The methodological quality of short-term/working memory treatments in post-stroke aphasia: A systematic review

Lilla Zakariás, Helen Kelly, Christos Salis, Chris Code

Author note

Lilla Zakariás, Department of Linguistics, University of Potsdam, Potsdam, Germany; Helen Kelly, Speech and Hearing Sciences, University College Cork, Cork, Ireland; Christos Salis, Speech and Language Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom; Chris Code, Department of Psychology, University of Exeter, United Kingdom

Correspondence concerning this article should be addressed to Lilla Zakariás, Department of Linguistics, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany

E-mail: zakarias@uni-potsdam.de

Abstract

Purpose

The aim of this systematic review is to provide a critical overview of short-term memory (STM) and working memory (WM) treatments in stroke aphasia and to systematically evaluate the internal and external validity of STM/WM treatments.

Method

A systematic search was conducted in 2014 February and then updated in 2016 December using 13 electronic databases. We provided descriptive characteristics of the included studies, and assessed their methodological quality using the Risk of Bias in N-of-1 Trials (RoBiNT) quantitative scale, which was completed by two independent raters.

Results

The systematic search and inclusion/exclusion procedure yielded 17 single case or case-series studies with 37 participants for inclusion. Nine studies targeted auditory STM consisting of repetition and/or recognition tasks, whereas eight targeted attention and WM, such as attention process training (ATP) including *n*-back tasks with shapes and clock faces, and mental math tasks. In terms of their methodological quality, quality scores on the RoBiNT scale ranged from 4 to 17 (mean = 9.5) on a 0–30 scale, indicating high risk of bias in the reviewed studies. Effects of treatment were most frequently assessed on STM, WM, and spoken language comprehension. Transfer effects on communication and memory in activities of daily living were tested in only 5 studies.

Conclusions

Methodological limitations of the reviewed studies make it difficult, at present, to draw firm conclusions about the effects of STM/WM treatments in post-stroke aphasia. Further studies with more rigorous methodology and stronger experimental control are needed to determine the beneficial effects of this type of intervention. To understand the underlying mechanisms of STM/WM treatment effects and how they relate to language functioning, a careful choice of outcome measures and specific hypotheses about potential improvements on these measures are required. Future studies need to include outcome measures

of memory functioning in everyday life and psychosocial functioning more generally to demonstrate the ecological validity of STM and WM treatments.

Stroke often results in several, long-term cognitive disabilities affecting language, memory, attention and executive functioning. The focus of this paper is on stroke aphasia (i.e., the language deficits that affects spoken and written communication) as well as verbal short-term and working memory deficits that interact with aphasia, in particular on behavioral treatments for these two types of memory deficits. As a construct, short-term memory (STM) refers to the ability to temporarily maintain and retrieve information, usually in serial order (Baddeley, 2012; Engle, Tuholski, Laughlin, & Conway, 1999). Relatedly, working memory (WM) refers to a more complex cognitive construct than STM in that it goes beyond the temporary maintenance of information by also supporting its mental manipulation (Baddeley, 2012; Engle, 2002; Miyake et al., 2000). Manipulation in WM involves various processes, such as shifting attentional control between tasks or mental sets, updating and monitoring WM representations, inhibiting dominant or automatic responses, and resolving different types of interference (Friedman & Miyake, 2004; Miyake et al., 2000). Several theoretical accounts describing the relationship between STM and WM are reported in the literature (see for example Engle et al., 1999). In the present study, it is not our purpose to provide a review of these theoretical accounts. Here, we use both terms to make a distinction between the simple storage buffer (STM) (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002) and the more complex memory system maintaining information in the face of concurrent processing, distraction, and/or attention shifts (WM) (Baddeley, 2012; Engle et al., 1999; Miyake & Shah, 1999).

Neurologically, STM and WM are associated with various brain regions, including the frontal lobes, and in particular, the dorsolateral prefrontal cortex (D'Esposito et al., 1995; Miyake et al., 2000; Smith, Jonides, Marshuetz, & Koeppe, 1998), the left inferior frontal gyrus (Botvinick, Cohen, & Carter, 2004; Smith et al., 1998), the premotor and the supplementary motor cortex (Smith & Jonides, 1998), the anterior cingulate cortex (Botvinick et al., 2004; D'Esposito et al., 1995), as well as the parietal cortex (Smith & Jonides, 1998). Given the large overlap of these regions with regions supporting language functions (Fedorenko, Duncan, & Kanwisher, 2012; Geranmayeh, Wise, Mehta, & Leech, 2014), it is not surprising that people with aphasia often present with pervasive post-stroke STM/WM impairments (Martin & Ayala, 2004; Murray, Salis, Martin, & Dralle, 2018; Warrington & Shallice, 1969), even in cases where aphasia has resolved (e.g., Vallat et al., 2005). This raises the issue of whether STM/WM deficits simply coincide with aphasia because of damage to shared anatomical representations.

Impairments of both verbal (as measured by, for example, serial forward or backward recall of verbal items, e.g., digit span task) and non-verbal STM/WM (as measured by serial forward or backward recall of figures or spatial locations, e.g., visuo-spatial span) can accompany aphasia. In addition, these can negatively influence individuals' language comprehension (Martin, Kohen, Kalinyak-Fliszar, Soveri, & Laine, 2012; Novick, Kan, Trueswell, & Thompson-Schill, 2009; Robinson, Blair, & Cipolotti, 1998; Sung et al., 2009; Wright, Downey, Gravier, Love, & Shapiro, 2007), reading (Caspari, Parkinson, LaPointe, & Katz, 1998), functional communication (Frankel, Penn, & Ormond-Brown, 2007; Fridriksson, Nettles, Davis, Morrow, & Montgomery, 2006; Keil & Kaszniak, 2002; Luna, 2011; Penn, Frankel, Watermeyer, & Russell, 2010; Ramsberger, 2005), and spontaneous recovery and treatment outcomes (Seniów, Litwin, & Leśniak, 2009).

Treatment of STM/WM in aphasia: Rationale, processes, and generalization to language

Throughout this article, we generally collapse the terms verbal and non-verbal STM and WM, and refer to these constructs as STM/WM, except where the constructs were required to be more specified or explicit (e.g., in the treatment and the outcome measures section in the Results).

STM/WM treatments in aphasia is a growing topic of interest, thanks to emerging evidence from two lines of research: First, studies in healthy populations highlight an overlap of language and STM/WM at cognitive and neural levels (e.g., Fedorenko et al., 2012; Geranmayeh et al., 2014), as noted above. Second, studies of people with aphasia provide evidence for a strong association between STM/WM functions and language performance (e.g., Novick et al., 2009; see also earlier work by Albert, 1976; Caramazza, Zurif, & Gardner, 1978). Studies of people with aphasia propose different underlying mechanisms to explain this association: Some suggest that verbal STM is very closely integrated with language (e.g., N. Martin & Gupta, 2004; N. Martin, Minkina, Kohen, & Kalinyak-Fliszar, 2018; N. Martin & Saffran, 1997), whereas others view STM/WM as a separate cognitive construct supporting language only under certain conditions (Caplan, Michaud, & Hufford, 2013; Haarmann, Just, & Carpenter, 1997; R. C. Martin, 2007; R. C. Martin & He, 2004; R. C. Martin & Romani, 1994; Thothathiri & Mauro, 2018; Wright et al., 2007).

The constructs of language and STM/WM have a particularly complex relationship in aphasia because the content of STM/WM, as Martin and Reilly (2012) put it, is often language. The important question is: how can one know that one is evaluating memory separately from language impairment? As De Renzi and Nichelli (1975) remark, in some cases spoken expression or comprehension deficits may be so severe as to make performance in STM/WM tasks difficult to interpret. However, these and other more recent studies (e.g., DeDe, Ricca, Knilans, & Trubl, 2014) have shown that people with aphasia are impaired in STM/WM tasks that minimize spoken output demands (e.g., response by pointing as opposed to spoken repetition). Nevertheless, such tasks do place demands on spoken comprehension and may be mediated by language, thus one can argue that phonological and semantic comprehension impairments interfere with such tasks. However, results of participants with aphasia showing that difficulties arise only when the number of words to be retained (that is, memory load) increases suggest that in some cases primarily STM/WM is impaired. For example, Waters, Caplan and Hildebrandt (1991) showed that although their participant was 100% correct when s/he had to recall by pointing to two-word lists, their performance decreased to 80% and 40% in three- and four-word lists, respectively. Similar findings have been reported for WM tasks (e.g., Gvion & Friedmann, 2012). Despite such findings, the issue of task impurity (cf. Miyake & Friedman, 2012), whereby a particular task which is thought to assess a particular construct (STM/WM in the context of this review) is likely to engage several other constructs simultaneously. Task impurity, therefore, could be conceived as a major weakness not only in this literature domain but also others. In summary, impairments of STM and/or WM in aphasia can be defined and also identified by below age- and education-appropriate performance in a STM and/or WM test, provided the person with aphasia has understood the demands of the test and is able to cope with the speech or other test demands (Salis, Kelly, & Code, 2015).

Nevertheless, different lines of, primarily observational (as listed above) studies galvanized a promising hypothesis: namely, that improvements in STM/WM would lead to improvements in language abilities that critically depend on STM/WM functions in aphasia, such as spoken sentence comprehension (Caplan et al., 2013), naming (Martin & Saffran, 1997), functional communication (Fridriksson et al., 2006), and reading (Caspari et al., 1998). This hypothesis is particularly pertinent in rehabilitation research because it relates directly to the concept of generalization (Webster, Whitworth, & Morris, 2015), or transfer of skills (Klingberg, Forssberg, & Westerberg, 2002), according to which treatment will enhance not only the targeted skill but also performance on similar skills or even language abilities (e.g., sentence comprehension, naming).

In the first reported STM treatment study for a person with aphasia, Peach (1987) delivered a treatment using repetition and pointing span tasks with words (a STM task) to a participant presenting with moderate conduction aphasia. Peach was primarily interested in whether the treatment would improve the participant's ability to repeat sentences, and the author concluded that this had been the case based on visual inspection of the data. More recently, a series of novel STM/WM treatments have been conducted, using various protocols and involving individuals with a range of aphasia types and severities, as well as different cognitive-linguistic profiles (e.g., Francis, Clark, & Humphreys, 2003; Harris, Olson, & Humphreys, 2014; Mayer & Murray, 2002). Treatment effects on further domains of language, for instance spoken sentence comprehension and functional communication, have also been examined (e.g., Murray, Keeton, & Karcher, 2006; Salis, 2012).

