 8. As in assessment of auditory-verbal STM, serial recall tasks were the most popular for evaluating visuo-spatial STM abilities. That is, of the nine different tests (or subtests from larger test batteries), five were used to measure visuo-spatial span or serial recall. With 365 (out of 898; 40.6%) of the participants with aphasia completing a visuo-spatial span test, such tests represent the most frequently used STM measure among the eligible studies. In contrast, immediate recall of complex designs was rarely used to evaluate visuo-spatial STM, with administration to only 1 (out of 898) participant with aphasia. The TEA Visual Elevator was the only standardised visuo-spatial test encountered among the eligible studies to place substantial demands on shifting and updating components of WM. The most frequent mode of response among the visuo-spatial STM and WM tests was pointing. The WMS-R Visual Reproduction subtest requires a drawing response and the TEA Visual Elevator subtest requires a spoken response (i.e., counting).

**Quality appraisal of standardised STM/WM tests**—Each standardised STM/WM test encountered in the eligible studies (except for tests with psychometric data published in unobtainable manuals or studies) was evaluated in terms of its psychometric properties (see Appendix 2 for the test appraisal tool). Quality ratings for the auditory-verbal STM tests are displayed in Table 9. In terms of validity, every auditory-verbal STM test except for the RTT appropriately documented discriminant validity. Across these tests, however, other types of validity were either not reported or received fair or poor ratings. Only five of the 16 auditory-verbal STM tests received an excellent rating for construct validity, and only the WAIS-III Digit Span received an excellent rating for content/face validity. Concurrent validity was rated as poor in 13 tests, with the remaining three receiving a fair rating. The tests fared poorly in terms of all types of reliability examined, with no excellent ratings. Only three tests were rated as having fair test-retest reliability, whereas only the WMS-R Digit Span received a fair rating for split-half reliability.

As Table 10 shows, there were issues with the psychometric characteristics of the auditory-verbal WM tests. Among the nine tests, seven received an excellent rating for their construct validity. Among the other types of validity, however, the only excellent rating was for the concurrent validity of the English version of the Eye Movement WM Span test (Ivanova & Hallowell, 2014). Only the Listening Span task of Tompkins et al. (1994) provided evidence of predictive validity. Discriminant validity was documented in six (out of nine) of these tests. All types of reliability and measurement error were either rated as poor or not reported.

Appraisal of the STM and WM visuo-spatial tests indicated that each had excellent construct validity (see Table 11). Other types of validity received less positive ratings, with only the WMS-III visual tapping test receiving an excellent rating for content/face validity and only the Block Tapping of Kessels et al. (2008) receiving an excellent rating for concurrent validity. Both predictive and discriminant validity were reported inconsistently across the eight visuo-spatial STM tests that were quality appraised: three provided evidence of predictive validity and six provided evidence of discriminant validity. In terms of reliability, the only excellent ratings were for the test-retest reliability of the WMS-R visual reproduction test and the split-half reliability of the WMS-R visual tapping test. The WMS-III visual tapping test received a fair rating for its test-retest and split-half reliability and also

was the only test to include inter-rater reliability information. All other reliability quality ratings were poor. Measurement error was rated as poor except for the WMS-R and WMS-III visual tapping tests and the WMS-R visual reproduction test, all of which received an excellent rating.

**Quality appraisal of final set of eligible studies**—Table 12 lists the quality ratings for each of the 36 studies in the areas of design, control for confounds, aphasia variables, assessment variables, STM/WM score interpretation, and overall study quality. The majority of studies received a high quality rating in the areas of design (23/36) and STM/WM score interpretation (25/36). Of concern were the majority of low quality ratings in the area of assessment variables (24/36), with only a few studies stating in what environment participants were evaluated and/or who administered the test(s) and their professional qualifications. Few studies received a high quality rating in the area of control for confounds (7/36), with more than half of the studies (20/36) failing to indicate whether the effects of age and education on test performance were controlled or considered (i.e., a low rating). Accordingly, keeping in mind that a low quality rating in any category resulted in a low overall study quality rating, only three studies (Chiou & Kennedy, 2009; Fucetola et al., 2009; Ivanova & Hallowell, 2014) received a high overall study quality rating, and three studies (DeDe et al., 2014; Kalbe et al., 2005; Meteyard et al., 2015) received a moderate overall study quality rating.

## Discussion

The purpose of this systematic review was to comprehensively analyse standardised tests of STM and WM, both verbal and non-verbal, used in the contemporary aphasia literature. Our review involved not only identifying STM and WM tests, but also critically appraising both the psychometric properties of these tests as well as the quality of the aphasia investigations in which the tests were used. Overall, although a wide variety of standardised tests have been used to characterise STM and WM in individuals with aphasia, those that measure serial recall appeared most common, and substantial issues with respect to the psychometric strength of the STM/WM tests as well as the quality of studies were identified. Below is a more detailed discussion of the quality appraisal. This is followed with recommendations for improving assessment of STM and WM in aphasia, in both research and clinical practice.

### Standardised STM and WM tests

**Quality appraisal of auditory-verbal STM tests**—Auditory-verbal STM tests were the most popular. Within this broad category, the most popular task was Digit Span that required spoken recall. It was used with 272 persons with aphasia across 15 different studies. Such popularity may reflect that, historically, Digit Span, as a measure of STM ability, was one of the very first to be included in intelligence testing scales, dating back to the late 19th century (Richardson, 2007). Since the late 1930s, it has been incorporated into the test batteries of Wechsler and is still present in their recent versions. Digit Span has also had a long history of use in aphasiology (Eling, 2015; Schuell et al., 1964). Indeed, the Digit Span subtest of the Wechsler batteries was the most popular version in the current systematic review compared to other, more recent, versions (CAT; Computerised Neurocognitive Test).

The four versions of Digit Span with documentation available for our quality appraisal were rated as having excellent construct validity and all had documentation of predictive and discriminant validity (Table 9). However, a mixed profile of quality was found for other aspects of validity and for reliability. For example, content/face validity was deemed excellent only in the WAIS-III while poor in the other three versions (CAT, WAIS-R, WMS-R). Measurement error was poor only in the CAT. The relatively low levels of test-retest reliability for Digit Span have been known for some time, making it customary to combine scores from its forward and backward recall versions to improve reliability (Richardson, 2007). There is evidence to suggest that reliability coefficients improve when scores from different tasks that relate to a particular psychometric property are combined (DeDe et al., 2014; Swinburn et al., 2004; Waters & Caplan, 2003). For example, Waters and Caplan (2003) showed that in non-brain-damaged adults (younger and older), test-retest reliability was acceptable ( $r = .70$ ) when individual memory test scores were combined.

One study (Caza et al., 2002) used versions of Digit Span with old normative data based on 1957 and 1969 editions of the Wechsler tests. The so-called Flynn effect refers to the increment of IQ scores as time progresses (Flynn, 1984, 2009). Accordingly, older normative data as reference points may jeopardise discriminant and predictive validity. The Flynn effect has been evident in Digit Span data (Wicherts et al., 2004) and could also operate in other STM and WM tests that use historical normative data. Loring and Bauer (2010) noted that what makes a test outdated is not necessarily the publication of a more recent version of it, but rather empirical evidence the new edition is more valid and reliable, always with reference to the clinical population for which the test is intended. To our knowledge, such empirical research for clinical use of Digit Span (not only the Wechsler but also other versions) with persons with aphasia, does not exist.

There were two additional versions of Digit Span that did not require speech production: the pointing and matching span versions of the FriGvi (Friedmann & Gvion, 2002), a test developed in Israel for speakers of Hebrew. Both tests had fair construct validity and did display discriminant validity in differentiating STM performance in people with aphasia. However, both tests were poor in other aspects of validity, reliability, and measurement error. We should note that unlike some Digit Span tests requiring spoken recall (Wechsler versions, CAT), which present only two trials per span length, the FriGvi pointing version presents five trials per list length. As Woods et al. (2011) noted, the two trial paradigm assumes that a person's true maximum length span can be assessed by only four list presentations: two at the maximum length and two above. However, this method may seriously underestimate the maximum length of persons who are distracted or encounter idiosyncratically difficult digit strings (e.g., permutations of their telephone area code) at a particular length.

Relying on Digit Span for assessing auditory-verbal STM in aphasia presents with other possible limitations. Numerical skills are often impaired in aphasia, so interpretation of Digit Span performance on its own may not truly reveal the integrity or decrement of STM (DeDe et al., 2014). Furthermore, in aphasia, STM has been found to be sensitive to the lexical processing characteristics of the words within the STM test (e.g., lexicality, frequency) (e.g., Howard & Nickels, 2005; Martin & Ayala, 2004; Martin, Saffran, & Dell, 1996). For example, because the lexical frequency of digits is high in comparison to other words

(Martin, Lesch, & Bartha, 1999), relying only on digits to evaluate STM may yield inaccurate results.

Only the non-word span and the probe word span of the FriGvi (Friedmann & Gvion, 2002) explicitly assess the influence of lexical variables in STM. Both tests were used in two studies, with a total of 17 participants with aphasia completing each test. Knowing if lexicality and other lexical variables influence STM has diagnostic and treatment implications. Studies have shown that the nature of auditory-verbal STM deficits in aphasia can vary along the phonological-semantic dichotomy and in some individuals can be differentially spared or impaired (e.g., Martin & Allen, 2008; Martin & Ayala, 2004).

Only two tests did not tap into serial aspects of STM, the immediate condition of the Rey Auditory Verbal Learning Test (RAVLT; Rey, 1964) and the probe word span of the FriGvi. The ability to process language effectively relies heavily on the ability to process information serially, and this may explain the popularity of serial STM tests. Regarding the RAVLT, we only included studies that reported results for the immediate recall condition, which assesses STM. Subsequent recall conditions rely on long-term memory. We were unable to obtain the 1964 version of the RAVLT used by Vukovic et al. (2008), so are not in a position to appraise it. Whereas we are not aware of studies on the Flynn effect in relation to the RAVLT, Baxendale (2010) found that among healthy adults from the UK, verbal learning ability as measured by a test similar to the RAVLT was relatively stable across time with no significant differences between the scores in the majority of age ranges, apart from the 31–45 years age group. However, it should also be noted that Vukovic et al. (2008) administered the RAVLT in Serbian and used the test materials but not the norms.

**Quality appraisal of auditory-verbal and visual-verbal WM tests**—Compared to auditory-verbal STM, a more limited number of standardised tests have been used to evaluate auditory-verbal or visual-verbal WM in individuals with aphasia (Table 7). Half of these WM tests were complex span tasks (Eye Movement WM task, Listening Span by spoken recall or written recognition), which place heavy demands on the WM submechanisms of rehearsal and shifting; the other half (i.e., TEA Elevator Counting with Distraction and with Reversal, *n*-back) evaluate WM more in terms of its monitoring and updating submechanisms (Conway et al., 2005; Kearney-Ramos et al., 2014; Salis et al., 2015; Wright & Fergadiotis, 2012). The complex span tasks were more popular in that they were used in a larger number of studies and with a larger number of participants with aphasia.

The auditory-verbal and visual-verbal WM tests also varied in terms of whether they did (e.g., Elevator Counting with Reversal) or did not require a verbal response (e.g., *n*-back). Within the group of tests not involving a verbal response, a variety of nonverbal response modalities was used (i.e., pointing, computer key press, eye movement). Regardless of response modality, the complex span tasks had greater language demands (i.e., all required sentence processing) compared to the updating tasks. In fact, Tompkins et al. (1994) warned that their complex listening span test was likely unsuitable for individuals with severe aphasia.

Research in healthy as well as other patient populations indicates cognitive demand differences between complex span versus updating WM tests (e.g., Jaeggi, Buschkuhl, Perrig, & Meier, 2010; Kane, Conway, Hambrick, & Engle, 2007), which in turn may lead examinees to use different strategies when completing such tests (Logie, 2011). Despite these findings, both types of tasks were used in only three studies (i.e., DeDe et al., 2014; Mayer & Murray, 2002, 2012b). Whereas debate persists concerning the theoretical architecture of WM, multidimensionality is a common feature (Logie, 2011; Wright & Fergadiotis, 2012), thus suggesting that a test examining a limited set of WM submechanisms may not fully characterise WM abilities. Consequently, as we and others (Conway et al., 2005; DeDe et al., 2014) have noted, until a more comprehensive verbal WM measure is developed, the practice of utilising just one test to characterise WM abilities should be avoided.

Quality appraisal findings further supported the conclusion that reliance on only one auditory-verbal or visual-verbal WM test is inadequate. Despite an excellent rating for construct validity across most of the verbal WM tests used in the eligible studies (Table 10), ratings for other aspects of validity indicated substantial problems. For example, all of the tests received poor ratings for content/face validity, and only one test (Listening Span of Tompkins et al., 1994) had evidence of predictive validity. Reliability and measurement error were uniformly problematic for all of these WM tests. The most common issues leading to less desirable quality ratings included insufficient description of procedures used to examine validity or reliability (e.g., stating a correlation was calculated, but not specifying if it was an intra-class, Pearson, or Spearman), failure to include information regarding certain psychometric properties (e.g., split-half reliability and measurement error were rarely mentioned), and restricted sample sizes (which compromise certain aspects of reliability). Thus, although complex span tests were found to be used more frequently in the aphasia literature, there does not appear to be a psychometric rationale for their popularity compared to the other types of WM measures (i.e., *n*-back; TEA subtests). More generally, as previous authors have noted (Salis et al., 2015; Wright & Fergadiotis, 2012), currently available auditory-verbal and visual-verbal WM tests require further empirical development (e.g., modifications to support performance of those with severe language difficulties) and evaluation to determine if their use with individuals with aphasia can yield psychometrically sound data.