In terms of the theoretical motivations that underpin STM/WM treatment studies, there are two broad categories. The first category of studies has drawn primarily upon the general associative patterns between STM/WM and language abilities in aphasia that have been demonstrated in observational as well as experimental studies in the aphasiological literature. Examples of studies that fall into this category are as follows: Francis et al. (2003), Mayer and Murray (2002), Peach (1987), and Murray et al. (2006). The second category of studies is also based on this principle but, additionally, studies in this category are motivated by a particular theoretical account of STM/WM functioning. For example, the treatment study by Harris et al. (2014) was motivated by the theoretical distinction of STM between phonological and semantic abilities (Randi C. Martin & Allen, 2008). The treatment study reported by Martin and colleagues (Kalinyak-Fliszar, Kohen, & Martin, 2011) are based on the spreading activation model of word processing that utilizes activation-decay parameters to explain word and STM processing impairments (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997). Finally, the treatments reported by Vallat-Azouvi and colleagues (Vallat et al., 2005; Vallat-Azouvi, Pradat-Diehl, & Azouvi, 2014) were guided by the multicomponent model of WM (Baddeley, 2012). Depending on the theory they rely on, studies may also differ in their expectations in terms of treatment outcomes. For example, Harris et al. hypothesized that phonological STM treatment would selectively improve sentence repetition, while semantic STM treatment would result in selective improvements in spoken sentence comprehension (i.e., performance on an anomaly judgment task). Kalinyak-Fliszar et al. (2011) hypothesized that treatment of verbal STM and executive functions would result in improvements in repetition of words, non-words, and sentences, as well as spoken sentence comprehension. Vallat-Azouvi et al. (2014) in turn hypothesized distinct, domain-specific and domain-general effects depending on what WM component the treatment had focused on. In summary, theoretical support of STM/WM treatments is either based on findings from broader or more specific conceptualizations of STM/WM.

Motivation of the present study

To date, little is known about the evidence-base of STM/WM treatments in aphasia. While there have been recent narrative reviews (Majerus, 2017; Minkina, Rosenberg, Kalinyak-Fliszar, & Martin, 2017; Murray, 2012; Salis et al., 2015), to our knowledge, systematic reviews of STM/WM treatment studies in stroke aphasia have not been reported previously. Also, previous reviews have not focused on broad and systematic evaluation of the methodological quality of STM/WM treatments, which is the main focus of the present systematic review. Majerus (2017) focused on identifying the statistical properties of effect sizes of STM/WM treatments in terms of near and far transfer. The review by Minkina and colleagues (2017) focused on treatment studies with a particular interest in word retrieval deficits in aphasia. Murray (2012) provided a comparative review of STM and WM treatment in aphasia as well as other populations. Finally, the review by Salis et al. (2015) was purely descriptive in terms of STM/WM tasks involved in STM/WM treatment studies. Importantly, none of these reviews carried out systematic searches, with a priori inclusion and exclusion criteria. Furthermore, they did not appraise the evidence systematically using an appraisal tool with documented psychometric properties (RoBiNT; Tate et al., 2015).

Crucially, the present systematic review will enhance and broaden the evidence-base of treatments in stroke aphasia and help identify the best available evidence that could be adopted in clinical practice (Greenhalgh, 2014). This is particularly important because two surveys of stroke survivors' needs in the UK (McKevitt et al., 2011) and Australia (Andrew et al., 2014) found that stroke survivors themselves reported that general memory problems (rather than STM/WM specifically) after stroke were an unmet long-term rehabilitation need. Furthermore, there is growing acknowledgement that STM/WM deficits can influence rehabilitation decisions and outcomes (Balasooriya-Smeekens, Bateman, Mant, & De Simoni, 2016; Suleman & Kim, 2015); for example, returning to and staying in work (Balasooriya-Smeekens et al., 2016).

The question arises as to whether STM/WM treatments are indeed efficacious; that is, does treatment-induced STM/WM improvements *transfer* to unpracticed STM/WM tasks and aspects of language processing (e.g., spoken sentence comprehension, functional communication). In the experimental psychology and neuroscience literature, two forms of transfer have been distinguished: 1) near transfer, which refers to improvements on tasks that are similar to the treatment task but were not practiced during treatment (e.g., new STM/WM tasks) and 2) far transfer, which refers to improvements on skills or abilities that were not targeted during treatment but critically depend on STM/WM (e.g., spoken sentence comprehension, functional communication). We adopt the terms near and far transfer effects in the present study and aim to identify such effects after STM/WM treatments in stroke aphasia. Identifying these effects would help better understand: (i) the underlying cognitive mechanisms of transfer of STM/WM treatments; and (ii) the nature of the relationship between STM/WM and aspects of language processing in aphasia.

Consequently, the specific aims of the present study were: 1) To identify and describe STM/WM treatments in stroke aphasia through a systematic review of relevant literature; 2) to appraise the internal and external validity of these STM/WM treatments; 3) to investigate whether measures of STM/WM (near transfer), language (e.g., spoken sentence comprehension) and aspects of communication and memory in activities of daily living (far transfer) can improve following STM/WM treatments in stroke aphasia.

Methods

We prepared the present systematic review in accordance with the International Prospective Register of Ongoing Systematic Reviews (PROSPERO) statement (Booth et al., 2011, 2012; registration number: CRD42017052334).

Literature Search, Screening and Eligibility

We conducted a systematic search on the following electronic databases –Academic Search Complete, CINAHL FT, Education FT, Medline, Omnifile FT, PsyARTICLES, PsycINFO, Psychology & Behavioural Sciences Collection, Social Sciences FT, Cochrane Database of Systematic Reviews and Cochrane Central Register of Controlled Trials, PsycBITE™ and SpeechBITE™– from online inception to 2014 February by H.K., and then updated in 2016 (February 2014-December 2016) by two authors (H.K. and L.Z.)[1].

The search strategy comprised MeSH (Medical Subject Headings) terms and free text words focusing on three components (for details see Appendix): (i) population (aphasia or dysphasia), (ii) working memory/short-term memory (and related terms), (iii) rehabilitation (and related terms). In addition, reference lists of included studies, conference abstracts (Clinical Aphasiology Conference) and

[1] The systematic review initially commenced in 2014 February, however, due to competing workloads of the international collaborators it was put on hold for almost two years. Thus, when the authors recommenced the review in 2016 December, it was necessary to include a further search for potential new publications.

reviews that focused on the efficacy of aphasia rehabilitation (Brady, Kelly, Godwin, Enderby, & Campbell, 2016), cognitive rehabilitation, including STM/WM (Murray, 2012), and stroke rehabilitation outcome measures (Salter et al., 2005) were screened for potentially eligible studies.

Our selection criteria for inclusion in this review were as follows: 1) Participants were over 18 years old; 2) participants were described as presenting with non-progressive, acquired aphasia as a result of stroke, or had made a good or full recovery of stroke aphasia but continued to present with STM/WM and communication difficulties, for example, self-reported difficulties holding conversations (Vallat et al., 2005; Vallat-Azouvi et al., 2014) or mild written text comprehension difficulties (Vallat-Azouvi et al., 2014); 3) intervention protocol included treatment of STM/WM; 4) outcomes included STM/WM data; 5) in cases of mixed etiology groups, it had to be possible to identify the treatment outcomes for participants with post-stroke aphasia; 6) the study was published (or available from authors) in English. Studies that involved STM/WM training tasks or principles, for example, spaced retrieval training (Fridriksson, Holland, Beeson, & Morrow, 2005), or attention training (Coelho, 2005), but did not report STM/WM abilities, were excluded. Similarly, studies that reported etiologies other than stroke (e.g., trauma, Paek & Murray, 2015) were also excluded.

One of two review authors (H.K. or L.Z.) screened study titles and abstracts of the records identified through the electronic searches described above and excluded studies that clearly met one or more exclusion criteria. In cases where neither the title nor the abstract indicated clear eligibility, the full text was screened by two of three authors (H.K., C.S., L.Z.). We obtained hard copies of all remaining studies that fulfilled the listed inclusion criteria and two review authors (H.K., C.S. and/or L.Z.) independently assessed the studies based on the inclusion criteria and decided whether to include or exclude studies. Any disagreements were resolved through discussion with all authors, and the rationale for excluding any studies was recorded.

Data Extraction

For each study meeting the eligibility criteria, the following data were extracted: 1) Study aims; 2) study method information (design, randomization, blinding participants/clinicians/assessors, inter-rater reliability of clinicians and assessors, treatment fidelity and adherence); 3) participant characteristics (demographic, neurologic, cognitive-linguistic); 4) information on treatment procedure and setting (e.g., rationale for task selection, task procedure and stimuli, dosage of treatment, feedback provided, location where treatment was delivered, professional qualification of the clinician); 5) information on outcome measures (name or description of the assessment tool, number of sampling, analyses, results, qualification of the assessor). Two of three authors (H.K., C.S., and/or L.Z.), blinded to each other, performed data extraction independently for each study and any differences were resolved through discussion.

Appraisal of Methodological Quality

We used the Risk of Bias in N-of-1 Trials (RoBiNT; Tate et al., 2015) to rate the methodological quality of the included studies. The RoBiNT scale was designed to evaluate intervention studies using single case methodology (i.e., single case experimental designs and quasi single case experimental designs) and is also applicable to pre-post intervention designs (non-experimental single case designs) (Tate et al., 2015). To create comparability and consistency in the appraisal, we also used the RoBiNT scale for studies that do not strictly follow the single case methodology but share elements (e.g., case-control designs). The RoBiNT comprises 15 items covering both internal (n = 7) and external (n = 8) validity, with items scored on a 3-point scale (range 0–2). The total score ranges from 0–30. Internal validity of studies reflects the extent to which changes in the dependent variable (e.g., performance on the outcome measures) are attributable to introduction of the independent variable (STM/WM treatment) and not some other factors (e.g., spontaneous recovery, charm of the clinician). Internal validity is influenced by several features of the study, such as design, randomization, sampling of behavior, etc. (for details see Tate et al., 2013, 2015). External validity of studies reflects the extent to which a particular study's findings with a given sample can be extended to the population (e.g., people with aphasia, or people with aphasia with a certain cognitive-linguistic profile), and settings beyond the original study (e.g., everyday conversation). External validity is influenced by features, such as whether the study provides detailed information about the population and the setting where the treatment was delivered as well as whether the study was replicated on a different sample or setting, etc. (see Tate et al., 2013, 2015). Importantly, assessment of both types of validities largely depends on detailed information provided by the studies.

An advantage of the RoBiNT scale is that it has published psychometric properties, such as excellent inter-rater reliability (ICC = .93-.95, Tate et al., 2015) and good construct validity (Tate et al., 2015). Two of three authors (H.K., C.S., and L.Z.) independently rated each included study blinded to the other raters. Three included studies were authored by C.S. (Salis, 2012; Salis, Hwang, Howard, & Lallini, 2017) and L.Z. (Zakariás, Keresztes, Marton, & Wartenburger, 2018), however, in these cases the authors were not involved in rating their own study. Any disagreements were resolved through discussion with all authors.

Results

The electronic searches generated 552 studies (381 and 171 studies, first and second search, respectively). Twenty further studies were identified through screening reference lists of the included articles. Following the removal of duplicates, 346 studies were screened on title and abstract. Studies that clearly did not meet the inclusion criteria (n = 301) were excluded. The remaining 45 full text articles were independently assessed for eligibility by two authors (C.S., H.K., and/or L.Z.). Following full text selection, 17 studies were included for analysis (15 research articles and two abstracts in conference proceedings). The selection process is shown in Figure 1. Three authors were contacted to obtain further details from their studies. One study was not available in the English language and another study did not meet the inclusion criteria and were excluded. We were unable to trace one study. Studies included in the review are shown in Table 1.

Participants

Table 1 provides details of the 37 participants with post-stroke aphasia who took part in the selected studies (24 male, 13 female). Aphasia type was specified for 27 participants (in 14 of the 17 studies), not reported for 7 participants, and three participants were reported to have recovered from aphasia by the time of the study. Assessment data on aphasia severity was provided in 11 studies. Two studies reported severity without providing supporting data. Six participants were native English speakers (Kalinyak-Fliszar et al., 2011; Peach, Nathan, & Beck, 2017; Salis, 2012), six Korean (Eom & Sung, 2016), three Spanish (Berthier et al., 2014), three Hungarian (Zakariás, Keresztes, et al., 2018), and one German (Koenig-Bruhin & Studer-Eichenberger, 2007). For 18 participants (49%), native language was not reported (four studies with eight participants were conducted in a US or UK setting, another two studies with two participants used test batteries in French). The majority of participants (27/37) were beyond eight months post-stroke (chronic aphasia). Time post-stroke for nine participants was not reported and one was 11 days post-onset (Peach, 1987). Information about coexisting impairments of STM/WM was provided in 11 studies involving 16 participants. STM and WM impairments were reported in seven and five studies, respectively. Only one study specified the severity of WM impairment (Francis et al., 2003) (see Table 1).

Table 1 insert about here

Treatments

Table 2 describes the characteristics of treatments. The most common treatments (reported in 9 of the 17 studies) were auditory-verbal STM treatments, consisting of repetition and/or recognition tasks with words, non-words, word pairs, or sentences. Among these, treatment involved delayed repetition in three studies, with a gradually increasing delay between presentation and the participant's response over the course of therapy (1 sec vs. 5 sec, immediate vs. 5 sec vs. 10-12 sec, immediate vs. 5 sec; Kalinyak-Fliszar et al., 2011; Koenig-Bruhin & Studer-Eichenberger, 2007; Majerus et al., 2005, respectively). Recognition tasks were matching listening span (Salis, 2012; Salis et al., 2017) or pointing listening span tasks with words (Harris et al., 2014; Peach, 1987). Only one STM treatment included visual verbal tasks (Harris et al., 2014), namely recognition tasks with written words and non-words. The rest of the treatments (8/17) targeted WM (see Table 2), such as the attention process training–3 (ATP–3; Sohlberg & Mateer, 2010) including *n*-back tasks with shapes and clock faces, and mental math tasks, the repetition based treatment protocol (Eom & Sung, 2016) including sentence repetition, reconstruction, and reading tasks, and the adaptive *n*-back training with letters (Zakariás et al., 2018).

Memory treatments were contrasted with traditional language treatments in two studies (Berthier et al., 2014; Mayer & Murray, 2002). One study contrasted treatment outcomes for phonological vs. semantic STM (Harris et al., 2014). This theoretical distinction of STM was based on work by Martin and Allen (2008). The treatment reported by Vallat-Azouvi et al. (2014) compared treatment outcomes for the three components of WM (phonological loop, visuo-spatial sketchpad, and central executive), based on Baddeley (2012).

Table 2 indicates that the frequency of treatment administration varied greatly across studies (0.7-5 times per week, average frequency: 2.4 times per week), with the duration of treatment ranging from 1-19 months in total (mean of 23 weeks per study). Treatment dose ranged from 12 to 360 hours (mean = 64 hours per study).

Table 2 insert about here

Methodological Quality Appraisal of Included Studies: Internal and External Validity

Table 3 provides the RoBiNT scores for the scientific quality of each study. The total quality score ranged from 4 to 17 (mean = 9.5), indicating high risk of bias in the reviewed studies. Bias can lead to under- or overestimation of the observed effects (i.e., performance change in the target behavior, for example, STM/WM and language functions).

Internal validity. Internal validity was poor across the studies, with a score between 0-8 (mean = 1.6). One study employed a multiple baseline design across conditions (Kalinyak-Fliszar et al., 2011); nine studies employed a single-case methodology (as defined by Tate et al., 2015); seven studies employed pre-post intervention design (not considered a single case experimental design and therefore ineligible for points). Randomization of intervention phases was possible for two studies (e.g., Berthier et al., 2014; Harris et al., 2014); however, this was not implemented by the authors. Randomization of phase order (baseline vs. intervention) was not possible for the remaining studies because of study design (i.e., baseline measurement necessarily preceded intervention). Only six studies reported a sufficient sampling of behavior (at least three data points) across the study phases. For all interventions, neither clinicians nor participants were "blinded" to intervention. This is because it is practically difficult, if not impossible, to design procedures for this type of intervention in ways that neither the person administering, nor the person receiving the treatment is aware of the purpose of the treatment. Two studies reported an independent assessor for post-treatment outcome measurements who was not aware of relevant pre-treatment results (Salis, 2012; Salis et al., 2017); one study collected data on a computer, reducing risk of experimenter bias (Zakariás et al., 2018). High reliability of treatment data was evident only in three studies, with two studies reporting high inter-rater reliability (Kalinyak-Fliszar et al., 2011; Mayer & Murray, 2002), and one study using a computerized treatment automatically recording the participants' responses (Zakariás et al., 2018). Treatment adherence met necessary criteria only in three studies, with two studies providing sufficient measures of treatment adherence (Lee & Sohlberg, 2013; Salis et al., 2017), and one study delivering the treatment on a computer, which automatically yielded a maximum score on this item (Zakariás et al., 2018).

External validity. Compared to the internal validity scores, external validity scores were generally higher, ranging from 4-12 (mean = 7.9). Studies generally described and analyzed baseline characteristics of the participants well. Only seven studies reported the intervention environment, broadly describing the general location (mostly university clinic or participants' homes). The majority of studies (15/17) provided information about the target behavior, specifically defining the skill or ability that was being treated, and/or describing how the skill was measured. Only 11 of the 17 studies provided sufficient detail on the content and procedure of delivery of the intervention for later replication. Although 12 studies reported some type of statistical analysis, a justification for the suitability of the statistical procedure or the results of the analyses were not reported in most studies. Six studies replicated their findings, with four studies involving four or more participants (Eom & Sung, 2016; Lee & Sohlberg, 2013; Peach et al., 2017; Salis et al., 2017) and two studies involving two or three participants (Harris et al., 2014; Zakariás et al., 2018). Measures indicating whether effects of the intervention transferred to skills not targeted during treatment (e.g., spoken sentence comprehension) were included in 14 studies. However, external validity of these transfer effects was weakened by the fact that only two studies (Kalinyak-Fliszar et al., 2011; Lee & Sohlberg, 2013) collected data on these measures throughout all study phases.

Table 3 about here

Outcome Measures

Before we summarize the results concerning the effects of STM/WM treatments, we describe the outcome measures used to detect near and far transfer effects. Review of Table 4 indicates that serial recall (forward or backward) was the most frequently used outcome measure to assess near transfer effects of treatment on STM/WM (14/17). The majority of studies measured effects on auditory-verbal STM/WM (14/17). In contrast, only four studies measured treatment effects on the visuo-spatial domain (4/17 studies).

Among the auditory-verbal STM outcome measures (used in 13 out of the 17 studies), digit span forward was the most popular task (nine studies). Other common tests were word span forward and sentence repetition, both used in seven studies. The remaining auditory-verbal STM tasks were repetition or recognition tasks comprising single words/non-words, word/non-word pairs or triplets, letter or digit strings. Treatment effects on WM were specifically investigated in 11 studies, using backward or complex span tasks (6/17), recall or recognition tasks with interference (3/17), or a general cognitive assessment (the Test of Everyday Attention, TEA; Robertson, Ward, Ridgeway, & Nimmo-Smith, 1994) (3/17) with subtests that assess WM. In terms of response demands, the majority of studies (14/17) involved spoken output; tasks in six studies required either a pointing response or a recognition judgment (e.g., yes/no response). The interpretation of results on outcome measures entailing spoken output requires caution for participants with concomitant apraxia of speech (reported in four of the 17 studies), as well as for participants with other motor disorders that can affect speech output (e.g., dystonia reported in one study). Similarly, aphasia (and aphasia severity) in general can affect performance on the STM/WM outcome measures, requiring caution when considering beneficial effects of the treatments.

Effects of STM/WM treatment on standard language measures (i.e., far transfer effects) were investigated in the majority of studies (16/17). However, hypotheses on task-specific performance changes and the underlying mechanisms behind these changes were provided only in 12 studies. Transfer effects on auditory comprehension were specifically measured in 11 studies (11/17 studies). Among these, nine investigated effects on spoken sentence comprehension, mainly using sentence-picture matching tasks, and two studies investigated effects on spoken discourse comprehension (Murray et al., 2006; Peach et al., 2017). Transfer effects on reading comprehension were tested in two studies (2/17 studies) (Lee & Sohlberg, 2013; Murray et al., 2006). Language production, in particular spoken discourse was investigated in only two studies (2/17 studies), using picture-description tasks (Berthier et al., 2014; Koenig-Bruhin & Studer-Eichenberg, 2007). Change in aphasia severity (Aphasia Quotient of the Western Aphasia Battery, WAB; Kertész, 1982) due to treatment was assessed in two studies (Berthier et al., 2014; Eom & Sung, 2016). Change in general language profile, assessed by the WAB (Kim & Na, 2001), the Boston Diagnostic Aphasia Examination (BDAE; Goodglass & Kaplan, 1983), or the Aphasia Diagnostic Profiles (ADP; Helm-Estabrooks, 1992) was reported in three studies (Eom & Sung, 2016; Peach, 1987; Murray et al., 2006, respectively) (3/17).