**Quality appraisal of visuo-spatial STM and WM tests**—Visuo-spatial span tests were the most popular type of test for assessing visuo-spatial STM and WM, with a total of 365 aphasic participants tested, and half of the eligible studies including one or more visuo-spatial STM or WM test (Table 8). Such popularity in the aphasia literature was expected given the relatively reduced language demands of visuo-spatial STM/WM tests compared to their auditory-verbal or visual-verbal counterparts. Among the types of visuo-spatial STM tests identified in the appraised literature, serial recall tasks were most prevalent. Of the four visuo-spatial serial recall tests reviewed, the WMS-III visual tapping subtest received the strongest quality ratings, although evidence of its discriminant validity was lacking (Table 11). Notably, the newer visual tapping (WMS-III) did represent an improved version of the older WMS-R visual tapping in several psychometric domains. Reliability and measurement



error were areas of significant concern for the visuo-spatial STM tests developed by Kessels et al. (2008), a version of a Corsi block tapping task, and DeDe et al. (2014): Both tests received poor quality ratings for these psychometric properties and neither reported inter-rater reliability. We should note that several studies (e.g., Berthier et al., 2011) used block tapping tests but were not included in this review because the wrong citations were provided. Milner (1971) was one of these erroneous citations: Milner (1971) referred to Corsi's doctoral research (i.e., Corsi, 1972), which involved block tapping as a Hebbian learning task rather than visuo-spatial STM span test per se. Another problematic citation for the block tapping test was that of De Renzi and Nichelli (1975) who referred to their block tapping task as a "spatial span task" (p. 344), but provided insufficient description of how the task was implemented. In contrast, Kessels et al. (2008) included the actual sequences for their block tapping test. Regardless, our quality appraisal findings suggest that the WMS-III visual tapping appeared to be the most appropriate choice when looking for a measure of visuo-spatial serial recall.

Three visuo-spatial STM tests did not require serial recall: Two involved the immediate recognition of complex designs via a pointing response (i.e., Helm-Estabrooks, 2001; Kalbe et al., 2005) and one involved the recall of designs via a drawing response (i.e., WMS-R Visual Reproduction I). Of the two involving immediate recognition of complex designs, the version by Kalbe and colleagues received a stronger validity appraisal; however, both of these tests received poor ratings in measurement error and across all types of reliability. Consequently, neither test would be appropriate for monitoring recovery or treatment effects. Compared to these recognition tests, the WMS-R Visual Reproduction I had stronger psychometric characteristics, despite concerns with certain types of validity and reliability. Among the eligible studies, this visuo-spatial STM test was used in only one study with one participant (i.e., Murray et al., 2006). It is possible that this test was used infrequently because drawing abilities in individuals with aphasia may be confounded by a number of concomitant conditions (e.g., dominant hand paresis; constructional apraxia; visual neglect; Murray & Clark, 2015).

The only standardised visuo-spatial WM test encountered in the eligible studies was the TEA Visual Elevator subtest, which evaluates updating submechanisms of WM (Kearney-Ramos et al., 2014). Our quality appraisal highlighted several psychometric concerns with this test including poor ratings of content and concurrent validity, measurement error, and test-retest and split-half reliability. Given that only one standardised test was identified, additional research is warranted to examine the visuo-spatial WM performance patterns of individuals with aphasia on other updating tests (e.g., *n*-back measures) as well as tests designed to evaluate shifting processes (e.g., complex span measures).

### Quality appraisal of studies

Our systematic review and quality appraisal identified only six studies with high (Chiou & Kennedy, 2009; Fucetola et al., 2009; Ivanova & Hallowell, 2014) or moderate (DeDe et al., 2014; Kalbe et al., 2005; Meteyard et al., 2015) overall study quality ratings, and thus revealed a number of concerns regarding the description, use, and interpretation of STM and WM tests in the aphasia literature (Table 12). Whereas study design was rated as high or

moderate in the vast majority of the papers, issues arose in terms of the other appraisal categories. Inadequate description of aphasia variables (i.e., low rating) was encountered in several studies. That is, in these studies, the presence of aphasia was mentioned but with nominal description and/or documentation of the aphasia profile (e.g., no information concerning aphasia severity). Failure to include aphasia profile information subverts determining to which segment of the aphasia population the STM/WM test(s) findings apply. Approximately half of the studies adequately described the language profiles of the participants with aphasia but included a restricted range of profiles; in some cases this was related to the small sample size (e.g., Francis et al., 2003) whereas in others, the study was designed to focus on a particular aphasia profile (e.g., Gvion & Friedmann, 2012). A restricted range of profiles limits the extent to which STM/WM test findings can be generalised to the broad aphasia population and may result in a lack of evidence for certain segments of that population. Indeed, individuals with severe aphasia or a Wernicke's aphasia type were under-represented in the studies reviewed.

With respect to the use and interpretation of the STM/WM tests, most studies failed to describe the assessment conditions, with only three studies specifying the characteristics of both the testing environment and the test administrator. Description of assessment variables is necessary to (1) allow replication of STM/WM test administration procedures not only in future research but also in clinical settings, and (2) aid in interpreting the test findings (e.g., different STM test scores at time point 1 and 2 could reflect administration differences versus a change in memory performance). Another major concern was the small number of investigations (i.e., 6 out of 36) in which age and education in concert with at least one other confounding factor were taken into account when administering and interpreting the STM/WM tests. Consideration of such factors is essential given the extensive literature documenting the substantial influence of demographic variables such as age, education, and ethnocultural background on cognitive test performances (e.g., Casaletto et al., 2015; Norman et al., 2011). Accordingly, STM/WM test outcomes become difficult to interpret when such factors have not been reported at all in a study or have been disregarded when scoring STM/WM tests or comparing patient and control groups. Relatedly, whereas most studies included the reference standard for the STM/WM test scores, close to 30% failed to do so. In these latter studies, whether the STM/WM test results indicate the presence or absence of impairment cannot be vetted.

## Recommendations

Based on our review of the standardised STM/WM tests and the studies utilising such tests, we recommend the following in future endeavours related to the evaluation of STM or WM in aphasia:

1. There is a need to obtain standardisation information from larger sample sizes to increase the power of STM/WM tests' psychometric properties. This would provide confidence to researchers and clinicians in adopting specific tests. With some notable exceptions (Kalbe et al., 2005; Swinburn et al., 2004), normative and validation sample sizes for individuals with aphasia were small. At the very least, age and education information must also be included in the normative and validation data, given the well-documented influence of such demographic

variables on cognitive test performance (e.g., Casaletto et al., 2015). Description of aphasia variables and testing environments is also recommended to allow determination of the range of patient profiles and administration settings in which the test can detect STM/WM difficulties or changes in STM/WM abilities.

2. There is a need to expand theoretical paradigms and study the psychometric properties of their tasks in both STM and WM. In auditory-verbal STM, there is a need to go beyond Digit Span and include tests that systematically manipulate word types and lexical variables (cf., Friedmann & Gvion, 2002). Such tests are needed to further delineate the role of phonological and semantic STM abilities in aphasia as well the role of item versus order deficits in STM (cf., Majerus et al., 2015; Martin, 2009). This issue has been addressed in experimental tasks that manipulate linguistic variables in verbal STM and WM tasks, but these experimental tasks have not yet undergone sufficient psychometric evaluation (e.g., Christensen & Wright, 2010; Martin, Kohen, & Kalinyak-Fliszar, 2010).
3. Several relatively new standardised cognitive test batteries have STM and WM subtests (e.g., WMS-IV Symbol Span; Wechsler, 2009), but have yet to be utilised in the aphasia literature (at least as of April, 2015).
4. Albeit one reviewed study solely used computerised STM tests (i.e., Lee & Pyun, 2014), expansion of computerised delivery of STM/WM tests appears an area in need of further exploration. Computerised tests afford timing precision, improve consistency of delivery, and minimise variability of presentation between different human assessors, ultimately improving testing (Noyes & Garland, 2008; Woods et al., 2011). However, practical limitations in terms of computer portability and availability could be prohibiting factors.
5. Investigations of staircase methods of presentation as opposed to the dominant “ascending” or “incremental” method of testing (i.e., from lists or blocks with fewer stimuli to lists or blocks with more stimuli) (cf., Ehrenstein & Ehrenstein, 1999) are needed. Although in the more traditional ascending testing method difficulty increases gradually, and thus possibly engages examinees in the testing process (because initial items are not too difficult), proactive interference also increases (May, Hasher, & Kane, 1999). May et al. showed that, particularly in older adults, the traditional ascending testing method can produce smaller WM scores because of increased proactive interference. Staircase methods could diminish such proactive interference. Computerised tests would allow automated adjustment of list presentations in terms of the staircase method, highlighting the need for more frequent collaboration between human computer interaction specialists and aphasiologists (Molero Martin, Laird, Hwang, & Salis, 2013; Salis & Hwang, 2016).
6. There should be more research on the possible influence of response modality in STM and WM testing (e.g., spoken response versus recognition; drawing versus recognition), in both non-brain-damaged adults as well as those with aphasia. For example, comparisons of matching span versus spoken recall tasks have revealed a distinction between (a) encoding and storage associated with language input

versus (b) retrieval associated with language output processes (e.g., Allport, 1984; Jacquemot, Dupoux, & Bachoud-Lévi, 2011; Romani, 1992). Nonetheless, issues of whether or not STM/WM tests that differ in response modality can be used interchangeably or should perhaps be used in concert to bolster the validity and reliability of STM/WM test results have not been systematically investigated.

7. As clinical researchers we recognise that research needs are different from clinical needs. To ensure that research findings make an impact on clinical practice, there should be more dialogue between stakeholders, that is, researchers, clinicians, and people with aphasia, to achieve the design of STM and WM tests that are psychometrically sound and discriminating, as well as appealing to clinicians who have limited time to derive a differential diagnosis before treatment or to measure improvements following treatment.

### Limitations

Some limitations must be acknowledged with respect to the current systematic review. First, only journal studies and test manuals in English were reviewed and appraised. Thus, selection bias is possible and our findings may not apply to STM/WM tests in other languages. Second, our ratings of the psychometric properties of tests were based on the sources available to us and the way those were reported. It may well be the case that if different statistical analyses were reported, the quality ratings might too have been different.

### Conclusions

The present systematic review identified use of a number of standardised auditory-verbal and visuo-spatial STM and WM tests in the contemporary aphasia literature. Further research is warranted, however, given problems in terms of these tests' validity, reliability, and measurement error, and in terms of how researchers documented their use and interpretation of such tests. That is, in concert with previous reviews (e.g., Salis et al., 2015), no gold standard for evaluating STM/WM abilities in people with aphasia was identified. Until such a gold standard STM/WM assessment tool has been ratified, reliance on just one test to characterise STM or WM in a given individual with aphasia appears an inadequate practice given (1) the psychometric concerns among the standardised tests currently being used in the aphasia literature, and (2) the multi-faceted nature of STM and WM specified in theoretical models of these memory constructs. Also, given the growing literature suggesting a crucial role for non-linguistic cognitive functions in mediating aphasia symptoms and outcomes (e.g., Nguyen et al., 2015; Nicholas, Hunsaker, & Guarino, 2015), our findings underscore the need not only to extend aphasia research regarding STM/WM standardised test development and validation, but also to review systematically standardised tests regularly being used in research and clinical practice to characterise other domains of cognitive functioning in individuals with aphasia.

### Acknowledgments

We would like to thank Helen Kelly (University College Cork, Ireland) for her advice and assistance with the literature searches and Paul Conroy (University of Manchester, UK) for providing us with information about the Rey Complex Figure Test.