With respect to the ecological validity, only five studies included measures of communication and memory in activities of daily living as reported by participants and/or their carers (5/17 studies). Note that treatment effects on daily living activities are also a form of far transfer effects according to our definition. Of these, four studies used a communication questionnaire. Two of these used a self-reported measure (Peach et al., 2017; Vallat et al., 2005) and two studies used both a self-reported and a spouse-reported measure (Murray et al., 2006; Salis et al., 2017). Two studies used a self-reported WM questionnaire (Vallat et al., 2005; Vallat-Azouvi et al., 2014).

The Effects of Treatments

Table 4 provides an overview of improvements in the outcome measures. Based on the adapted five-phase health care model of clinical outcome research in aphasia (Robey, 2004; Robey & Schultz, 1998), we classify all studies included in the present review as Phase 1 research. Phase 1 studies explore the application and the effects of a novel treatment for the first time and they make inferences about the effects of treatment applying at the level of an individual (or more individuals, but typically not at the level of a population)[2]. According to Robey and Schultz (1998), the terms efficacy and effectiveness can only be used at later stages of the clinical outcome research (i.e., Phase 2, or higher). Therefore, we use the term "treatment effects" when referring to possible benefits of an intervention.

Because of the methodological limitations of the reviewed studies (see the section on quality appraisal above), any conclusions regarding the beneficial effects of STM/WM treatments have to be viewed with caution. Hereinafter we summarize the effects of STM/WM treatments as reported by the authors, which was based on statistical or systematic visual analysis of data in 82% of the studies (14/17). The remaining three studies relied on visual inspection of the data, weakening the external validity of their findings.

[2] There is not always a one-to-one correspondence between features of the phases described by Robey and Schultz (1998) and features of the studies reviewed here. Some studies integrate features of more than one phase. For example, Peach et al. (2017) reported validity and reliability measures for the applied treatment (L-SAT), which is a typical feature of Phase 2 studies, but necessary background information about the included participants (i.e., definition of population), a core feature of Phase 2 studies is lacking (possibly due to space limitations). Similarly, although Salis et al. (2017) attempted to replicate positive effects of a previously used treatment (Salis, 2012) and they established a reliable and valid administration of the treatment, the lack of information regarding participants' selection does not allow for classification as a Phase 2 study.

Eighty-five percent and 82% of the studies that investigated effects on STM and WM, respectively, reported improvements in these domains. Similarly, seven (77%) of the nine studies that included outcome measures on spoken sentence comprehension reported improvements in these measures. Understanding of spoken paragraphs showed a small, subtle improvement in one study as measured by increased reaction times (Murray et al., 2006). The two studies that assessed effects on reading comprehension reported promising results: Higher accuracy and increased reading rate in paragraph reading tasks after treatment (Lee & Sohlberg, 2013; Mayer & Murray, 2002). Two studies reported nominal improvements in spoken discourse, using a picture description task (Berthier et al., 2014; Koenig-Bruhin & Studer-Eichenberger, 2007). However, among the four studies using a communication questionnaire, only one was successful in demonstrating improvements (Vallat et al., 2005). Results in measures of memory in daily living (i.e., WM questionnaires) varied across the two studies. One study reported improvements in WM (Vallat et al., 2005;), whereas the other did not report any changes in this domain (Vallat-Azouvi et al., 2014).

Table 4 about here

Discussion

The aim of this review was threefold: 1) To identify and describe STM/WM treatments in stroke aphasia; 2) to appraise the internal and external validity of these STM/WM treatments, and 3) to systematically review evidence concerning the effects of STM/WM treatments on measures of STM/WM (near transfer), language (e.g., spoken sentence comprehension) and aspects of communication and memory in activities of daily living (far transfer) in stroke aphasia. Seventeen studies were included and assigned quality scores using the RoBiNT scale (Tate et al., 2015), and their treatment effects were summarized. Because the methodological quality of studies significantly affects the interpretation of results concerning the treatments and their effects, we begin with discussing the methodological quality and the most important characteristics of the included studies.

Methodological Quality

Among the studies reviewed, only 53% (9/17) used a single-case methodology incorporating design features (e.g., multiple baseline measurements, inclusion of untreated control measures; Tate et al., 2015) that introduce experimental control and help mitigate or isolate sources of bias. However, in most of the studies (8/9) design features were not implemented appropriately (Tate et al., 2015), thus rendering any observed treatment effects questionable. Forty-one percent of the studies (7/17) described individuals in a pre-post intervention design. Such a design is typically used in group studies where experimental control is provided by a control group (Tate et al., 2015). Applying such a design to an individual seriously undermines internal validity because in the absence of the control group nothing serves as experimental control (Backman, Harris, Chisholm, & Monette, 1997; Tate et al., 2015). The remaining one study (Zakariás et al., 2018) applied a case-control design. This design allows for comparing each individual's data to a matched control group, and thus is strongly recommended for use with neuropsychological populations (Crawford & Garthwaite, 2005; Crawford, Garthwaite, & Porter, 2010). Note that the scoring method of RoBiNT scale does not acknowledge such case-control designs and thus its advantages are not reflected in the design score.

With respect to the sampling of behavior, Tate et al. (2015) recommend a minimum of five data points sampled in every phase (in the baseline as well as in the intervention phase) to establish stability in performance and control for variability within individuals (for alternative views see Howard, Best, & Nickels, 2015). While most of the studies in the present review did sample more than five times in the intervention phase, insufficient sampling during the baseline phase resulted in a score of 0 for 11 studies.

Overall, scorings on the RoBiNT scale revealed low internal validity for the studies reviewed. Strikingly, 65% of the included studies (11/17) scored zero on design, randomization, and sampling of behavior. Similarly, external validity scores, although higher than for internal validity, were still low. Low internal and external validity makes it difficult to draw conclusions about the beneficial effects of STM/WM treatments in stroke aphasia. A recent, related review by Majerus (2017) described the efficacy of STM/WM treatments using calculations of effect sizes for individual participants (Beeson & Robey,

2006) and a Bayesian one-sample t test to calculate overall effects across participants in the included studies. The conclusion drawn by Majerus, who acknowledged issues about specificity of treatment and content validity, was that STM/WM treatment studies appear to show satisfactory levels of efficacy. Without a proper appraisal of issues pertaining to the internal and external validity of studies, we are less certain about the effects of STM/WM to date because of the major issues our review has identified. Furthermore, the method used by Majerus to calculate effect sizes (based on Beeson and Robey, 2006), while generally accepted as appropriate, has been criticized as being susceptible to auto-correlation and therefore to Type I error (Howard et al., 2015). Also, it is not clear how Majerus calculated effect sizes in studies that implemented pre-post intervention designs, that is, studies with only one measurement before and after treatment (see Table 2). Beeson and Robey (2006) state that calculating effect sizes in such designs is problematic because these calculations require some variance in performance (i.e., a minimum of two, ideally three or more measurements) in the baseline phase. This issue does diminish the robustness of the statistical analyses Majerus reported.

Participants and Treatments

With respect to the participants included in the reviewed studies, 73% (N=27) of the individuals presented with mild or moderate aphasia. Because only 8% (three participants) had severe aphasia, the review cannot evaluate the impact of STM/WM treatments for people with severe aphasia. It is possible that those with severe aphasia (or severe language comprehension deficit) could benefit more from STM/WM treatments. This assumption is supported by a recent study (Zakariás, Salis, & Wartenburger, 2018), in which more severe spoken sentence comprehension deficits at the beginning of training were associated with larger improvements after training. Zakariás, Salis, et al. (2018) pooled the data of their WM treatment study and the data of Zakariás, Keresztes, et al.'s (2018) WM treatment study, and performed non-parametric correlational analyses (N=6) to investigate the potential relationship of initial WM and language comprehension ability (i.e., participants' WM and language comprehension ability when entering the studies) to the amount of improvements on spoken sentence comprehension after treatment. They found a negative relationship between initial spoken sentence comprehension ability and the amount of improvement on spoken sentence comprehension after treatment, suggesting that the more severe the spoken sentence comprehension deficit is at the beginning of treatment, the more it improves after treatment. It is possible that the language comprehension profile – e.g., lexical access and syntactic processing in spoken sentence comprehension – also affects STM/WM treatment outcomes. This hypothesis needs to be investigated in future STM/WM treatment studies.

Participants' cognitive profile (i.e., the presence and severity of attention, STM, and/or WM deficits) was considered in 13 studies when selecting participants for inclusion in the treatment. The rest of the studies did not report these abilities when describing their participants. Comprehensively assessing initial cognitive profiles may provide a more fine-grained evaluation of treatment effects, because it is possible that treatments of STM/WM produce better outcomes for people presenting with severe STM/WM deficit through providing extra computational resources or alternative routes for resolving processing conflicts encountered during language processing (Caplan et al., 2013; Fedorenko, 2014).

Approximately 60% of the treatments focused on auditory-verbal STM and used repetition or recognition tasks, mainly comprising words or sentences. The remaining treatments focused on attention as well as WM and were usually complex in terms of their structure (e.g., ATP-2; Sohlberg, Johnson, Raskin, & Mateer, 2001). Treatment dosage and intensity varied greatly across studies, ranging from 12 to 360 hours (mean = 64 hours) and 0.7-5 therapy sessions per week (mean = 2.4 sessions per week), respectively. Given suggestions in the treatment literature on aphasia (Bhogal, Teasell, & Speechley, 2003; Brady et al., 2016; Teasell et al., 2009) as well as literature of WM training in healthy populations (e.g., Jaeggi, Buschkuehl, Shah, & Jonides, 2014), one can hypothesize that treatment with high dose and great intensity (i.e., five times a week) may lead to better outcomes (but see Brady et al. 2016, reporting higher drop-out for groups receiving high-intensity compared to low-intensity aphasia therapy). This assumption, however, is not confirmed by our data due to large differences in treatment outcomes both across and within studies. Improvements after treatment were reported in both high and low intensity studies (e.g., Berthier et al., 2014 and Koenig-Bruhin & Studer-Eichenberger, 2007, respectively). The assumption that high intensity STM/WM treatments are more beneficial compared to low intensity treatments should be specifically tested in future studies.

The Effects of Treatments

Overall, participants seemed to improve in the treatment tasks, suggesting that STM/WM functions can be improved in stroke aphasia. Improvements were also noted on the STM/WM tasks that were not practiced during treatment in the majority of studies. Results of treatment on language and everyday functioning were highly variable. Around 75% of the studies that investigated effects on spoken sentence comprehension reported substantial improvements in this domain. Only a few studies investigated effects on spoken discourse and functional communication; improvements in this domain were reported in some but not all studies. Studies rarely included patient-reported outcome measures on memory in activities of daily living (e.g., questionnaires assessing everyday life problems related to deficits of STM/WM). When they did, results were inconclusive. As noted earlier, due to low methodological validity of the reviewed studies, any conclusions regarding the positive effects of treatments have to be viewed with caution. Related reviews (e.g., Murray, 2012; Salis et al., 2015) reached similar conclusions. As clinicians, we are very aware of principles of evidence-based practice (e.g., Greenhalgh, 2014) and the need for high-level evidence to support adoption of STM/WM treatments in clinical practice. As Murray (2012) concluded in her review, translation of preliminary findings to clinical practice must await further research aimed at delineating which individuals with aphasia might benefit most from STM/WM treatments. The findings from our review point to a similar conclusion.

Another question that arises is how the improvements (or absences of in some cases) that have been shown on STM/WM and language functioning can be reconciled in terms of the theoretical support that motivated these treatments, explicitly or implicitly. One issue that our review has highlighted systematically across all reported studies is that the methodological quality of the studies needs to be considered before upholding any hypothesis, be it loosely based on prior evidence or directly connected to a particular model of STM/WM (as discussed in the Introduction). Because of the methodological shortcomings in the reported STM/WM treatments, relatively small evidence-base, and limited replication, the current evidence-base of STM/WM neither supports, nor refutes the hypothesized relations between STM/WM treatments and language functioning (see also Nickels, Rapp, & Kohnen, 2015 for a discussion of challenges in the use of treatment to investigate cognitive functioning).

Limitations of the present systematic review

As we noted in the methods section, the RoBiNT scale was designed to evaluate studies with single case methodology (i.e., single case experimental designs and quasi single case experimental designs) and is also applicable to pre-post intervention designs (non-experimental single case studies) (Tate et al., 2015). To create comparability and consistency in the appraisal, we also used this scale for studies that did not strictly follow the single case methodology, but shared elements with this methodology. However, the use of the RoBiNT scale, which is particularly sensitive to features of single case methodology, may result in underrating case-control studies (Zakariás et al., 2018). Future studies may consider developing and validating scales or checklists for appraisal of differing designs. These attempts may consume considerable amount of resources as such development requires specific expertise complemented with great care and caution to make sure that checklists include all relevant and crucial information for a rigorous assessment.

Another potential limitation of our review relates to the fact that methodological properties (e.g., validity and reliability) are unknown for many outcome measures used to detect improvements after STM/WM treatments in the included studies. This information would be particularly useful in the case of studies where design features (related to experimental control) were not implemented in an appropriate way. In these studies established psychometric properties of outcome measures could help exclude confounding factors and attribute observed improvements to the treatment.

Conclusions and Recommendations for Future Research

This systematic review suggests that, currently, the evidence for the beneficial effects of STM/WM treatment in post-stroke aphasia is limited. Improvements in the treatment tasks are common, however, the results regarding improvements in the outcome measures of language and everyday functions (transfer effects) are mixed. Moreover, the validity of transfer effects is questionable in some studies. Further studies with rigorous methodology and larger sample sizes are needed to determine the positive effects of these interventions. Rigorous methodology should include the use of designs with strong experimental control, such as single-case experimental designs (e.g., multiple baseline design across conditions and/or participants) (Howard et al., 2015; Tate at al., 2015), case-control designs (Crawford & Garthwaite, 2005; Crawford et al., 2010), and randomized control trials (Kendall, 2003). Appropriate statistical analyses should be carefully selected to reduce statistical artifacts (e.g., regression toward the mean, autocorrelation) leading to potential misinterpretations of treatment effects (Howard et al., 2015). For all of the above-mentioned designs, appropriate statistical methods are available. The group of statistics called WEighted STatistics (WEST-ROC and WEST-COL) is recommended to use in single-case experimental designs (Howard et al., 2015), whereas the Regbuild (Crawford & Garthwaite, 2007) and the Revised Standardized Difference Test (RSDT; Crawford & Garthwaite, 2005), and the Wilcoxon signed-rank test (Field, 2009) can be used in case-control designs and randomized control trials, respectively. The WEST statistics control for performance changes (e.g., spontaneous recovery, practice effects, intra-individual variability) potentially occurring in the multiple-baseline phase. Specifically, these methods determine whether the amount of change in the treatment phase differs from that in the baseline phase and whether there is a difference between the rate of change across baseline and treatment phases (Howard et al., 2015). Regbuild and RSDT are used to compare a single case's performance to the performance of a control group. These statistics are applicable even for a modest-size control group, or non-normally distributed group data (Crawford & Garthwaite, 2005, 2007). A matched control group can serve as experimental control and can increase the internal validity of any findings.

Investigation of the effects of STM/WM treatments should be extended to people with severe aphasia. To better understand the underlying mechanisms of potential transfer effects, a careful choice of outcome measures and a justification of hypotheses about potential improvements on these measures are required. Outcome measures on memory and communication activities in daily life (e.g., self-reported memory and communication questionnaires) should be included in future studies to demonstrate clinically significant improvements, in particular, the General Health Questionnaire (GHQ-12; Goldberg, 1972) and Stroke and Aphasia Quality of Life Scale (SAQOL-39; Hilari, Byng, Lamping, & Smith, 2003) which measure emotional wellbeing and quality of life respectively, and have been recommended as core outcome measures for inclusion in aphasia trials (Wallace et al, 2018). Finally, given the heterogeneous outcome measures reported in STM/WM research, future studies should include recommendations about a Core Outcome Set related to STM/WM trials with aphasia, that is, the minimum set of outcomes that should be measured and reported in STM/WM treatments. This would potentially enable comparison of equivalent research and use of this data in meta-analysis.

References

Albert, M. L. (1976). Short-term memory and aphasia. *Brain & Language*, *3*(1), 28–33. https://doi.org/10.1016/0093-934X(76)90003-1

Andrew, N. E., Kilkenny, M., Naylor, R., Purvis, T., Lalor, E., Moloczij, N., & Cadilhac, D. A. (2014). Understanding long-term unmet needs in Australian survivors of stroke. *International Journal of Stroke*, *9*(SA100), 106–112. https://doi.org/10.1111/ijs.12325

Backman, C. L., Harris, S. R., Chisholm, J.-A. M., & Monette, A. D. (1997). Single-subject research in rehabilitation: A review of studies using AB, withdrawal, multiple baseline, and alternating treatments designs. *Archives of Physical Medicine and Rehabilitation*, *78*(10), 1145–1153. https://doi.org/10.1016/S0003-9993(97)90142-8

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*(1), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Balasooriya-Smeekens, C., Bateman, A., Mant, J., & De Simoni, A. (2016). Barriers and facilitators to staying in work after stroke: insight from an online forum. *BMJ Open*, *6*(4), e009974. https://doi.org/10.1136/bmjopen-2015-009974

Beeson, P. M., & Robey, R. R. (2006). Evaluating Single-Subject Treatment Research: Lessons Learned from the Aphasia Literature. *Neuropsychology Review*, *16*(4), 161–169. https://doi.org/10.1007/s11065-006-9013-7

Berthier, M. L., Dávila, G., Green-Heredia, C., Torres, I. M., Mier, R. J. y R. de, De-Torres, I., & Ruiz-Cruces, R. (2014). Massed sentence repetition training can augment and speed up recovery of speech production deficits in patients with chronic conduction aphasia receiving donepezil treatment. *Aphasiology*, *28*(2), 188–218. https://doi.org/10.1080/02687038.2013.861057

Bhogal, S. K., Teasell, R., & Speechley, M. (2003). Intensity of aphasia therapy, impact on recovery. *Stroke*, *34*(4), 987–993. https://doi.org/10.1161/01.STR.0000062343.64383.D0

Booth, A., Clarke, M., Dooley, G., Ghersi, D., Moher, D., Petticrew, M., & Stewart, L. (2012). The nuts and bolts of PROSPERO: An international prospective register of systematic reviews. *Systematic Reviews*, *1*, 2. https://doi.org/10.1186/2046-4053-1-2

Booth, A., Clarke, M., Ghersi, D., Moher, D., Petticrew, M., & Stewart, L. (2011). An international registry of systematic-review protocols. *The Lancet*, *377*(9760), 108–109. https://doi.org/10.1016/S0140-6736(10)60903-8

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*(12), 539–546. https://doi.org/10.1016/j.tics.2004.10.003

Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. In *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd. https://doi.org/10.1002/14651858.CD000425.pub4

Caplan, D., Michaud, J., & Hufford, R. (2013). Short-term memory, working memory, and syntactic comprehension in aphasia. *Cognitive Neuropsychology*, *30*(2), 77–109. https://doi.org/10.1080/02643294.2013.803958

Caramazza, A., Zurif, E. B., & Gardner, H. (1978). Sentence memory in aphasia. *Neuropsychologia*, *16*(6), 661–669. https://doi.org/10.1016/0028-3932(78)90001-5

Caspari, I., Parkinson, S. R., LaPointe, L. L., & Katz, R. C. (1998). Working memory and aphasia. *Brain and Cognition*, *37*(2), 205–223. https://doi.org/10.1006/brcg.1997.0970

Coelho, C. (2005). Direct attention training as a treatment for reading impairment in mild aphasia. *Aphasiology*, *19*(3–5), 275–283. https://doi.org/10.1080/02687030444000741

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163–183. https://doi.org/10.1016/S0160-2896(01)00096-4

Crawford, J. R., & Garthwaite, P. H. (2005). Testing for suspected impairments and dissociations in single-case studies in neuropsychology: Evaluation of alternatives using monte carlo simulations and revised tests for dissociations. *Neuropsychology*, *19*(3), 318–331. https://doi.org/10.1037/0894-4105.19.3.318

Crawford, J. R., & Garthwaite, P. H. (2007). Using regression equations built from summary data in the neuropsychological assessment of the individual case. *Neuropsychology*, *21*(5), 611–620. https://doi.org/10.1037/0894-4105.21.5.611

Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology*, *27*(3), 245–260. https://doi.org/10.1080/02643294.2010.513967

De Renzi, E., & Nichelli, P. (1975). Verbal and Non-Verbal Short-Term Memory Impairment Following Hemispheric Damage. *Cortex*, *11*(4), 341–354. https://doi.org/10.1016/S0010-9452(75)80026-8

DeDe, G., Ricca, M., Knilans, J., & Trubl, B. (2014). Construct validity and reliability of working memory tasks for people with aphasia. *Aphasiology*, *28*(6), 692–712. https://doi.org/10.1080/02687038.2014.895973

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*(4), 801–838. https://doi.org/10.1037//0033-295X.104.4.801

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature*, *378*(6554), 279–281. https://doi.org/10.1038/378279a0

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23. https://doi.org/10.1111/1467-8721.00160

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology. General*, *128*(3), 309–331.