### Funding

This work was supported in part by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under award number [R01DC013196]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

### References

- Abou El Ella M, Alloush T, El-Shobary A, El-Dien Hafez N, El-Halim A, El-Rouby I. Modification and standardization of Arabic version of the comprehensive aphasia test. *Aphasiology*. 2013; 27:599–614.
- Adrover-Roig D, Galparsoro-Izagirre N, Marcotte K, Ferré P, Wilson MA, Inéés Ansaldo A. Impaired L1 and executive control after left basal ganglia damage in a bilingual Basque-Spanish person with aphasia. *Clinical Linguistics & Phonetics*. 2011; 25:480–498. [PubMed: 21453016]
- Allen CM, Martin RC, Martin N. Relations between short-term memory deficits, semantic processing, and executive function. *Aphasiology*. 2012; 26(3–4):428–461. [PubMed: 22736889]
- Allport DA. Auditory-verbal short-term memory and conduction aphasia. In: Bouma H, Bouwhuis DG, editors *Attention and performance X: Control of language processes*. Hove, UK: Lawrence Erlbaum; 1984. 313–326.
- Anastasi A, Urbina S. *Psychological testing*. 7. Upper Saddle River, NJ: Prentice Hall; 2009.
- Ardila A, Concha M, Rosselli M. Angular gyrus syndrome revisited: Acalculia, finger agnosia, right-left disorientation and semantic aphasia. *Aphasiology*. 2000; 14:743–754.
- Attard MC, Rose ML, Lanyon L. The comparative effects of multi-modality aphasia therapy and constraint-induced aphasia therapy-plus for severe chronic Broca's aphasia: An in-depth pilot study. *Aphasiology*. 2013; 27:80–111.
- Baddeley A. Working memory: Theories, models, and controversies. *Annual Review of Psychology*. 2012; 63:1–29.
- Barbeau E, Didic M, Tramoni E, Felician O, Joubert S, Sontheimer A, ... Poncet M. Evaluation of visual recognition memory in MCI patients. *Neurology*. 2004; 62:1317–1322. [PubMed: 15111668]
- Basilakos A, Rorden C, Bonilha L, Moser D, Fridriksson J. Patterns of poststroke brain damage that predict speech production errors in apraxia of speech and aphasia dissociate. *Stroke*. 2015; 46:1561–1566. [PubMed: 25908457]
- Baxendale S. The Flynn effect and memory function. *Journal of Clinical and Experimental Neuropsychology*. 2010; 32:699–703. [PubMed: 20119877]
- Beeson PM, Rising K, Volk J. Writing treatment for severe aphasia: Who benefits? *Journal of Speech Language and Hearing Research*. 2003; 46:1038–1060.
- Berthier ML, Davila G, Garcia-Casares N, Green C, Juarez R, Ruiz-Cruces R, ... Barbancho MA. Atypical conduction aphasia and the right hemisphere: Cross-hemispheric plasticity of phonology in a developmentally dyslexic and dysgraphic patient with early left frontal damage. *Neurocase*. 2011; 17:93–111. [PubMed: 20818576]
- Biddle A, Watson L, Hooper C. Using evidence-based research in speech-language pathology. Presentation at the Academy of Neurologic Communication Disorders and Sciences Annual Meeting; Atlanta, GA. 2002.
- Biran M, Fisher S. Structured treatment can improve predicate argument structure impairment. *Aphasiology*. 2015; 29:29–56.
- Bose A, van Lieshout P. Effects of utterance length on lip kinematics in aphasia. *Brain and Language*. 2008; 106:4–14. [PubMed: 18440061]
- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, ... De Vet HC. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Clinical Chemistry and Laboratory Medicine*. 2003; 41:68–73. [PubMed: 12636052]
- Butler RA, Ralph MAL, Woollams AM. Capturing multidimensionality in stroke aphasia: Mapping principal behavioural components to neural structures. *Brain*. 2014; 137:3248–3266. [PubMed: 25348632]

- Campanella F, Crescentini C, Mussoni A, Skrap M. Refractory semantic access dysphasia resulting from resection of a left frontal Glioma. *Neurocase*. 2013; 19:27–35. [PubMed: 22519645]
- Cappelletti M, Freeman ED, Cipolotti L. Numbers and time doubly dissociate. *Neuropsychologia*. 2011; 49:3078–3092. [PubMed: 21807010]
- Carragher M, Sage K, Conroy P. The effects of verb retrieval therapy for people with non-fluent aphasia: Evidence from assessment tasks and conversation. *Neuropsychological Rehabilitation*. 2013; 23:846–887. [PubMed: 24028211]
- Casaletto KB, Umlauf A, Beaumont J, Gershon R, Slotkin J, Akshoomoff N, Heaton RK. Demographically corrected normative standards for the English version of the NIH toolbox cognition battery. *Journal of the International Neuropsychological Society*. 2015; 21:378–391. [PubMed: 26030001]
- Caspari I, Parkinson SR, LaPointe LL, Katz RC. Working memory and aphasia. *Brain and Cognition*. 1998; 37:205–223. [PubMed: 9665743]
- Caza N, Belleville S, Gilbert B. How loss of meaning with preservation of phonological word form affects immediate serial recall performance: A linguistic account. *Neurocase*. 2002; 8:255–273. [PubMed: 12221140]
- Centre for Reviews and Dissemination. *Systematic reviews: CRD's guidance for undertaking reviews in health care*. York: Centre for Reviews and Dissemination, University of York; 2008.
- Chiou HS, Kennedy MRT. Switching in adults with aphasia. *Aphasiology*. 2009; 23:1065–1075.
- Christensen SC, Wright HH. Verbal and non-verbal working memory in aphasia: What three n-back tasks reveal. *Aphasiology*. 2010; 24:752–762.
- Coelho C. Direct attention training as a treatment for reading impairment in mild aphasia. *Aphasiology*. 2005; 19:275–283.
- Conroy P, Sage K, Lambon Ralph MA. Errorless and errorful therapy for verb and noun naming in aphasia. *Aphasiology*. 2009; 23:1311–1337.
- Conway AR, Kane MJ, Bunting MF, Hambrick DZ, Wilhelm O, Engle RW. Working memory span tasks: A methodological review and user's guide. *Psychological Bulletin Review*. 2005; 12:769–786.
- Coolican H. *Research methods and statistics in psychology*. 6. Hove, UK: Psychology Press; 2014.
- Corsi PM. Unpublished PhD dissertation. McGill University; Canada: 1972. Memory and medial temporal region of the brain.
- Corsten S, Mende M, Cholewa J, Huber W. Treatment of input and output phonology in aphasia: A single case study. *Aphasiology*. 2007; 21:587–603.
- Cowan N. Multiple concurrent thoughts: the meaning and developmental neuropsychology of working memory. *Developmental Neuropsychology*. 2010; 35:447–474. [PubMed: 20721769]
- Crescentini C, Lunardelli A, Mussoni A, Zadini A, Shallice T. A left basal ganglia case of dynamic aphasia or impairment of extra-language cognitive processes? *Neurocase*. 2008; 14:184–203. [PubMed: 18569743]
- Crocket D, Clark C, Spreen S, Klonoff H. Severity of impairment or specific types of aphasia: An empirical investigation. *Cortex*. 1981; 17:83–95. [PubMed: 7273805]
- DeDe G, Ricca M, Knilans J, Trubl B. Construct validity and reliability of working memory tasks for people with aphasia. *Aphasiology*. 2014; 28:692–712.
- De Renzi E, Nichelli P. Verbal and non-verbal short-term memory impairment following hemispheric damage. *Cortex*. 1975; 11:341–354. [PubMed: 1222578]
- De-Torres I, Dávila G, Berthier ML, Walsh SF, Moreno-Torres I, Ruiz-Cruces R. Repeating with the right hemisphere: Reduced interactions between phonological and lexical-semantic systems in crossed aphasia? *Frontiers in Human Neuroscience*. 2013; 7doi: 10.3389/fnhum.2013.00675
- Dewar BK, Patterson K, Wilson BA, Graham KS. Re-acquisition of person knowledge in semantic memory disorders. *Neuropsychological Rehabilitation*. 2009; 19:383–421. [PubMed: 18982524]
- Dotan D, Friedmann N. Steps towards understanding the phonological output buffer and its role in the production of numbers, morphemes, and function words. *Cortex*. 2015; 63:317–351. [PubMed: 25460497]



- Ehrenstein WH, Ehrenstein A. Psychophysical methods. In: Windhorst U, Johansson H, editors *Modern techniques in neuroscience research*. Berlin: Springer; 1999. 1211–1241.
- Edwards S, Salis C, Meteyard L. Aphasia. In: Aronoff M, editor *Oxford bibliographies in linguistics*. 2. New York: Oxford University Press; 2015.
- Eling P. Kurt Goldstein's test battery. *Cortex*. 2015; 63:16–26. [PubMed: 25241395]
- Fastenau PS, Denburg NL, Hufford BJ. Adult norms for the Rey-Osterrieth Complex Figure test and for supplemental recognition and matching trials from the extended Complex Figure test. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D)*. 1999; 13:30–47.
- Fillingham JK, Sage K, Lambon Ralph MA. The treatment of anomia using errorless learning. *Neuropsychological Rehabilitation*. 2006; 16(2):129–154. [PubMed: 16565031]
- Flynn JR. The mean IQ of Americans: Massive gains 1932–1978. *Psychological Bulletin*. 1984; 95:29–51.
- Flynn JR. The WAIS-III and WAIS-IV: Daubert motions favor the certainly false over the approximately true. *Applied Neuropsychology*. 2009; 16:98–104. [PubMed: 19430991]
- Francis D, Clark N, Humphreys G. The treatment of an auditory working memory deficit and the implications for sentence comprehension abilities in mild receptive aphasia. *Aphasiology*. 2003; 17:723–750.
- Francis DR, Clark N, Humphreys GW. Circumlocution-induced naming (CIN): A treatment for effecting generalisation in anomia? *Aphasiology*. 2002; 16:243–259.
- Friedmann N, Gvion A. FriGvi: Friedmann Gvion battery for assessment of phonological working memory. Israel: Tel Aviv University; 2002.
- Friedmann N, Gvion A. As far as individuals with conduction aphasia understood these sentences were ungrammatical: Garden path in conduction aphasia. *Aphasiology*. 2007; 21:570–586.
- Fucetola R, Connor LT, Perry J, Leo P. Aphasia, severity, semantics, and depression predict functional communication in acquired aphasia. *Aphasiology*. 2006; 20:449–461.
- Fucetola R, Connor LT, Strube MJ, Corbetta M. Unravelling nonverbal cognitive performance in acquired aphasia. *Aphasiology*. 2009; 23:1418–1426.
- Galling MA, Goorah N, Berthier ML, Sage K. A clinical study of the combined use of bromocriptine and speech and language therapy in the treatment of a person with aphasia. *Aphasiology*. 2014; 28:171–187.
- Greenhalgh T. How to read a paper: Papers that report diagnostic or screening tests. *BMJ Clinical Research*. 2014; 315:540–543.
- Grindrod CM, Baum SR. Hemispheric contributions to lexical ambiguity resolution in a discourse context: Evidence from individuals with unilateral left and right hemisphere lesions. *Brain and Cognition*. 2005; 57:70–83. [PubMed: 15629218]
- Gualtieri CT, Johnson LG. Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs. *Archives of Clinical Neuropsychology*. 2006; 21:623–643. [PubMed: 17014981]
- Gvion A, Friedmann N. Does phonological working memory impairment affect sentence comprehension? A study of conduction aphasia. *Aphasiology*. 2012; 26:494–535.
- Harnish SM, Lundine J. Nonverbal working memory as a predictor of anomia treatment success. *American Journal of Speech-Language Pathology*. 2015; 24(4):S880–S894. [PubMed: 26383918]
- Härtig C, Markowitsch H-J, Neufeld H, Calabrese P, Deisinger K, Kessler J. *Wechsler memory scale – Revised Edition, German Edition*. Bern: Huber; 2000.
- Helm-Estabrooks N. *Cognitive linguistic quick test*. San Antonio, TX: The Psychological Corporation; 2001.
- Helm-Estabrooks N. Cognition and aphasia: A discussion and a study. *Journal of Communication Disorders*. 2002; 35:171–186. [PubMed: 12036150]
- Hendricks CT, Nicholas ML, Zipse L. Effects of phonological neighbourhood on the treatment of naming in aphasia. *Aphasiology*. 2014; 28:338–358.
- Hoffman P, Jefferies E, Haffey A, Littlejohns T, Lambon Ralph MA. Domain-specific control of semantic cognition: A dissociation within patients with semantic working memory deficits. *Aphasiology*. 2013; 27:740–764.

- Howard D, Franklin S. Memory without rehearsal. In: Vallar G, Shallice T, editors *Neuropsychological impairments of short-term memory*. Cambridge: Cambridge University Press; 1990. 287–318.
- Howard D, Nickels L. Separating input and output phonology: Semantic, phonological, and orthographic effects in short-term memory impairment. *Cognitive Neuropsychology*. 2005; 22:42–77. [PubMed: 21038240]
- Ivanova M, Hallowell B. A new modified listening span task to enhance validity of working memory assessment for people with and without aphasia. *Journal of Communication Disorders*. 2014; 52:78–98. [PubMed: 24986153]
- Ivanova MV, Dragoy OV, Kuptsova SV, Ulicheva AS, Laurinavichyute AK. The contribution of working memory to language comprehension: Differential effect of aphasia type. *Aphasiology*. 2015; 29:645–664.
- Jacquemot C, Dupoux E, Bachoud-Lévi AC. Is the word-length effect linked to subvocal rehearsal? *Cortex*. 2011; 47:484–493. [PubMed: 20804979]
- Jaeggi SM, Buschkuhl M, Perrig WJ, Meier B. The concurrent validity of the N-back task as a working memory measure. *Memory*. 2010; 18(4):394–412. [PubMed: 20408039]
- Kalbe E, Reinhold N, Brand M, Markowitsch HJ, Kessler J. A new test battery to assess aphasic disturbances and associated cognitive dysfunctions – German normative data on the aphasia check list. *Journal of Clinical and Experimental Neuropsychology*. 2005; 27:779–794. [PubMed: 16183613]
- Kambanaros M, Weekes BS. Phonological dysgraphia in bilingual aphasia: Evidence from a case study of Greek and English. *Aphasiology*. 2013; 27:59–79.
- Kane MJ, Conway ARA, Hambrick DZ, Engle RW. Variation in working memory capacity as variation in executive attention and control. In: Conway ARA, Jarrold C, Kane MJ, Miyake A, Towse JN, editors *Variation in working memory*. New York: Oxford University Press; 2007. 21–48.
- Kasselimis DS, Panagiotis GS, Economou A, Peppas C, Evdokimidis I, Potagas C. Are memory deficits dependent on the presence of aphasia in left brain damaged patients? *Neuropsychologia*. 2013; 51:1773–1776. [PubMed: 23770384]
- Kay J, Lesser R, Coltheart M. *Psycholinguistic assessments of language processing in Aphasia*. Hove, UK: Psychology Press; 1992.
- Kearney-Ramos TE, Fausett JS, Gess JL, Reno A, Peraza J, Kilts CD, James GA. Merging clinical neuropsychology and functional neuroimaging to evaluate the construct validity and neural network engagement of the *n*-back task. *Journal of the International Neuropsychological Society*. 2014; 20:736–750. [PubMed: 24963641]
- Kendall D, Conway T, Rosenbek J, Gonzalez-Rothi L. Case study of phonological rehabilitation of acquired phonologic alexia. *Aphasiology*. 2003; 17:1073–1095.
- Kendall DL, Nadeau SE, Conway T, Fuller RH, Riestra A, Gonzalez Rothi LJ. Treatability of different components of aphasia – Insights from a case study. *The Journal of Rehabilitation Research and Development*. 2006; 43:323–335. [PubMed: 17041818]
- Kessels RP, van den Berg E, Ruis C, Brands AM. The backward span of the Corsi block-tapping task and its association with the WAIS-III digit span. *Assessment*. 2008; 15:426–434. [PubMed: 18483192]
- Khan KS, Kunz R, Kleijnen J, Antes G. Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine*. 2003; 96:118–121. [PubMed: 12612111]
- Lang CJG, Quitz A. Verbal and nonverbal memory impairment in aphasia. *Journal of Neurology*. 2012; 259:1655–1661. [PubMed: 22258478]
- Larsen J, Baynes K, Swick D. Right hemisphere reading mechanisms in a global alexic patient. *Neuropsychologia*. 2004; 42:1459–1476. [PubMed: 15246284]
- Lavoie M, Routhier S, Légaré A, Macoir J. Treatment of verb anomia in aphasia: Efficacy of self-administered therapy using a smart tablet. *Neurocase*. 2016; 22(1):109–118. [PubMed: 26007615]
- Law SP, Wong W, Chiu KMY. Preserved reading aloud with semantic deficits: Evidence for a non-semantic lexical route for reading Chinese. *Neurocase*. 2005; 11:167–175. [PubMed: 16006339]
- Lee B, Pyun SB. Characteristics of cognitive impairment in patients with post-stroke aphasia. *Annals of Rehabilitation Medicine*. 2014; 38(6):759–765. [PubMed: 25566474]

- Lee JB, Sohlberg MM. Evaluation of attention training and metacognitive facilitation to improve reading comprehension in aphasia. *American Journal of Speech-Language Pathology*. 2013; 22:318–333.
- Lee TM, Yuen KS, Chan CC. Normative data for neuropsychological measures of fluency, attention, and memory measures for Hong Kong Chinese. *Journal of Clinical and Experimental Neuropsychology (Neuropsychology, Development and Cognition: Section A)*. 2002; 24:615–632.
- Leff AP, Schofield TM, Crinion JT, Seghier ML, Grogan A, Green DW, Price CJ. The left superior temporal Gyrus is a shared substrate for auditory short-term memory and speech comprehension: Evidence from 210 patients with stroke. *Brain*. 2009; 132:3401–3410. [PubMed: 19892765]
- Lehman MT, Tompkins CA. Reliability and validity of an auditory working memory measure: Data from elderly and right-hemisphere damaged adults. *Aphasiology*. 1998; 12:771–785.
- Léonard B, de Partz MP, Grandin C, Pillon A. Domain-specific reorganization of semantic processing after extensive damage to the left temporal lobe. *NeuroImage*. 2009; 45:572–586. [PubMed: 19100847]
- Léonard B, Pillon A, de Partz M. Reacquisition of semantic knowledge by errorless learning in a patient with a semantic deficit and anterograde amnesia. *Aphasiology*. 2008; 22:447–488.
- Lidzba K, Staudt M, Zieske F, Schwilling E, Ackermann H. Prestroke/poststroke fMRI in aphasia: Perilesional hemodynamic activation and language recovery. *Neurology*. 2012; 78:289–291. [PubMed: 22238411]
- Loring DW, Bauer RM. Testing the limits: Cautions and concerns regarding the new Wechsler IQ and Memory scales. *Neurology*. 2010; 74:685–690. [PubMed: 20177123]
- Lott SN, Sperling AJ, Watson NL, Friedman RB. Repetition priming in oral text reading: a therapeutic strategy for phonologic text alexia. *Aphasiology*. 2009; 23:659–675. [PubMed: 20664804]
- Logie RH. The functional organization and capacity limits of working memory. *Current Directions in Psychological Science*. 2011; 20:240–245.
- Majerus S, Attout L, Artielle M, Van der Kaa M. The heterogeneity of verbal short-term memory impairment in aphasia. *Neuropsychologia*. 2015; 77:165–176. [PubMed: 26275964]
- Martin N. The role of semantic processing in short-term memory and learning: Evidence from aphasia. In: Thorn A, Page M, editors *Interactions between short-term and long-term memory in the verbal domain*. Hove, UK: Psychology Press; 2009. 220–243.
- Martin N, Ayala J. Measurements of auditory-verbal STM span in aphasia: Effects of item, task, and lexical impairment. *Brain and Language*. 2004; 89:464–483. [PubMed: 15120538]
- Martin N, Gupta P. Exploring the relationship between word processing and verbal short-term memory: Evidence from association and dissociations. *Cognitive Neuropsychology*. 2004; 21:213–228. [PubMed: 21038201]
- Martin N, Kohen FP, Kalinyak-Fliszar M. A processing approach to the assessment of language and verbal short-term memory abilities in aphasia. Paper presented at the Clinical Aphasiology Conference; 2010.
- Martin N, Saffran EM, Dell GS. Recovery in deep dysphasia: Evidence for a relation between auditory-verbal STM capacity and lexical errors in repetition. *Brain and Language*. 1996; 52:83–113. [PubMed: 8741977]
- Martin RC, Allen CM. A disorder of executive function and its role in language processing. *Seminars in Speech and Language*. 2008; 29:201–210. [PubMed: 18720317]
- Martin RC, Lesch MF, Bartha MC. Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*. 1999; 41:3–29.
- May CP, Hasher L, Kane MJ. The role of interference in memory span. *Memory & Cognition*. 1999; 27:759–767. [PubMed: 10540805]
- Mayer J, Murray L. Approaches to the treatment of alexia in chronic aphasia. *Aphasiology*. 2002; 16:727–743.
- Mayer JF, Murray LL. Measuring working memory deficits in aphasia. *Journal of Communication Disorders*. 2012; 45(5):325–339. [PubMed: 22771135]
- McNeil MR, Prescott TE. *The revised token test*. Baltimore, MD: University Park Press; 1978.

- Meteyard L, Bruce C, Edmundson A, Oakhill J. Profiling text comprehension impairments in aphasia. *Aphasiology*. 2015; 29:1–28.
- Meyers JE, Meyers KR. *Rey complex figure test and recognition trial: Professional manual*. Odessa, FL: Psychological Assessment Resources; 1995.
- Milner B. Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*. 1971; 27:272–277. [PubMed: 4937273]
- Miyake A, Friedman NP. The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*. 2012; 21:8–14. [PubMed: 22773897]
- Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A. The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*. 2000; 41:49–100. [PubMed: 10945922]
- Moher D, Liberati A, Tetzlaff J, Altman DG. The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*. 2009; 6(6):e1000097. doi: 10.1371/journal.pmed1000097 [PubMed: 19621072]
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, ... De Vet HC. The COSMIN checklist manual. Amsterdam: VU University Medical Centre; 2009.
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, ... de Vet HC. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research*. 2010; 19:539–549. [PubMed: 20169472]
- Molero Martin D, Laird R, Hwang F, Salis C. Computerized short-term memory treatment for older adults with aphasia: Feedback from clinicians. *Proceedings of the 15th international ACM SIGACCESS Conference on Computers and Accessibility*; Bellevue, Washington. New York, NY: ACM; 2013. 44:1–44:2.
- Morris N, Jones DM. Memory updating in working memory: The role of the central executive. *British Journal of Psychology*. 1990; 81:111–121.
- Murray LL. Direct and indirect treatment approaches for addressing short-term or working memory deficits in aphasia. *Aphasiology*. 2012a; 26:317–337.
- Murray LL. Attention and other cognitive deficits in aphasia: Presence and relation to language and communication measures. *American Journal of Speech-Language Pathology*. 2012b; 21:51–64.
- Murray L, Ballard K, Karcher L. Linguistic specific treatment: Just for Broca’s aphasia? *Aphasiology*. 2004; 18:785–809.
- Murray LL, Clark HM. *Neurogenic disorders of language and cognition: Evidence-based clinical practice*. Austin, TX: Pro-Ed; 2015.
- Murray LL, Keeton RJ, Karcher L. Treating attention in mild aphasia: Evaluation of attention process training-II. *Journal of Communication Disorders*. 2006; 39:37–61. [PubMed: 16039661]
- Murray L, Timberlake A, Eberle R. Treatment of underlying forms in a discourse context. *Aphasiology*. 2007; 21:139–163.
- Myers PS. Toward a definition of RHD syndrome. *Aphasiology*. 2001; 15:913–918.
- Nicholas M, Hunsaker E, Guarino AJ. *Aphasiology*. 2015. The relation between language, non-verbal cognition and quality of life in people with aphasia; 1–14. ahead-of-print
- Nicholas M, Sinotte M, Helm-Estabrooks N. Using a computer to communicate: Effect of executive function impairments in people with severe aphasia. *Aphasiology*. 2005; 19:1052–1065.
- Nguyen VQ, PrvuBettger J, Guerrier T, Hirsch MA, Thomas JG, Pugh T, Rhoads C. Factors associated with discharge to home versus discharge to institutional care after inpatient stroke rehabilitation. *Archives of Physical Medicine and Rehabilitation*. 2015; 96:1297–303. [PubMed: 25823940]
- Norman MA, Moore DJ, Taylor M, Franklin D Jr, Cysique L, Ake C. HNRC Group. Demographically corrected norms for African Americans and caucasians on the hopkins verbal learning test-revised, brief visuospatial memory test-revised, Stroop color and Word test, and Wisconsin card sorting test 64-card version. *Journal of Clinical and Experimental Neuropsychology*. 2011; 33:793–804. [PubMed: 21547817]
- Noyes JM, Garland KJ. Computer- versus paper-based tasks: Are they equivalent? *Ergonomics*. 2008; 51:1352–1375. [PubMed: 18802819]

- Orsini A, Capitani E, Laiacona M, Papagno C, Vallar G. Verbal and spatial immediate memory span: Normative data from 1355 adults and 1112 children. *The Italian Journal of Neurological Sciences*. 1987; 8:537–548.
- Otsuka Y, Suzuki K, Fujii T, Miura R, Endo K, Kondo H, Yamadori A. Proper name anomia after left temporal subcortical hemorrhage. *Cortex*. 2005; 41:39–47. [PubMed: 15633705]
- Peterson LR, Peterson MJ. Short-term retention of individual verbal items. *Journal of Experimental Psychology*. 1959; 58:193–198. [PubMed: 14432252]
- Plaza M, Gatignol P, Leroy M, Duffau H. Speaking without Broca's area after tumor resection. *Neurocase*. 2009; 15:294–310. [PubMed: 19274574]
- Renvall R, Laine M, Laakso M, Nadine Martin N. Anomia treatment with contextual priming: A case study. *Aphasiology*. 2003; 17:305–328.
- Rey A. *L'examen clinique en psychologie*. Paris: Presses Universitaires de France; 1964.
- Richardson JTE. Measures of short-term memory: A historical review. *Cortex*. 2007; 43:635–650. [PubMed: 17715798]
- Robertson IH, Ward T, Ridgeway V, Nimmo-Smith I. *The Test of Everyday Attention (TEA)*. Bury St Edmunds, UK: Thames Valley Test Company; 1994.
- Robson H, Sage K, Lambon Ralph MA. Wernicke's aphasia reflects a combination of acoustic-phonological and semantic control deficits: A case-series comparison of Wernicke's aphasia, semantic dementia and semantic aphasia. *Neuropsychologia*. 2012; 50:266–275. [PubMed: 22178742]
- Romani C. Are there distinct input and output buffers? Evidence from an aphasic patient with an impaired output buffer. *Language and Cognitive Processes*. 1992; 7:131–162.
- Sage K, Snell C, Lambon Ralph MA. How intensive does anomia therapy for people with aphasia need to be? *Neuropsychological Rehabilitation*. 2011; 21(1):26–41. [PubMed: 21181603]
- Salis C. Short-term memory treatment: Patterns of learning and generalisation to sentence comprehension in a person with aphasia. *Neuropsychological Rehabilitation*. 2012; 22:428–448. [PubMed: 22300398]
- Salis C, Hwang F. Digital technology and aphasia. *Aphasiology*. 2016; 30:109–111.
- Salis C, Kelly H, Code C. Assessment and treatment of short-term and working memory impairments in stroke aphasia: A practical tutorial. *International Journal of Language and Communication Disorders*. 2015; 50:721–736. [PubMed: 26123369]
- Sarno MT. The nature of verbal impairment after closed head injury. *The Journal of Nervous and Mental Disease*. 1980; 168:685–692. [PubMed: 7441232]
- Schellig D. *Block-tapping test*. Lisse: Swets & Zeitlinger; 1997.
- Schlosser RW, Wendt O, Sigafos J. Not all systematic reviews are created equal: Considerations for appraisal. *Evidence-Based Communication Assessment and Intervention*. 2007; 1:138–150.
- Schuell H, Jenkins JJ, Jimenez-Pabon E. *Aphasia in adults: Diagnosis, prognosis and treatment*. London: Harper Row; 1964.
- Seniów J, Litwin M, Le niak M. The relationship between non-linguistic cognitive deficits and language recovery in patients with aphasia. *Journal of the Neurological Sciences*. 2009; 283:91–94. [PubMed: 19268973]
- Sidiropoulos K, Ackermann H, Wannke M, Hertrich I. Temporal processing capabilities in repetition conduction aphasia. *Brain and Cognition*. 2010; 73:194–202. [PubMed: 20621742]
- Sidiropoulos K, de Bleser R, Ablinger I, Ackermann H. The relationship between verbal and nonverbal auditory signal processing in conduction aphasia: Behavioral and anatomical evidence for common decoding mechanisms. *Neurocase*. 2015; 21:377–393. [PubMed: 24679121]
- Sidiropoulos K, de Bleser R, Ackermann H, Preilowski B. Pre-lexical disorders in repetition conduction aphasia. *Neuropsychologia*. 2008; 46:3225–3238. [PubMed: 18761023]
- Sierpowska J, Gabarrós A, Ripollés P, Juncadella M, Castañer S, Camins Á, ... Rodríguez-Fornells A. Intraoperative electrical stimulation of language switching in two bilingual patients. *Neuropsychologia*. 2013; 51:2882–2892. [PubMed: 24029376]
- Sinotte MP, Coelho CA. Attention training for reading impairment in mild aphasia: A follow-up study. *NeuroRehabilitation*. 2007; 22:303–310. [PubMed: 17971621]