Eom, B., & Sung, J. E. (2016). The effects of sentence repetition–based working memory treatment on sentence comprehension abilities in individuals with aphasia. *American Journal of Speech-Language Pathology*, *25*(4S), S823–S838. https://doi.org/10.1044/2016_AJSLP-15-0151

Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Frontiers in Psychology*, *5*, 1–17. https://doi.org/10.3389/fpsyg.2014.00335

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Current Biology*, *22*(21), 2059–2062. https://doi.org/10.1016/j.cub.2012.09.011

Field, A. (2009). *Discovering statistics using SPSS*. London: Sage publications.

Francis, D., Clark, N., & Humphreys, G. (2003). The treatment of an auditory working memory deficit and the implications for sentence comprehension abilities in mild "receptive" aphasia. *Aphasiology*, *17*(8), 723–750. https://doi.org/10.1080/02687030344000201

Frankel, T., Penn, C., & Ormond-Brown, D. (2007). Executive dysfunction as an explanatory basis for conversation symptoms of aphasia: A pilot study. *Aphasiology*, *21*(6–8), 814–828. https://doi.org/10.1080/02687030701192448

Fridriksson, J., Holland, A., Beeson, P., & Morrow, L. (2005). Spaced retrieval treatment of anomia. *Aphasiology*, *19*(2), 99–109. https://doi.org/10.1080/02687030444000660

Fridriksson, J., Nettles, C., Davis, M., Morrow, L., & Montgomery, A. (2006). Functional communication and executive function in aphasia. *Clinical Linguistics & Phonetics*, *20*(6), 401–410. https://doi.org/10.1080/02699200500075781

Friedman, N. P., & Miyake, A. (2004). The Relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, *133*(1), 101–135. https://doi.org/10.1037/0096-3445.133.1.101

Geranmayeh, F., Wise, R. J. S., Mehta, A., & Leech, R. (2014). Overlapping networks engaged during spoken language production and its cognitive control. *Journal of Neuroscience*, *34*(26), 8728–8740. https://doi.org/10.1523/JNEUROSCI.0428-14.2014

Goldberg, D. (1972). *The detection of psychiatric illness by questionnaire.* In: Maudsley Monographs. Oxford: Oxford University Press.

Goodglass, H., & Kaplan, E. (1983). Boston diagnostic examination for aphasia. Philadelphia: Lea & Febiger.

Greenhalgh, T. (2014). *How to read a paper: The basics of evidence-based medicine*. John Wiley & Sons.

Gvion, A., & Friedmann, N. (2012). Does phonological working memory impairment affect sentence comprehension? A study of conduction aphasia. *Aphasiology*, *26*(3–4), 494–535. https://doi.org/10.1080/02687038.2011.647893

Haarmann, H. J., Just, M. A., & Carpenter, P. A. (1997). Aphasic sentence comprehension as a resource deficit: A computational approach. *Brain & Language*, *59*(1), 76–120. https://doi.org/10.1006/brln.1997.1814

Harris, L., Olson, A., & Humphreys, G. (2014). The link between STM and sentence comprehension: A neuropsychological rehabilitation study. *Neuropsychological Rehabilitation*, *24*(5), 678–720. https://doi.org/10.1080/09602011.2014.892885

Helm-Estabrooks, N. (1992). *ADP: Aphasia Diagnostic Profiles*. Riverside Publishing Company.

Hilari, K., Byng, S., Lamping, D. L., & Smith, S. C. (2003). Stroke and Aphasia Quality of Life Scale-39 (SAQOL-39): Evaluation of acceptability, reliability, and validity. *Stroke*, *34*(8), 1944-1950.

Howard, D., Best, W., & Nickels, L. (2015). Optimising the design of intervention studies: Critiques and ways forward. *Aphasiology*, *29*(5), 526–562. https://doi.org/10.1080/02687038.2014.985884

Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition*, *42*(3), 464–480. https://doi.org/10.3758/s13421-013-0364-z

Kalinyak-Fliszar, M., Kohen, F., & Martin, N. (2011). Remediation of language processing in aphasia: Improving activation and maintenance of linguistic representations in (verbal) short-term memory. *Aphasiology*, *25*(10), 1095–1131. https://doi.org/10.1080/02687038.2011.577284

Keil, K., & Kaszniak, A. W. (2002). Examining executive function in individuals with brain injury: A review. *Aphasiology*, *16*(3), 305–335. https://doi.org/10.1080/02687030143000654

Kendall, J. M. (2003). Designing a research project: Randomised controlled trials and their principles. *Emergency Medicine Journal*, *20*(2), 164–168. https://doi.org/10.1136/emj.20.2.164

Kertész, A. (1982). *Western Aphasia Battery*. New York: The Psychological Corporation.

Kim, H., & Na, D. L. (2001). *Korean version of the Western Aphasia Battery (K-WAB)*. Seoul, Korea: Paradise Welfare Foundation.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, *24*(6), 781–791. https://doi.org/10.1076/jcen.24.6.781.8395

Koenig-Bruhin, M., & Studer-Eichenberger, F. (2007). Therapy of short-term memory disorders in fluent aphasia: A single case study. *Aphasiology*, *21*(5), 448–458. https://doi.org/10.1080/02687030600670593

Lee, J. B., & Sohlberg, M. M. (2013). Evaluation of attention training and metacognitive facilitation to improve reading comprehension in aphasia. *American Journal of Speech-Language Pathology*, *22*(2), S318-333. https://doi.org/10.1044/1058-0360(2013/12-0099)

Luna, C. Z. (2011). Functional communication in chronic aphasia and executive function: The effect of treating cognitive flexibility [Clinical Aphasiology Paper]. Retrieved April 18, 2017, from http://aphasiology.pitt.edu/2243/

Majerus, S. (2017). Working memory treatment in aphasia: A theoretical and quantitative review. *Journal of Neurolinguistics*, 1–19. https://doi.org/10.1016/j.jneuroling.2017.12.001

Majerus, S., Kaa, V. D., A, M., Renard, C., Van der Linden, M., & Poncelet, M. (2005). Treating verbal short-term memory deficits by increasing the duration of temporary phonological representations: A case study. *Brain & Language*, *95*(1). https://doi.org/10.1016/j.bandl.2005.07.094

Martin, N., & Gupta, P. (2004). Exploring the relationship between word processing and verbal short-term memory: Evidence from associations and dissociations. *Cognitive Neuropsychology*, *21*(2–4), 213–228. https://doi.org/10.1080/02643290342000447

Martin, N., Kohen, F., Kalinyak-Fliszar, M., Soveri, A., & Laine, M. (2012). Effects of working memory load on processing of sounds and meanings of words in aphasia. *Aphasiology*, *26*(3–4), 462–493. https://doi.org/10.1080/02687038.2011.619516

Martin, N., Minkina, I., Kohen, F. P., & Kalinyak-Fliszar, M. (2018). Assessment of linguistic and verbal short-term memory components of language abilities in aphasia. *Journal of Neurolinguistics*, 1–27. https://doi.org/10.1016/j.jneuroling.2018.02.006

Martin, N., & Reilly, J. (2012). Short-term/working memory impairments in aphasia: Data, models, and their application to aphasia rehabilitation. *Aphasiology*, *26*(3–4), 253–257. https://doi.org/10.1080/02687038.2011.648163

Martin, N., & Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*, *14*(5), 641–682. https://doi.org/10.1080/026432997381402

Martin, R. C. (2007). Semantic short-term memory, language processing, and inhibition. In A. Meyer, L. Wheeldon, & A. Krott (Eds.), *Automaticity and control in language processing* (pp. 161–191). Hove: Psychology Press.

Martin, R. C., & He, T. (2004). Semantic short-term memory and its role in sentence processing: A replication. *Brain & Language*, *89*(1), 76–82. https://doi.org/10.1016/S0093-934X(03)00300-6

Martin, R. C., & Romani, C. (1994). Verbal working memory and sentence comprehension: A multiple-components view. *Neuropsychology*, *8*(4), 506–523. https://doi.org/10.1037/0894-4105.8.4.506

Martin, Randi C., & Allen, C. M. (2008). A Disorder of executive function and its role in language processing. *Seminars in Speech and Language*, *29*(3), 201–205. https://doi.org/10.1055/s-0028-1082884

Mayer, J. F., & Murray, L. L. (2002). Approaches to the treatment of alexia in chronic aphasia. *Aphasiology*, *16*(7), 727–743. https://doi.org/10.1080/02687030143000870

McKevitt, C., Fudge, N., Redfern, J., Sheldenkar, A., Crichton, S., Rudd, A. R., … Wolfe, C. D. A. (2011). Self-reported long-term needs after stroke. *Stroke*, *42*(5), 1398–1403. https://doi.org/10.1161/STROKEAHA.110.598839

Minkina, I., Rosenberg, S., Kalinyak-Fliszar, M., & Martin, N. (2017). Short-term memory and aphasia: From theory to treatment. *Seminars in Speech and Language*, *38*(1), 17–28. https://doi.org/10.1055/s-0036-1597261

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge University Press.