- Solcà M, Di Pietro M, Schnider A, Leemann B. Impairment of semantic memory after basal forebrain and fornix lesion. *Neurocase*. 2015; 21:198–205. [PubMed: 24498851]
- Sulleman S, Kim E. Decision-making, cognition, and aphasia: Developing a foundation for future discussions and inquiry. *Aphasiology*. 2015; 29:1409–1425.
- Sung JE, McNeil MR, Pratt SR, Dickey MW, Hula WD, Szuminsky NJ, Doyle PJ. Verbal working memory and its relationship to sentence-level reading and listening comprehension in persons with aphasia. *Aphasiology*. 2009; 23:1040–1052.
- Swinburn K, Porter G, Howard D. *Comprehensive aphasia test*. Hove, UK: Psychology Press; 2004.
- Thompson HE, Jefferies E. Semantic control and modality: An input processing deficit in aphasia leading to deregulated semantic cognition in a single modality. *Neuropsychologia*. 2013; 51:1998–2015. [PubMed: 23851292]
- Tompkins CA, Bloise CGR, Timko ML, Baumgaertner A. Working memory and inference revision in brain damaged and normally aging adults. *Journal of Speech Language and Hearing Research*. 1994; 37:896–896.
- Turkstra LS, Coelho C, Ylvisaker M. The use of standardized tests for individuals with cognitive-communication disorders. *Seminars in Speech and Language*. 2005; 26:215–222. [PubMed: 16278794]
- Valle F, Cuetos F. *EPLA: Evaluación del Procesamiento Lingüísticos en la Afasia*. Hove, UK: Lawrence Erlbaum; 1995.
- de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine: A practical guide*. Cambridge: Cambridge University Press; 2011.
- Vukovic M, Vuksanovic J, Vukovic I. Comparison of the recovery patterns of language and cognitive functions in patients with post-traumatic language processing deficits and in patients with aphasia following a stroke. *Journal of Communication Disorders*. 2008; 41:531–552. [PubMed: 18571195]
- Wagner R, Torgesen J, Rashotte C. *Comprehensive test of phonological processing*. Austin, TX: Pro-Ed; 1999.
- Waters GS, Caplan D. The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers*. 2003; 35:550–564.
- Wechsler D. A standardized memory scale for clinical use. *The Journal of Psychology*. 1945; 19:87–95.
- Wechsler D. *Échelle d'Intelligence Ottawa-Wechsler*. Ottawa: Institut de psychologie de l'Université d'Ottawa; 1957.
- Wechsler D. *Échelle Clinique de Wechsler*. Paris: Les Éditions de Psychologie Appliquée; 1969.
- Wechsler D. *Wechsler adult intelligence scale – Revised*. San Antonio, TX: Psychological Corporation; 1981.
- Wechsler D. *Wechsler memory scale – Revised*. San Antonio, TX: Psychological Corporation; 1987.
- Wechsler D. *Wechsler memory scale – III*. San Antonio, TX: Psychological Corporation; 1997.
- Wechsler D. *WAIS-III: Manuel de l' échelle d' intelligence de Wechsler pour adultes*. Paris: Centre de Psychologie Appliquée; 2000.
- Wechsler D. *Wechsler memory scale – IV*. San Antonio, TX: Pearson; 2009.
- Wicherts JM, Dolan CV, Hessen DJ, Oosterveld P, van Baal GCM, Boomsma DI, Span MM. Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence*. 2004; 32:509–537.
- Woods D, Kishiyama MM, Yund EW, Herron TJ, Edwards B, Poliva O, Hink RF, Reed B. Improving digit span assessment of short-term verbal memory. *Journal of Clinical and Experimental Neuropsychology*. 2011; 33:101–111. [PubMed: 20680884]
- Wright HH, Fergadiotis G. Conceptualizing and measuring working memory and its relationship to aphasia. *Aphasiology*. 2012; 26:258–278. [PubMed: 22639480]



## Appendix 1. Adapted Study Quality Rating Tool

| Quality Categories              | Ratings  |  |  |
|---------------------------------|--|--|--|
|                                 | <i>High</i>  | <i>Moderate</i>  | <i>Low</i>   |
| Design                          | Single-subject across participants; relatively large group (i.e., >10)   | Single-subject 1 participant; small group (i.e., <10)  | Case study   |
| Control for confounding factors | Adjustment for at least 3 confounding factors (e.g., ethnocultural background, gender), including age and education  | Adjustment for at least age and education  | Adjustment for 1 or 0 confounding factors  |
| Aphasia variables               | Specification of aphasia severity and description of language profile; range of aphasia profiles included  | Specification of aphasia severity and description of language profile; restricted range of aphasia profiles included (e.g., only mild aphasia) | Specification of presence of aphasia but limited description of language profile |
| Assessment variables            | Specification of assessor qualifications AND assessment conditions (e.g., same assessor across testing sessions; tested in quiet room) sufficient to allow replication | Specification of assessor qualifications OR assessment conditions sufficient to allow replication  | No specification of assessment variables   |
| STM/WM test interpretation      | Reference standard for the STM/WM test score(s) specified (e.g., compared to appropriate control group; utilised standard scores)                                      | Reference standard for the STM/WM test score(s) specified  | No specification of reference standard   |

This Study Quality Rating Tool is based on information in NIHR York University Guidelines and Criteria for Appraising Diagnostic Test Studies; Khan et al. (2003), STARD and COSMIN checklists. A study must score high in 4 out of 5 categories for an overall High rating (with no low rating); an overall moderate rating for a study cannot include any low rating.

## Appendix 2. Test Psychometric Properties Quality Rating Tool

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

### A. Validity

- Construct validity – statistical analysis (e.g., item analysis; factor analysis; correlating with performance of other STM/WM tests; sensitive to changes with recovery; can discriminate those with and without STM/WM deficit) has indicated that STM/WM test data match the intended structure of the test

Excellent = If there is a statistical correlation/regression of any sort, even if simple, between target instrument and other instruments

Fair = If there is no statistical analysis but just a discussion

Low = If there is no discussion whatsoever
- Content/Face validity – looks like it should assess STM/WM; degree to which test is model-based; has items with graded difficulty; possible systematic sources

of bias have been examined (e.g., cultural); experts designed the test; theoretical/empirical foundation of test and characteristics of items/subtests specified

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

Excellent = Assessed if all 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 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 (  $\geq 10$  )

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

- 3.** Criterion-related validity: concurrent (STM/WM test outcomes have been found consistent with other valid measures of memory; have scores been found significantly related to other indices of STM/WM 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

4. Criterion related validity: predictive validity

The test should predict performance on other measures/contexts to which the results will be generalised; does STM/WM test predict performance on other measures beyond the construct of STM/WM?

YES NO

5. Discriminant validity

The STM/WM test has been shown to discriminate those with and without typical STM/WM abilities

YES NO

## B. Reliability

1. Test-retest reliability information, e.g., length between testing (e.g., days, months)

- 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 2 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 test conditions similar for both measurements? (e.g., type of administration, environment, instructions)

Excellent = Test conditions were similar (evidence provided)

Good = Assumable that test conditions were similar

Fair = Unclear if test conditions were similar

Poor = Test conditions were NOT similar OR no reported

- h.** for continuous scores: was an intraclass correlation coefficient (ICC) calculated?

Excellent = ICC calculated and model or formula of the ICC is described

Good = ICC calculated but model or formula of the ICC not described or not optimal

Fair = Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred (i.e., strict simple  $r =$  or  $> .90$ ; relaxed  $.80$ )

Poor = No ICC or Pearson or Spearman correlations calculated OR method not reported

- i. for dichotomous/nominal/ordinal scores: was kappa calculated?

Excellent = Kappa calculated and reported

Fair = other statistical analysis calculated and reported

Poor = Only percentage agreement calculated OR method not reported

2. Split-half reliability/internal consistency

- a. was the sample size included in the internal consistency analysis adequate?

Excellent = Adequate sample size ( $\geq 100$ )

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 (i.e., need to extract from manual or article)

- d. Split half procedure not applicable

3. Inter-rater reliability Yes \_\_\_\_% No not applicable

Strict criteria: simple correlation ( $r$ ) for 2 ratings = or  $> .90$ ;  $.80$  for Kappa

Relaxed criteria:  $r .80$ ; Kappa  $.70$

## C. 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

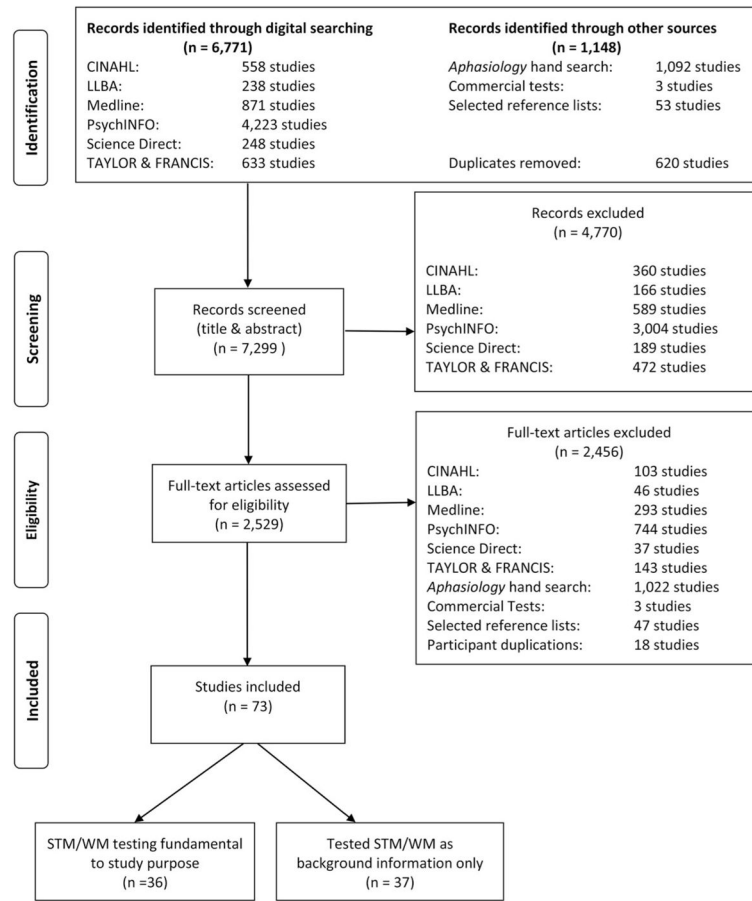
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript





**Figure 1.**  
Flow diagram of the identification-inclusion process.

**Table 1**

Search terms.

| <b>Construct related</b>  | <b>Population related</b>  | <b>Topic related</b>   |
|---|--|--|
| acoustic, active, attention, auditory, buffer, capacity, continuous performance, echoic, free, immediate, listening, memory, non-verbal, "non verbal", "nonverbal", phonological, primary, read <sup>a</sup> , recall, recognition, repetition, retention, sensory, serial, short-term "short term", semantic, spatial, tapping temporary, tonal, transient, verbal, visual, visuo-spatial, "visuospatial", working | acquired, adult <sup>a</sup> , aneurysm, aphasia, brain, cerebro-vascular, "cerebrovascular" cortical, CVA, dysphasia, head, h? emorrhage, injury, isch?emic, stroke, subcortical, traumatic, tumo?r, vascular | assess <sup>a</sup> , diagnos <sup>a</sup> , evaluation instrument, propert <sup>a</sup> , reliab <sup>a</sup> measure <sup>a</sup> , psychometric <sup>a</sup> , sensitivity, specificity, standard <sup>a</sup> , task <sup>a</sup> , test <sup>a</sup> , tool, valid <sup>a</sup> |

<sup>a</sup>Truncation of search term; internal character indicating UK spelling.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Summary of main studies (in alphabetical order).

| Study   | Focus of study   | Test   | Type of STM/WM assessed   | Participants with aphasia <sup>a</sup>              | Control participants <sup>a,b</sup>  |
|---|--|--|---|---|--|
| Abou El Ella et al. (2013)                                    | Modification and standardisation of the CAT in Arabic      | CAT Digit Span   | Auditory-verbal serial recall   | N = 100, age = 50, education from none to graduate  | N = 50, age = 45, education from none to graduate                                      |
| Allen et al. (2012)   | Links between STM, inhibition and semantics                | WAIS-R Digit Span  | Auditory-verbal serial recall   | N = 20, age = 63, ed = 15                           | N = 6, age = 69, ed = NR   |
| Butler, Ralph, and Woolliams (2014)                           | Neuroimaging of cognitive-linguistic processing            | WMS-R Digit Span   | Auditory-verbal serial recall   | N = 31, age = 63, ed = 12                           | N = 19, age = 68, ed = 13  |
| Caza, Belleville, and Gilbert (2002)                          | Semantic contribution to STM for words/non-words           | * WAIS Digit Span  | Auditory-verbal serial recall   | N = 1, age = 47, Masters educated                   | Not included   |
| Chiou and Kennedy (2009)                                      | Attention switching in aphasia                             | TEA Visual Elevator  | Visual WM   | N = 14, age = 64, ed = 15                           | N = 14, age = 66, ed = 16  |
| Coelho et al. (2005)  | Treatment study of attention                               | TEA Elevator Counting with Distraction; Visual Elevator  | Auditory and visual WM  | N = 1, age = 50, law school                         | Not included   |
| Crescentini, Lunardelli, Mussoni, Zadini, and Shallice (2008) | Subcortical language functions (in dynamic aphasia)        | * WAIS-R Digit Span  | Auditory-verbal serial recall   | N = 1, age = 67, ed = 8                             | Not included   |
| DeDe et al. (2014)  | Psychometric validation of several STM/WM tests            | Listening and Reading Spans; Picture Span; Square Span (forward, backward); N-back; Alphabet Span; Subtract-2 Span; WAIS-R Digit Span  | Auditory-verbal recall for words, non-words; sentence processing-word storage in WM; updating | N = 12, age = 64, ed = 14                           | N = 47: younger group n = 21, age = 21, ed = 14; older group n = 23, age = 65; ed = 14 |
| Fillingham, Sage, and Lambon Ralph (2006)                     | Errorless learning in anomia treatment                     | TEA Elevator Counting with Distraction   | Auditory WM   | N = 11, age = 68, ed = NR                           | Not included   |
| Francis, Clark, and Humphreys (2003)                          | Treatment of auditory-verbal STM                           | WMS-R Digit Span   | Auditory-verbal serial recall   | N = 1, age = 69, education information not provided | Not included   |
| Friedmann and Gvion (2007)                                    | Syntactic comprehension and STM in conduction aphasia      | FrGvi (Friedmann & Gvion, 2002) Word Span; Long Word Span; Similar Word Span; Non-word Span; Digit Span; Listening Span (recall and recognition probe test); Digit and Word Matching Spans | Auditory-verbal serial recall and recognition; auditory-verbal WM                             | N = 5, age = 56, ed = 12                            | N = 15, age = 54, ed = 12  |
| Fucetola, Connor, Perry, and Leo (2006)                       | Predictors of functional communication in aphasia recovery | WMS-III Visual Span  | Visuo-spatial serial recall   | N = 57, age = 58, ed = 14                           | Not included   |
| Fucetola, Connor, Strube, and Corbetta (2009)                 | Confirmatory factor analysis of some of                    | WMS-III Visual Span  | Visuo-spatial serial recall   | N = 136, age = 59, ed = 14                          | Not included   |

| Study  | Focus of study   | Test   | Type of STM/WM assessed  | Participants with aphasia <sup>a</sup>   | Control participants <sup>a,b</sup>   |
|--|--|--|--|--|---|
| Galling, Goorah, Berthier, and Sage (2014)                       | the WAIS-III and WMS-III nonverbal tasks in stroke aphasia<br>Impact of bromocriptine on the behaviour, cognition and linguistic skills of a person with aphasia | TEA Elevator Counting with Distraction   | Auditory WM  | N = 1, age = 58, ed = NR   | Not included  |
| Gvion and Friedmann (2012)                                       | Phonological STM in input and output conduction aphasia  | FriGvi, Word Span; Long Word Span; Similar Word Span; Non-word Span; Digit Span; Listening Span (recall and recognition probe test); Digit and Word Matching Spans | Auditory-verbal serial recall and recognition; auditory-verbal WM                    | N = 14, age = 52, ed = 13  | N = 269, range = 20–82 (only range reported), education at least 12 years   |
| Helm-Estabrooks (2002)   | Non-linguistic and linguistic cognitive skills in aphasia  | CLQT Design Memory   | Non-verbal visuo-spatial STM recognition; auditory-verbal serial recall; auditory WM | N = 13, age = 62, ed = 14  | Not included  |
| Hoffman, Jefferies, Haffey, Littlejohns, and Lambon Ralph (2013) | Semantic control and domain-general executive function in semantic aphasia   | WMS-R Digit Span; TEA Elevator Counting with Distraction   | Auditory-verbal WM   | N = 3, ages = 52, 54, 74, ed = left school at 15 (no other data provided)                    | Not included  |
| Howard and Nickels (2005)  | Input and output phonological stores in STM  | WAIS-R Digit Span  | Auditory-verbal serial recall  | N = 2, age = NR, ed = NR   | Not included  |
| Ivanova, Dragoy, Kupisova, Ulicheva, and Laurinavichyute (2015)  | Differential impact of WM impairments in individuals with fluent versus non-fluent aphasia types   | Eye-movement WM (Ivanova & Hallowell, 2014)  | Auditory-verbal WM   | N = 35, age = 54; n = 16 (non-fluent), age = 53, ed = 13; n = 19 (fluent), age = 55, ed = 13 | N = 36, age = 50, ed = 15   |
| Ivanova and Hallowell (2014)                                     | Validation of novel, eye-tracking auditory WM test   | Novel Eye-Movement WM  | Auditory-verbal WM   | N = 27, age = 56, ed = 5   | N = 33, age = 55, ed = 6  |
| Kalbe, Reinhold, Brand, Markowitsch, and Kessler (2005)          | Standardisation of the Aphasia Checklist   | Immediate Recognition of Geometric Figures   | Visual STM   | N = 154, age = 63, range of education abilities  | N = 106, age = 58, range of education abilities   |
| Kasselimis et al. (2013)   | Link between left-hemisphere, memory deficits and aphasia  | Block Tapping (Kessels, van den Berg, Ruis, & Brands, 2008)  | Auditory-verbal serial recall  | N = 49 (who could complete span test), age = 60, ed = 11                                     | N = 15, age = 58, ed = 10   |
| Lang and Quitz (2012)  | Repetition in conduction aphasia in relation to STM  | * WMS-R Digit and Visual Span  | Auditory-verbal and visuo-spatial serial recall                                      | N = 49, age = 68, ed = <9 years 80%, > 9 years 30% of the sample                             | Other non-aphasic left or right brain-damaged controls: N = 50, age = 66.58, ed = <9 years 80%, > 9 years 30% of the sample |

| Study   | Focus of study  | Test  | Type of STM/WM assessed   | Participants with aphasia <sup>a</sup>   | Control participants <sup>a,b</sup>   |
|---|---|---|---|--|---|
| Lee and Pyun (2014)                                     | Cognitive status in post-stroke aphasia   | Digit and Visual Span, Computerized Neurocognitive Test (MaxMedica, Seoul, Korea)   | Auditory-verbal serial recall, visuo-spatial serial recall                                  | N = 26, age = 54.7, ed = 10  | Other non-aphasic brain-damaged control: N = 68: n = 36 RHD, n = 32 LHD no aphasia, age = 60 RHD, 61 LHD, ed = 12 RHD, 10 LHD |
| Lee and Sohlberg (2013)                                 | Effect of attention training combined with metacognitive facilitation on reading comprehension in aphasia | TEA Elevator Counting with Distraction; Visual Elevator; Elevator Counting Reversal   | Auditory WM; visual WM; updating incoming information                                       | N = 4; n = 3 (anomic); n = 1 (conduction); n = 3 (mild); n = 1 (moderate), age = 71; ed = 17   | Not included  |
| Mayer and Murray (2002)                                 | Investigation of WM and reading treatment for individual with aphasia                                     | WMS-R Digit and Visual Span; TEA Visual Elevator, Elevator Counting with Distraction; Listening Span (Tompkins, Bloise, Timko, & Baumgaertner, 1994)                            | Auditory-verbal and visuo-spatial serial recall, auditory WM, auditory-verbal WM            | N = 1 (anomic), age = 62, ed = 18+   | Not included  |
| Meteyard, Bruce, Edmondson, and Oakhill (2015)          | Text reading in aphasia   | Pointing Span (Kay, Lesser, & Coltheart, 1992)  | Auditory-verbal serial recall by pointing   | N = 4; n = 2 (anomic), n = 1 (conduction), n = 1 (Broca's), age = 11, ed = 17  | N = 8, age = 62.6, ed = matched but no details provided   |
| Murray (2012b)  | Attention deficits and aphasia  | WMS-R Visual; TEA Elevator Counting with Distraction and Visual Elevator; Listening Span (Tompkins et al., 1994)  | Visuo-spatial serial recall, auditory and visual WM, auditory-verbal WM                     | N = 39; n = 15 (anomic), n = 8 (Broca's), n = 4 (TSA), n = 3 (TMA), n = 2 (Wernicke's), n = 3 (conduction), n = 2 (borderline fluent), n = 2 (mixed non-fluent), severity: n = 29 (mild); n = 18 (moderate), age = 60; ed = 15 | N = 39 healthy controls; age = 63; ed = 15  |
| Murray, Keeion, and Karcher (2006)                      | Attention processing training in mild aphasia   | WMS-R Digit and Visual Span, Visual Reproduction I; TEA Elevator Counting with Distraction; Visual Elevator; Elevator Counting Reversal; Listening Span (Tompkins et al., 1994) | Auditory-verbal and visuo-spatial serial recall, Auditory and visual WM, Auditory-verbal WM | N = 1 (mild conduction aphasia), age = 57; ed = university graduate  | Not included  |
| Nicholas, Sinotte, and Helm-Estabrooks (2005)           | Treatment study based on alternative communication  | CLQT Design Memory  | Non-verbal visuo-spatial STM recognition  | N = 5 non-fluent, age = 52, ed = 16  | Not included  |
| Sage, Snell, and Lamborn Ralph (2011)                   | Intensive and non-intensive therapy in the relearning of words in aphasia                                 | TEA Elevator Counting with Distraction  | Auditory-verbal WM  | N = 8, n = 5 fluent, n = 3 non-fluent, age = 61, education not reported  | Not included  |
| Salis (2012)  | STM training for STM and sentence comprehension   | WMS-R Digit Span and Visual Reproduction I; Token Test (McNeil & Prescott, 1978)  | Auditory-verbal and visuo-spatial serial recall; Auditory-verbal STM                        | N = 1, age = 73, university graduate   | Not included  |
| Sidiropoulos, de Bleser, Ablinger, and Ackermann (2015) | Verbal and nonverbal auditory signal  | WMS-R Digit Span  | Auditory-verbal serial recall   | N = 17, age = 59, education not reported   | N = 13 non-aphasic LHD, age = 59  |

| Study                                   | Focus of study   | Test   | Type of STM/WM assessed       | Participants with aphasia <sup>a</sup>         | Control participants <sup>a,b</sup> |
|---|--|--|-------------------------------|--|-------------------------------------|
| Sinotte and Coelho (2007)               | processing in conduction aphasia<br>Attention training to treat reading ability in mild aphasia                | TEA Elevator Counting with Distraction; Visual Elevator          | Auditory-verbal WM; Visual WM | <i>N</i> = 1, age = 60, education not reported | Not included                        |
| Sung et al. (2009)                      | WM and sentence comprehension in aphasia   | Listening Span (Tompkins et al., 1994)                           | Auditory-verbal WM            | <i>N</i> = 20, age = 63, ed = 15               | Not included                        |
| Vukovic, Vuksanovic, and Yukovic (2008) | Recovery of language and cognitive functions in post-traumatic language processing deficits and stroke aphasia | Rey Auditory Verbal Learning Test – immediate recall (Rey, 1964) | Auditory-verbal free recall   | <i>N</i> = 34, age = 47, ed = 12               | <i>N</i> = 37, age = 33, ed = 10    |

<sup>a</sup> mean age and education [ed] in years, rounded figures.

<sup>b</sup> control participants were adults with no brain damage, unless otherwise indicated;

RHD = right hemisphere damage; LHD = left hemisphere damage; *N* = total number of participants; *n* = number of participants in sub-samples; NR = not reported; TSA = transcortical sensory aphasia; TMA = transcortical motor aphasia;

\* = used non-English test version;

CAT = Comprehensive Aphasia Test; CLQT = Cognitive Linguistic Quick Test; TEA = Test of Everyday Attention; WAIS = Wechsler Adult Intelligence Scale; WAIS-R = Wechsler Adult Intelligence Scale – Revised; WAIS-III = Wechsler Adult Intelligence Scale – Revised; WMS = Wechsler Memory Scale; WMS-R = Wechsler Memory Scale – Revised; WMS-III = Wechsler Memory Scale 3rd Edition.