Murray, L. L. (2012). Direct and indirect treatment approaches for addressing short-term or working memory deficits in aphasia. *Aphasiology*, *26*(3–4), 317–337. https://doi.org/10.1080/02687038.2011.589894

Murray, L. L., Keeton, R. J., & Karcher, L. (2006). Treating attention in mild aphasia: Evaluation of attention process training-II. *Journal of Communication Disorders*, *39*(1), 37–61. https://doi.org/10.1016/j.jcomdis.2005.06.001

Nickels, L., Rapp, B., & Kohnen, S. (2015). Challenges in the use of treatment to investigate cognition. *Cognitive Neuropsychology*, *32*(3–4), 91–103. https://doi.org/10.1080/02643294.2015.1056652

Novick, J. M., Kan, I. P., Trueswell, J. C., & Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology*, *26*(6), 527–567. https://doi.org/10.1080/02643290903519367

Paek, E. J., & Murray, L. L. (2015). Working memory approach for language and memory deficits in an individual with chronic aphasia: A case study. *Ehearsay: Electronic Journal of the Ohio Speech-Language Hearing Association*, *5*(1), 86–98.

Peach, R. K. (1987). A Short-term memory treatment appraoch to the repetition deficit in conduction aphasia. *Clinical Aphasiology*, *17*, 35–45.

Peach, R. K., Nathan, M. R., & Beck, K. M. (2017). Language-Specific Attention Treatment for aphasia: Description and preliminary findings. *Seminars in Speech and Language*, *38*(01), 005–016. https://doi.org/10.1055/s-0036-1597260

Penn, C., Frankel, T., Watermeyer, J., & Russell, N. (2010). Executive function and conversational strategies in bilingual aphasia. *Aphasiology*, *24*(2), 288–308. https://doi.org/10.1080/02687030902958399

Ramsberger, G. (2005). Achieving conversational success in aphasia by focusing on non-linguistic cognitive skills: A potentially promising new approach. *Aphasiology*, *19*(10–11), 1066–1073. https://doi.org/10.1080/02687030544000254

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). The Test of Everyday Attention (TEA). San Antonio, TX: Psychological Corporation.

Robey, R. R. (2004). A five-phase model for clinical-outcome research. *Journal of Communication Disorders*, *37*(5), 401–411. https://doi.org/10.1016/j.jcomdis.2004.04.003

Robey, R. R., & Schultz, M. C. (1998). A model for conducting clinical-outcome research: An adaptation of the standard protocol for use in aphasiology. *Aphasiology*, *12*(9), 787–810. https://doi.org/10.1080/02687039808249573

Robinson, G., Blair, J., & Cipolotti, L. (1998). Dynamic aphasia: An inability to select between competing verbal responses? *Brain*, *121*(1), 77–89. https://doi.org/10.1093/brain/121.1.77

Salis, C. (2012). Short-term memory treatment: Patterns of learning and generalisation to sentence comprehension in a person with aphasia. *Neuropsychological Rehabilitation*, *22*(3), 428–448. https://doi.org/10.1080/09602011.2012.656460

Salis, C., Hwang, F., Howard, D., & Lallini, N. (2017). Short-term and working memory treatments for improving sentence comprehension in aphasia: A review and a replication study. *Seminars in Speech and Language*, *38*(1), 29–39. https://doi.org/10.1055/s-0036-1597262

Salis, C., Kelly, H., & Code, C. (2015). Assessment and treatment of short-term and working memory impairments in stroke aphasia: A practical tutorial. *International Journal of Language & Communication Disorders*, *50*(6), 721–736.

Salter, K., Jutai, J. W., Teasell, R., Foley, N. C., Bitensky, J., & Bayley, M. (2005). Issues for selection of outcome measures in stroke rehabilitation: ICF activity. *Disability and Rehabilitation*, *27*(6), 315–340. https://doi.org/10.1080/09638280400008545

Seniów, J., Litwin, M., & Leśniak, M. (2009). The relationship between non-linguistic cognitive deficits and language recovery in patients with aphasia. Journal of the *Neurological Sciences*, *283*(1-2), 91-94. https://doi.org/10.1016/j.jns.2009.02.315

Smith, E. E., & Jonides, J. (1998). Neuroimaging analyses of human working memory. *Proceedings of the National Academy of Sciences*, *95*(20), 12061–12068. https://doi.org/10.1073/pnas.95.20.12061

Smith, E. E., Jonides, J., Marshuetz, C., & Koeppe, R. A. (1998). Components of verbal working memory: Evidence from neuroimaging. *Proceedings of the National Academy of Sciences*, *95*(3), 876–882.

Sohlberg, M. M., Johnson, L. P., Raskin, S. A., & Mateer, C. A. (2001). *Attention process training-II: A program to address attentional deficits for persons with mild cognitive dysfunction* (2nd ed.). Wake Forest, NC: Lash & Associates.

Suleman, S., & Kim, E. (2015). Decision-making, cognition, and aphasia: Developing a foundation for future discussions and inquiry. *Aphasiology*, *29*(12), 1409–1425. https://doi.org/10.1080/02687038.2015.1049584

Sung, J. E., McNeil, M. R., Pratt, S. R., Dickey, M. W., Hula, W. D., Szuminsky, N. J., & Doyle, P. J. (2009). Verbal working memory and its relationship to sentence-level reading and listening comprehension in persons with aphasia. *Aphasiology*, *23*(7–8), 1040–1052. https://doi.org/10.1080/02687030802592884

Tate, R. L., Perdices, M., Rosenkoetter, U., Wakim, D., Godbee, K., Togher, L., & McDonald, S. (2013). Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: The 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. *Neuropsychological Rehabilitation*, *23*(5), 619–638. https://doi.org/10.1080/09602011.2013.824383

Tate, R. L., Rosenkoetter, U., Wakim, D., Sigmundsdottir, L., Doubleday, J., Togher, L., … Perdices, M. (2015). *The risk-of-bias in N-of-1 trials (RoBiNT) scale: An expanded manual for the critical appraisal of single-case reports*. Sydney, Australia: Author.

Teasell, R., Foley, N., Salter, K., Bhogal, S., Jutai, J., & Speechley, M. (2009). Evidence-based review of stroke rehabilitation: Executive summary (12th ed.). *Topics in Stroke Rehabilitation*, *16*(6), 463–488. https://doi.org/10.1310/tsr1606-463

Thothathiri, M., & Mauro, K. L. (2018). The relationship between short-term memory, conflict resolution, and sentence comprehension impairments in aphasia. *Aphasiology*, *32*(3), 264–289. https://doi.org/10.1080/02687038.2017.1350630

Vallat, C., Azouvi, P., Hardisson, H., Meffert, R., Tessier, C., & Pradat-Diehl, P. (2005). Rehabilitation of verbal working memory after left hemisphere stroke. *Brain Injury*, *19*(13), 1157–1164. https://doi.org/10.1080/02699050500110595

Vallat-Azouvi, C., Pradat-Diehl, P., & Azouvi, P. (2014). Modularity in rehabilitation of working memory: A single-case study. *Neuropsychological Rehabilitation*, *24*(2), 220–237. https://doi.org/10.1080/09602011.2014.881294

Wallace, S. J., Worrall, L., Rose, T., Dorze, G. L., Breitenstein, C., Hilari, K., … Webster, J. (2018). A core outcome set for aphasia treatment research: The ROMA consensus statement. *International Journal of Stroke*. https://doi.org/10.1177/1747493018806200

Waters, G., Caplan, D., & Hildebrandt, N. (1991). On the Structure of Verbal Short-term Memory and its Functional Role in Sentence Comprehension: Evidence from Neuropsychology. *Cognitive Neuropsychology*, *8*(2), 81–126. https://doi.org/10.1080/02643299108253368

Webster, J., Whitworth, A., & Morris, J. (2015). Is it time to stop “fishing”? A review of generalisation following aphasia intervention. *Aphasiology*, *29*(11), 1240–1264. https://doi.org/10.1080/02687038.2015.1027169

Wright, H. H., Downey, R. A., Gravier, M., Love, T., & Shapiro, L. P. (2007). Processing distinct linguistic information types in working memory in aphasia. *Aphasiology*, *21*(6–8), 802–813. https://doi.org/10.1080/02687030701192414

Zakariás, L., Keresztes, A., Marton, K., & Wartenburger, I. (2018). Positive effects of a computerised working memory and executive function training on sentence comprehension in aphasia. *Neuropsychological Rehabilitation*, *28*(3), 369–386. https://doi.org/10.1080/09602011.2016.1159579

Zakariás, L., Salis, C., & Wartenburger, I. (2018). Transfer effects on spoken sentence comprehension and functional communication after working memory training in stroke aphasia. *Journal of Neurolinguistics*, 1–17. https://doi.org/10.1016/j.jneuroling.2017.12.002

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

Identification

Records identified through database searching
(n =552)

Academic Search Complete (2002-Dec 2016):	98 studies
CINAHL Plus FT (1996-Dec 2016):	39 studies
CCRCT (Dec 2016):	30 studies
CDSR (Dec 2016):	30 studies
Education FT (2011-Dec 2016):	2 study
Medline (1971-Dec 2016):	101 studies
Omnifile FT (2003-Dec 2016):	4 studies
PsycBITE (Dec 2016)	43 studies
Psychology & Behavioural Sciences (2003-Dec 2016):	16 studies
PsycINFO (1967-Dec2016):	83 studies
SpeechBITE (Dec 2016):	104 studies
PsychARTICLES (Dec 2016):	1 study
Social Sciences FT (Dec 2016):	1 study

Additional records identified through other sources
(n = 20)

Screening

Records after duplicates removed
(n =346)

Records screened on title and abstract
(n =346)

Records excluded based on title and abstract
(n =301)

Eligibility

Full-text articles assessed for eligibility
(n =45)

Full-text articles excluded:
(n =28)

Population: 6 studies
Intervention protocol: 18 studies
Intervention protocol and outcome measures: 3 studies
Unable to trace paper: 1 study

Included

Studies included in the review
(n =17)