**Table 3**

STM/WM as background testing: standardised auditory-verbal STM tests (listed alphabetically by test type).

| Test types  | Task summary                                      | Studies and test publication (test, author, year)  |
|---|---|--|
| Digit Span – spoken recall                        | Serial forward and backward recall                | Adrover-Roig et al. (2011): WAIS-R (Wechsler, 1981) Ardila, Concha, and Rosselli (2000): WAIS-R (Wechsler, 1981) Berthier et al. (2011): WMS (Wechsler, 1945) Campanella, Crescentini, Mussioni, and Skrap (2013): Digit Span (Orsini, Capitani, Laiacona, Papagno, & Vallar, 1987) Corsten, Mende, Cholewa, and Huber (2007): WMS-R (Härting et al., 2000) De-Torres et al. (2013): EPLA (Valle & Cuetos, 1995) Dewar, Patterson, Wilson, and Graham (2009): WAIS-R (Wechsler, 1981) Dotan and Friedmann (2015): FriGvi (Friedmann & Gvion, 2002) Francis, Clark, and Humphreys (2002): WMS-R (Wechsler, 1987) Kendall, Conway, Rosenbek, and Gonzalez-Rothi (2003): CTOPP (Wagner, Torgesen, & Rashotte, 1999) Kendall et al. (2006): CTOPP (Wagner et al., 1999) Law, Wong, and Chiu (2005): WMS-R (Lee, Yuen, & Chan, 2002) Léonard, de Partz, Grandin, and Pillon (2009): WAIS-III (Wechsler, 2000) Léonard, Pillon, and de Partz (2008): WAIS-III (Wechsler, 2000) Lott, Sperling, Watson, and Friedman (2009): WMS-III (Wechsler, 1997) Otsuka et al. (2005): WMS-R (Wechsler, 1987) Plaza, Gatignol, Leroy, and Duffau (2009): WAIS-R (Wechsler, 1981) Renvall, Laine, Laakso, and Nadine Martin (2003): WMS-R (Wechsler, 1987) Sidiropoulos, de Bleser, Ackermann, and Preilowski (2008): WMS-R (Wechsler, 1987) Sidiropoulos, Ackermann, Wannke, and Hertrich (2010): WMS-R (Wechsler, 1987) Sierpowska et al. (2013): WAIS-III (Wechsler, 1997) Solcà, Di Pietro, Schneider, and Leemann (2015): WMS-III (Wechsler, 1997) |
| Digit Matching                                    | Span Serial recognition of spoken lists of digits | De-Torres et al. (2013): EPLA (Valle & Cuetos, 1995)   |
| CNS Vital Signs Memory Test – immediate condition | Recognition of words                              | Kambanaros and Weekes (2013): CNS Vital Signs Memory Test (Gualtieri & Johnson, 2006)  |
| Non-word Span                                     | Spoken serial recall of non-words                 | Dotan and Friedmann (2015): FriGvi (Friedmann & Gvion, 2002)   |
| Word Span – matching order                        | Serial recognition of spoken lists of words       | Biran and Fisher (2015): FriGvi (Friedmann & Gvion, 2002) Dotan and Friedmann (2015): FriGvi (Friedmann & Gvion, 2002)   |

WAIS-R = Wechsler Adult Intelligence Scale – Revised; WMS = Wechsler Memory Scale; WMS-R = Wechsler Memory Scale – Revised; CTOPP = Comprehensive Test of Phonological Processing; WAIS-III = Wechsler Adult Intelligence Scale 3rd Edition; EPLA = Evaluación del Procesamiento Lingüísticos en la Afasia.

**Table 4**

STM/WM as background testing: standardised auditory-verbal WM tests.

| <b>Test type</b>                   | <b>Task &amp; stimuli</b>                         | <b>Studies and test publication (test, author, year)</b>   |
|------------------------------------|---|--|
| Elevator Counting With Distraction | Filtering tones, selective attention and updating | Cappelletti, Freeman, and Cicolotti (2011): TEA (Robertson, Ward, Ridgeway, & Nimmo-Smith, 1994) Conroy, Sage, and Lambon Ralph (2009): TEA (Robertson et al., 1994) Robson, Sage, and Lambon Ralph (2012): TEA (Robertson et al., 1994) Thompson and Jefferies (2013): TEA (Robertson et al., 1994) |
| Listening Span – by spoken recall  | Sentence processing, word storage in WM           | Grindrod and Baum (2005): Listening Span (Tompkins et al., 1994) Murray, Timberlake, and Eberle (2007): Listening Span (Tompkins et al., 1994) Murray, Ballard, and Karcher (2004): Listening Span (Tompkins et al., 1994)   |

TEA = Test of Everyday Attention.

**Table 5**

STM/WM as background testing: standardised visuo-spatial STM and WM tests (listed alphabetically by test type).

| Test type   | Task & stimuli                            | Studies and test publication (test, author, year)   |
|---|---|---|
| CNS Vital Signs Memory Test – immediate condition | Recognition of geometric figures          | Kambanaros and Weekes (2013): CNS Vital Signs Memory Test (Gualtieri & Johnson, 2006)   |
| Design Recognition                                | Recognition of complex geometric figures  | Hendricks, Nicholas, and Zipse (2014): Design Memory of CLQT (Helm-Estabrooks, 2001)  |
| DMS-48 – immediate recognition                    | Recognition of visual objects             | Lavoie, Routhier, Légaré, and Macoir (2016): DMS-48 immediate recognition (Barbeau et al., 2004)  |
| Figure or Shape Drawing – immediate recall        | Drawing of complex figures                | Attard, Rose, and Lanyon (2013): Rey Complex Figure Test (Fastenau, Denburg, & Hufford, 1999) Larsen, Baynes, and Swick (2004): WMS-III Visual Reproduction (Wechsler, 1997)  |
| Visuo-spatial Span                                | Visual serial forward and backward recall | Beeson, Rising, and Volk (2003): WMS-R (Wechsler, 1987) Lidzba, Staudt, Zieske, Schwilling, and Ackermann (2012): Block Tapping Test (Schellig, 1997) Lott et al. (2009): WMS-III (Wechsler, 1997) Otsuka et al. (2005): WMS-R (Wechsler, 1987) Sidiropoulos et al. (2008): WMS-R (Wechsler, 1987) Sidiropoulos et al. (2010): WMS-R (Wechsler, 1987) |

CNS = central nervous system; DMS-48 = Delayed Matching-to-Sample-48; CLQT = Cognitive Linguistic Quick Test; WMS = Wechsler Memory Scale; WMS-R = Wechsler Memory Scale – Revised; WMS-III = Wechsler Memory Scale 3rd Edition.

Table 6

Standardised auditory-verbal STM tests, listed alphabetically by test type.

| Test types   | Task summary  | Studies and test publication (test, author, year)  | N (participants with aphasia with whom test was used) |
|--|---|--|---|
| Digit Span – spoken recall                           | Serial forward and backward recall  | Abou ElElla et al. (2013): CAT (Swinburn, Porter, & Howard, 2004) Allen et al. (2012): WAIS-R (Wechsler, 1981) Butler et al. (2014): WMS-R (Wechsler, 1987) Caza et al. (2002): Échelle Clinique de Wechsler (Wechsler, 1969) <sup>4</sup> ; Échelle d' Intelligence Ottawa-Wechsler (Wechsler, 1957) <sup>4</sup> Crescentini et al. (2008): WAIS-R (Wechsler, 1981) DeDe et al. (2014): WAIS-R (Wechsler, 1981) Francis et al. (2003): WMS-R (Wechsler, 1987) Hoffman et al. (2013): WMS-R (Wechsler, 1987) Howard and Nickels (2005): WAIS-R (Wechsler, 1981) Lang and Qutz (2012): WMS-R (Härtig et al., 2000) <sup>4</sup> Lee and Pyun (2014): Computerised Neurocognitive Test (MaxMedica, Seoul, Korea) <sup>4</sup> Mayer and Murray (2002): WMS-R (Wechsler, 1987) Murray et al. (2006): WMS-R (Wechsler, 1987) Salis (2012): WMS-R (Wechsler, 1987) Sidropoulos et al. (2015): WMS-R (Wechsler, 1987) | 272   |
| Digit Span -pointing                                 | Serial recognition of written digits, presented aurally                                   | Friedmann and Gvion (2007): FriGvi (Friedmann & Gvion, 2002) Gvion and Friedmann (2012): FriGvi (Friedmann & Gvion, 2002)  | 17  |
| Digit Span – matching order                          | Serial recognition of spoken lists of digits  | Friedmann and Gvion (2007): FriGvi (Friedmann & Gvion, 2002) Gvion and Friedmann (2012): FriGvi (Friedmann & Gvion, 2002)  | 17  |
| Pointing Span  | Forward and backward serial recall of spoken words by picture pointing                    | DeDe et al. (2014): Picture span (DeDe et al., 2014) Meteyard et al. (2015): Object-action word pointing (Kay et al., 1992)  | 16  |
| Rey Auditory Verbal Learning Test – Immediate recall | Immediate free recall of words  | Vukovic et al. (2008): Rey Auditory Verbal Learning Test (Rey, 1964)   | 34  |
| Sentence Comprehension                               | Comprehension of short vs. long sentences   | Salis (2012): Revised Token Test (McNeil & Prescott, 1978)   | 1   |
| Word Span – short words                              | Spoken serial recall of short words   | Friedmann and Gvion (2007): FriGvi (Friedmann & Gvion, 2002) Gvion and Friedmann (2012): FriGvi (Friedmann & Gvion, 2002)  | 17  |
| Word Span – long words                               | Spoken serial recall of long words  | Friedmann and Gvion (2007): FriGvi (Friedmann & Gvion, 2002) Gvion and Friedmann (2012): FriGvi (Friedmann & Gvion, 2002)  | 17  |
| Word Span – phonological similarity                  | Spoken serial recall of phonologically similar words                                      | Friedmann and Gvion (2007): FriGvi (Friedmann & Gvion, 2002) Gvion and Friedmann (2012): FriGvi (Friedmann & Gvion, 2002)  | 17  |
| Non-word Span  | Spoken serial recall of non-words   | Friedmann and Gvion (2007): FriGvi (Friedmann & Gvion, 2002) Gvion and Friedmann (2012): FriGvi (Friedmann & Gvion, 2002)  | 17  |
| Word Span -recognition                               | Serial recognition of written words, previously presented aurally                         | Friedmann and Gvion (2007): FriGvi (Friedmann & Gvion, 2002) Gvion and Friedmann (2012): FriGvi (Friedmann & Gvion, 2002)  | 17  |
| Word Span – probe                                    | Recognition of spoken words (non-serially) varying by frequency, imageability, word class | Friedmann and Gvion (2007): FriGvi (Friedmann & Gvion, 2002) Gvion and Friedmann (2012): FriGvi (Friedmann & Gvion, 2002)  | 17  |
| Word Span – matching order                           | Serial recognition of spoken lists of digits  | Friedmann and Gvion (2007): FriGvi (Friedmann & Gvion, 2002) Gvion and Friedmann (2012): FriGvi (Friedmann & Gvion, 2002)  | 17  |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

These tests were not evaluated in terms of quality as they were unobtainable.

CAT = Comprehensive Aphasia Test; WAIS-R = Wechsler Adult Intelligence Scale – Revised; WMS-R = Wechsler Memory Scale – Revised.

**Table 7**

Standardised auditory-verbal and visual-verbal WM tests (listed alphabetically by test type).

| Test type                               | Task and stimuli  | Studies and test publication (test, author, year)   | N (participants with aphasia with whom test was used) |
|---|---|---|---|
| Elevator Counting With Distraction      | Filtering tones, selective attention and updating   | Coelho (2005); TEA (Robertson et al., 1994) Fillingham et al. (2006); TEA (Robertson et al., 1994) Galling et al. (2014); TEA (Robertson et al., 1994) Hoffman et al. (2013); TEA (Robertson et al., 1994) Lee and Sohlberg (2013); TEA (Robertson et al., 1994) Mayer and Murray (2002); TEA (Robertson et al., 1994) Murray et al. (2006); TEA (Robertson et al., 1994) Murray (2012b); TEA (Robertson et al., 1994) Sage et al. (2011); TEA (Robertson et al., 1994) Sinotte and Coelho (2007); TEA (Robertson et al., 1994) | 70  |
| Elevator Counting with Reversal         | Monitoring and updating   | Lee and Sohlberg (2013); TEA (Robertson et al., 1994) Murray et al. (2006); TEA (Robertson et al., 1994) Sinotte and Coelho (2007); TEA (Robertson et al., 1994)  | 6   |
| Eye Movement WM                         | Sentence processing, word storage in WM   | Ivanova and Hallowell (2014) in English: Eye Movement WM task (Ivanova & Hallowell, 2014) Ivanova et al. (2015) (in Russian): Eye Movement WM task (Ivanova & Hallowell, 2014)  | 67  |
| Listening Span – by spoken recall       | Sentence processing, word storage in WM Listening Span (DeDe et al., 2014; Tompkins et al., 1994) | DeDe et al. (2014): Listening Span (DeDe et al., 2014) Mayer and Murray (2002): Listening Span (Tompkins et al., 1994) Murray (2012b): Listening Span (Tompkins et al., 1994) Sung et al. (2009): Listening Span (Tompkins et al., 1994)  | 72  |
| Listening Span – by written recognition | Sentence processing, word storage in WM   | Friedmann and Gvion (2007): FriGvi (Friedmann & Gvion, 2002) Gvion and Friedmann (2012): FriGvi (Friedmann & Gvion, 2002)   | 17  |
| <i>n</i> -back                          | Monitoring and updating   | DeDe et al. (2014): 1-back, 2-back (DeDe et al., 2014)  | 12  |

TEA = Test of Everyday Attention.

**Table 8**

Standardised visuo-spatial STM and WM tests (listed alphabetically by test type).

| Test type                        | Task & stimuli  | Studies and publication (test, author, year)   | N (participants with aphasia with whom test was used) |
|----------------------------------|---|--|---|
| Shape Drawing – immediate recall | Immediate drawing recall of figures                             | Murray et al. (2006): Visual Reproduction (Wechsler, 1987)   | 1   |
| Design Recognition               | Immediate recognition of complex geometric figures              | Helm-Estabrooks (2002): Design Memory of CLQT (Helm-Estabrooks, 2001) Kalbe et al. (2005): Nonverbal Memory of Aphasia Check List (Kalbe et al., 2005) Nicholas et al. (2005): Design Memory of CLQT (Helm-Estabrooks, 2001)   | 172   |
| Visuo-spatial Span               | Visual serial forward and backward recall                       | DeDe et al. (2014): Square Span (DeDe et al., 2014) Fucetola et al. (2006): WMS-III (Wechsler, 1997) Fucetola et al. (2009): WMS-III (Wechsler, 1997) Kasselimis et al. (2013): Block Tapping (Kessels et al., 2008) Lang and Quitz (2012): WMS-R (Härting et al., 2000) <sup>a</sup> Lee and Pyun (2014): Computerised Neurocognitive Test (MaxMedica, Seoul, Korea) Mayer and Murray (2002): WMS-R (Wechsler, 1987) Murray et al. (2006): WMS-R (Wechsler, 1987) Murray (2012b): WMS-R (Wechsler, 1987) Salis (2012): WMS-R (Wechsler, 1987) | 365   |
| Visual Elevator                  | Working memory/ updating, attentional switching with pictograms | Chiou and Kennedy (2009): TEA (Robertson et al., 1994) Coelho et al. (2005): TEA (Robertson et al., 1994) Lee and Sohlberg (2013): TEA (Robertson et al., 1994) Mayer and Murray (2002): TEA (Robertson et al., 1994) Murray et al. (2006): TEA (Robertson et al., 1994) Murray (2012b): TEA (Robertson et al., 1994) Sinotte and Coelho (2007): TEA (Robertson et al., 1994)  | 61  |

<sup>a</sup>These tests were not evaluated in terms of quality as they were unobtainable.

CLQT = Cognitive Linguistic Quick Test; WMS-III = Wechsler Memory Scale 3rd Edition; WMS-R = Wechsler Memory Scale – Revised; TEA = Test of Everyday Attention.



**Table 9**

Evaluation of auditory-verbal STM tests, listed alphabetically by test type.

| Test                                       | Validity  |              |            |            | Reliability  |             |            | Measurement error |             |
|--|-----------|--------------|------------|------------|--------------|-------------|------------|-------------------|-------------|
|  | Construct | Content/Face | Concurrent | Predictive | Discriminant | Test-retest | Split-half |                   | Inter-rater |
| FriGvi – Digit Span pointing               | Fair      | Poor         | Poor       | No         | Yes          | Poor        | Poor       | No                | Poor        |
| FriGvi – Digit Span, matching              | Fair      | Poor         | Poor       | No         | Yes          | Poor        | Poor       | No                | Poor        |
| FriGvi – Word Span short words             | Fair      | Poor         | Poor       | No         | Yes          | Poor        | Poor       | No                | Poor        |
| FriGvi – Word Span long words              | Fair      | Poor         | Poor       | No         | Yes          | Poor        | Poor       | No                | Poor        |
| FriGvi – Word Span phonological similarity | Fair      | Poor         | Poor       | No         | Yes          | Poor        | Poor       | No                | Poor        |
| FriGvi – Non-word span                     | Fair      | Poor         | Poor       | No         | Yes          | Poor        | Poor       | No                | Poor        |
| FriGvi – Word Span recognition             | Fair      | Poor         | Poor       | No         | Yes          | Poor        | Poor       | No                | Poor        |
| FriGvi – Word Span – probe                 | Fair      | Poor         | Poor       | No         | Yes          | Poor        | Poor       | No                | Poor        |
| FriGvi – Word Span matching order          | Fair      | Poor         | Poor       | No         | Yes          | Poor        | Poor       | No                | Poor        |
| Picture Span (DeDe et al., 2014)           | Excellent | Poor         | Fair       | No         | Yes          | Poor        | Poor       | No                | Poor        |
| Picture Span (Kay et al., 1992)            | Fair      | Poor         | Poor       | No         | Yes          | Poor        | Poor       | No                | Poor        |
| Revised Token Test                         | Poor      | Poor         | Poor       | No         | No           | Poor        | Poor       | Yes               | Poor        |
| <sup>#</sup> CAT Digit Span                | Excellent | Poor         | Poor       | Yes        | Yes          | Fair        | Poor       | Yes               | Poor        |
| <sup>#</sup> WAIS-III Digit Span           | Excellent | Excellent    | Fair       | Yes        | Yes          | Fair        | Poor       | Yes               | Excellent   |
| <sup>#</sup> WAIS-R Digit Span             | Excellent | Poor         | Fair       | Yes        | Yes          | Fair        | Poor       | No                | Excellent   |
| <sup>#</sup> WMS-R Digit Span              | Excellent | Poor         | Poor       | No         | Yes          | Poor        | Fair       | No                | Excellent   |

<sup>#</sup> the same ratings apply for both forward and backward versions of these tests; Rey (1964) unobtainable.

CAT = Comprehensive Aphasia Test; WAIS-III = Wechsler Adult Intelligence Scale 3rd Edition; ; WMS-R = Wechsler Memory Scale – Revised.

**Table 10**  
 Evaluation of auditory-verbal and visual-verbal WM tests, listed alphabetically by test type.

| Test  | Validity  |              |            |            |              | Reliability |            |             | Measurement error |
|---|-----------|--------------|------------|------------|--------------|-------------|------------|-------------|-------------------|
|   | Construct | Content/Face | Concurrent | Predictive | Discriminant | Test-retest | Split-half | Inter-rater |                   |
| Auditory Elevator (TEA)                               | Excellent | Poor         | Poor       | No         | Yes          | Poor        | Poor       | No          | Poor              |
| Elevator Counting with Reversal (TEA)                 | Excellent | Poor         | Poor       | No         | Yes          | Poor        | Poor       | No          | Poor              |
| Eye Movement WM (English version)                     | Excellent | Poor         | Excellent  | No         | No           | Poor        | Poor       | No          | Poor              |
| Eye Movement WM (Russian version)                     | Fair      | Poor         | Poor       | No         | No           | Poor        | Poor       | No          | Poor              |
| FriGvi – Listening Span, written recognition          | Fair      | Poor         | Poor       | No         | Yes          | Poor        | Poor       | No          | Poor              |
| Listening Span, spoken recall (DeDe et al., 2014)     | Excellent | Poor         | Fair       | No         | Yes          | Poor        | Poor       | No          | Poor              |
| Listening Span, spoken recall (Tompkins et al., 1994) | Excellent | Poor         | Fair       | Yes        | Yes          | Poor        | Poor       | No          | Poor              |
| <i>n</i> -back 1-back (DeDe et al., 2014)             | Excellent | Poor         | Fair       | No         | No           | Poor        | Poor       | No          | Poor              |
| <i>n</i> -Back 2-back (DeDe et al., 2014)             | Excellent | Poor         | Fair       | No         | Yes          | Poor        | Poor       | No          | Poor              |

TEA = Test of Everyday Attention.

Table 11

Evaluation of standardised visuo-spatial STM and WM tests, listed alphabetically by test type.

| Test   | Validity  |              |            |            |              | Reliability |            |             | Measurement error |
|--|-----------|--------------|------------|------------|--------------|-------------|------------|-------------|-------------------|
|  | Construct | Content/Face | Concurrent | Predictive | Discriminant | Test-retest | Split-half | Inter-rater |                   |
| Block Tapping (Kessels et al., 2008)         | Excellent | Fair         | Excellent  | Yes        | Yes          | Poor        | Poor       | No          | Poor              |
| Design Memory (CLQT)                         | Excellent | Fair         | Poor       | No         | No           | Poor        | Poor       | No          | Poor              |
| Nonverbal memory (Aphasia Check List)        | Excellent | Poor         | Fair       | Yes        | Yes          | Poor        | Poor       | No          | Poor              |
| <sup>a</sup> Square Span (DeDe et al., 2014) | Excellent | Poor         | Fair       | No         | Yes          | Poor        | Poor       | No          | Poor              |
| Visual Elevator (TEA)                        | Excellent | Poor         | Poor       | No         | Yes          | Poor        | Poor       | No          | Poor              |
| Visual Reproduction (WMS-R)                  | Excellent | Poor         | Poor       | No         | Yes          | Excellent   | Poor       | Yes         | Excellent         |
| <sup>a</sup> Visual tapping (WMS-R)          | Excellent | Poor         | Poor       | No         | Yes          | Poor        | Excellent  | No          | Excellent         |
| <sup>a</sup> Visual tapping (WMS-III)        | Excellent | Excellent    | Fair       | Yes        | No           | Fair        | Fair       | Yes         | Excellent         |

<sup>a</sup> same ratings for both forward and backward recall. The visual span subtests of the Computerised Neurocognitive Test (MaxMedica, Seoul, Korea) were unobtainable.

CLQT = Cognitive Linguistic Quick Test; TEA = Test of Everyday Attention; WMS-R = Wechsler Memory Scale – Revised; WMS-III = Wechsler Memory Scale 3rd Edition.

Table 12

Study quality ratings: high; moderate low (see Appendix 2 for a description of these rating categories).

| Study                        | Design   | Control for confounds | Aphasia variables | Assessment variables | STM/WM score interpretation | Overall rating |
|------------------------------|----------|-----------------------|-------------------|----------------------|-----------------------------|----------------|
| Abou El Ella et al. (2013)   | High     | Low                   | Moderate          | Low                  | High                        | Low            |
| Allen et al. (2012)          | High     | Low                   | Low               | Low                  | High                        | Low            |
| Butler et al. (2014)         | High     | Moderate              | High              | Low                  | High                        | Low            |
| Caza et al. (2002)           | Low      | Low                   | Moderate          | Low                  | High                        | Low            |
| Chiou and Kennedy (2009)     | High     | Moderate              | High              | High                 | High                        | High           |
| Coelho et al. (2005)         | High     | Low                   | Moderate          | Low                  | Low                         | Low            |
| Crescentini et al. (2008)    | Low      | Low                   | Moderate          | Moderate             | High                        | Low            |
| DeDe et al. (2014)           | High     | Moderate              | High              | Moderate             | High                        | Moderate       |
| Fillingham et al. (2006)     | High     | Low                   | High              | Low                  | High                        | Low            |
| Francis et al. (2003)        | Low      | Low                   | Low               | Low                  | Low                         | Low            |
| Friedmann and Gvion (2007)   | Moderate | Low                   | Moderate          | Low                  | High                        | Low            |
| Fucetola et al. (2006)       | High     | Low                   | High              | High                 | High                        | Low            |
| Fucetola et al. (2009)       | High     | High                  | High              | High                 | High                        | High           |
| Galling et al. (2014)        | Low      | Low                   | Moderate          | Low                  | Low                         | Low            |
| Gvion and Friedmann (2012)   | High     | Low                   | Moderate          | Low                  | High                        | Low            |
| Helm-Estabrooks (2002)       | High     | Moderate              | Low               | Low                  | High                        | Low            |
| Hoffman et al. (2013)        | Moderate | Low                   | Moderate          | Low                  | Low                         | Low            |
| Howard and Nickels (2005)    | Low      | Low                   | Low               | Low                  | High                        | Low            |
| Ivanova et al. (2015)        | High     | High                  | High              | Moderate             | Low                         | Low            |
| Ivanova and Hallowell (2014) | High     | High                  | High              | Moderate             | High                        | High           |
| Kalbe et al. (2005)          | High     | Moderate              | High              | Moderate             | High                        | Moderate       |
| Kasselimis et al. (2013)     | High     | Moderate              | Low               | Low                  | Low                         | Low            |
| Lang and Quitz (2012)        | High     | Moderate              | Low               | Moderate             | Low                         | Low            |
| Lee and Pyun (2014)          | High     | High                  | Moderate          | Low                  | Low                         | Low            |
| Lee and Sohlberg (2013)      | High     | Low                   | Moderate          | Low                  | Low                         | Low            |
| Mayer and Murray (2002)      | Moderate | High                  | Moderate          | Low                  | High                        | Low            |
| Meteyard et al. (2015)       | Moderate | Moderate              | Moderate          | Moderate             | High                        | Moderate       |
| Murray (2012b)               | High     | High                  | High              | Low                  | High                        | Low            |

| Study                      | Design   | Control for confounds | Aphasia variables | Assessment variables | STM/WM score interpretation | Overall rating |
|----------------------------|----------|-----------------------|-------------------|----------------------|-----------------------------|----------------|
| Murray et al. (2006)       | Moderate | Low                   | Moderate          | Moderate             | High                        | Low            |
| Nicholas et al. (2005)     | High     | Low                   | Moderate          | Low                  | High                        | Low            |
| Sage et al. (2011)         | Moderate | Low                   | Low               | Low                  | High                        | Low            |
| Salis (2012)               | Moderate | Low                   | Moderate          | Low                  | High                        | Low            |
| Sidiropoulos et al. (2015) | High     | Moderate              | Low               | Low                  | High                        | Low            |
| Simotte and Coelho (2007)  | Moderate | Low                   | Moderate          | Low                  | High                        | Low            |
| Sung et al. (2009)         | High     | High                  | Low               | Low                  | Low                         | Low            |
| Vukovic et al. (2008)      | High     | Low                   | Moderate          | Moderate             | Low                         | Low            |