Table 1. Participant characteristics

Study	*N*	*Age (mean)*	*Education (mean)*	*Gender*	*Handedness*	*TPO (mean)*	*Aphasia type*	*Aphasia severity*	*Reported coexisting STM and WM impairments*	*Reported coexisting impairments*
Berthier et al. (2014)	3	51-72 (53.3)	Left school at the age of 15 and 16; NR	M	R	13-22 m (17.3 m)	Conduction	1 Moderate 2 Mild (WAB)	NR	1 Mild AOS 1 R Dystonia
Eom and Sung (2016)	6	17-59 (45)	0.5-12 y (7.17 y)	3 F 3 M	R	11-35 m (19.5 m)	3 Anomic 2 Broca's 1 Wernicke	3 Mild 3 Moderate (WAB)	NR	NR
Francis et al. (2003)	1	69	NR	F	R	Unclear	NR	NR	Severe impairment of auditory WM	Verbal dyspraxia
Harris et al. (2014)	2	73, 74	Non-university diploma; Law school education	M	NR	NR	1 Broca's 1 Recovered	NR Recovered (BCoS)	1 pSTM impairment 1 sSMT impairment	NR
Kalinyak-Fliszar et al. (2011)	1	55	High school diploma	F	R	29 m	Conduction	Moderate (WAB)	Verbal STM impairment	NR
Koenig-Bruhin and Studer-Eichenberger (2007)	1	47	University degree	M	R	34 m	Conduction (AAT)	NR	Verbal STM impairment	NR
Lee and Sohlberg (2013)	4	57-83 (71)	14-23 y (17.25 y)	2 M 2 F	NR	18-79 m (43 m)	3 Anomic 1 Conduction	3 Mild 1 Moderate (WAB)	NR	1 Mild AOS 4 Impaired attention
Majerus et al. (2005)	1	50	NR	F	NR	NR	NR	NR	pSTM impairment	NR
Mayer and Murray (2002)	1	62	University degree and some work towards doctorate	M	R	NR	Fluent	Mild (WAB)	Impaired STM and WM	R-sided weakness, oral-motor deficits, impaired attention

Table 1. (Continued)

Study	*N*	*Age (mean)*	*Education (mean)*	*Gender*	*Handedness*	*TPO (mean)*	*Aphasia type*	*Aphasia severity*	*Reported coexisting STM and WM impairments*	*Reported coexisting impairments*
Murray et al. (2006)	1	57	University degree	M	R	4 y	Conduction	Mild (ADP)	Verbal WM deficit	Mildly impaired attention
Peach (1987)	1	53	NR	F	NR	11 d	Conduction	Moderate (BDAE)	NR	NR
Peach et al. (2017)	4	39-81 (62)	NR	3 M 1 F	R	NR	3 Non-fluent 1 Fluent	3 Mild 1 Moderate (WAB-R)	NR	2 Attention impairment
Salis (2012)	1	73	University degree	F	L	5 y	TMA	Severe (WAB)	Verbal and visuo-spatial STM impairment	R Hemiplegia
Salis et al. (2017)	5	47-86 (63.2)	8-16 y (12 y)	4 M 1 F	NR	8-180 m (85.6 m)	NR	2 Severe 1 Moderate 1 Mild-moderate 1 Mild (BDAE)	5 Verbal STM impairment 4 Visuo-spatial STM impairment	3 Mild AOS 2 Moderate AOS
Vallat et al. (2005)	1	53	High school diploma	Male	R	14 m	Recovered from conduction	Recovered	WM deficit	NR
Vallat-Azouvi et al. (2014)	1	38	NR	F	NR	> 4 y	Recovered from conduction	Recovered	WM deficit	NR
Zakariás et al. (2018)	3	57-64 (61.3)	10-11 y (10.7 y)	2 M 1 F	R	8-12 m (10.7 m)	1 Anomic 2 TMA	Moderate (WAB)	NR	NR

Note. WAB = Western Aphasia Battery; BCoS = Birmingham Cognitive Screen; AAT = Aachen Aphasia Test; ADP = Aphasia Diagnostic Profiles; BDAE = Boston Diagnostic Aphasia Examination; y = years; M = male; F = female; R = right; L = left; m = months; NR = not reported; TMA = transcortical motor aphasia; AOS = apraxia of speech; WM = working memory; STM = short-term memory pSTM = phonological STM; sSTM = semantic STM.

Table 2. Treatment characteristics

Study	*Design*	*Target cognitive construct*	*Treatment procedure*	*Schedule and treatment duration*
Berthier et al. (2014)	Pre-post intervention	Tx1: Language skills such as naming, repetition, spoken word comprehension and conversation Tx2: Auditory-verbal STM	Tx1. Distributed speech-language therapy (DSLT) combined with the cholinesterase inhibitor donepezil (DP) Tx2. Massed sentence repetition therapy (MSRT) combined with DP	Tx1: ~2.5hrs per week, for 16 weeks (40 hours) Tx2: 5×1h per week, for 8 weeks (40 hours)
Eom and Sung (2016)	Pre-post intervention	Verbal WM	Repetition-based treatment protocol (repetition of sentences after auditory presentation, reconstruction of sentences by using word cards, and reading sentences aloud)	3×1h per week, for 4 weeks (12 hours)
Francis et al. (2003)	Pre-post intervention	Auditory-verbal STM	Sentence repetition	Home practice 2× per day, 5 days per week + 12 clinical sessions, for 17 weeks
Harris et al. (2014)	Pre-post intervention	Tx1: Phonological STM Tx2: Semantic STM	Tx1. Repetition and recognition tasks with non-words Tx2. Repetition and recognition tasks with words	1×90-min therapy session per week + 20 home trials per week, for 10 weeks (for both Tx1 and Tx2)
Kalinyak-Fliszar et al. (2011)	Multiple baseline across conditions with probe tasks	Auditory-verbal STM	Repetition with delay paradigm using words and non-words	3×45-60 min per week, 137 sessions

Table 2. (Continued)

Study	*Design*	*Target cognitive construct*	*Treatment procedure*	*Schedule and treatment duration*
Koenig-Bruhin and Studer-Eichenberger (2007)	Pre-post intervention with probe task	Auditory-verbal STM	Immediate and delayed repetition of compound nouns and sentences	2× per week, for 17 weeks
Lee and Sohlberg (2013)	AB+follow-up multiple baseline with probe task	Sustained, selective, and alternating attention, working memory	Attention Process Training-3 (APT-3, Sohlberg & Mateer, 2010)	4×30-45 min per week, for 8 weeks
Majerus et al. (2005)	Pre-post intervention	Auditory-verbal STM	Immediate and delayed repetition of word and non-word pairs	8 sessions per month, for 16 months
Mayer and Murray (2002)	Alternating treatment design with baseline	Tx1: Text level reading Tx2: attention and WM	Tx1. Repeated oral reading of text (Modified MOR, based on Beeson, 1998) Tx2. Sequence Exercises for Working Memory (modified reading span task with grammaticality judgments and spoken recall of semantic categories)	2hs/session, 11 sessions (for both Tx1 and Tx2)
Murray et al. (2006)	AB+follow-up multiple baseline with probe task	Sustained, selective, divided, and alternating attention, WM	ATP-2 (Sohlberg et al., 2010)	1×1h therapy session + 20–60-min home practice per week, for ~30 weeks

Table 2. (Continued)

Study	*Design*	*Target cognitive construct*	*Treatment procedure*	*Schedule and treatment duration*
Peach (1987)	AB+follow-up multiple baseline with probe task	Auditory-verbal STM	Auditory word sequencing (pointing to two-three pictures in a fieled of 10) and oral word sequencing (repetition of three words)	32 treatment sessions
Peach et al. (2017)	Pre-post intervention	Executive attention, attentional switching, visual selective attention/speed, sustained and divided attention, auditory-verbal working memory, auditory processing speed	Language Specific Attention Treatment (L-SAT)	45-71 sessions (including baseline and posttest), for 4-5-weeks
Salis (2012)	AB+follow-up multiple baseline	Auditory-verbal STM	Matching listening span	2×30 min per week, for 13 weeks
Salis et al. (2017)	AB+follow-up multiple baseline with probe task	Auditory-verbal STM	Matching listening span with feedback	~27-30 sessions
Vallat et al. (2005)	AB+follow-up multiple baseline across behaviors	Central executive and phonological loop	Reconstruction of words from oral spelling, reconstruction of words from oral spelling with a letter omitted, oral spelling, reconstruction of words from syllables, “alphabet way”, word sorting in alphabetic order, acronyms	3×1h per week, for 6 months

Table 2. (Continued)

Study	*Design*	*Target cognitive construct*	*Treatment procedure*	*Schedule and treatment duration*
Vallat-Azouvi et al. (2014)	AB+follow-up multiple baseline across behaviors	Tx1: Phonological loop Tx2: Visuo-spatial sketchpad Tx3: Central executive	Tx1. Reconstruction of words from oral spelling, reconstruction of words from oral spelling with a letter omitted, oral spelling, reconstruction of words from syllables, "alphabet way", word sorting in alphabetic order, acronyms Tx2. 2-D mental imagery on a chessboard, 2-D mental imagery on a calculator keyboard, 3-D mental imagery, visual *n*-back Tx3. *N*-back with words, *n*-back questions, reading span tasks	1×2hs per week for altogether 19 months (6, 6, and 7 months for Tx1, Tx2, and Tx3, respectively)
Zakariás et al. (2018)	Case-control (control group: N = 5)	WM and executive functions (i.e., interference control)	Adaptive *n*-back with letters	3-4×20 min per week, for 4-5 weeks

Note. Names of the designs are based on Tate et al. (2015). STM = short-term memory; WM = working memory; Tx=treatment.

Table 3. RoBiNT scores of included studies

Study	*Design with control*	*Randomisation*	*Sampling of behaviour*	*Blinding of people involved in the intervention*	*Blinding of assessor(s)*	*Interrater agreement*	*Treatment adherence*	***Internal validity subscale (max. 14)***	*Baseline characteristics*	*Setting*	*Dependent variable (target behavior)*	*Independent variable (therapy/intervention)*	*Raw data record*	*Data analysis*	*Replication*	*Generalisation*	***External validity subscale (max. 16)***	***Total score (max. 30)***
Berthier et al. (2014)	0	0	0	0	0	0	0	**0**	2	1	2	2	0	1	1	1	**10**	**10**
Eom and Sung (2016)	0	0	0	0	0	0	0	**0**	1	1	2	2	0	1	2	1	**10**	**10**
Francis et al. (2003)	0	0	0	0	0	0	0	**0**	2	1	1	1	0	1	0	1	**7**	**7**
Harris et al. (2014)	0	0	0	0	0	0	0	**0**	2	1	2	2	0	1	1	1	**10**	**10**
Kalinyak-Fliszar et al. (2011)	2	0	1	0	0	2	0	**5**	2	0	2	2	2	2	0	2	**12**	**17**
Koenig-Bruhin and Studer-	0	0	1	0	0	0	0	**1**	2	0	1	1	0	1	0	0	**5**	**6**

Eichenberger (2007)																		
Lee and Sohlberg (2013)	0	0	2	0	0	0	2	**4**	1	1	2	2	0	2	2	2	**12**	**16**
Majerus et al. (2005)	0	0	0	0	0	0	0	**0**	2	0	0	1	0	1	0	0	**4**	**4**

Study	*Design with control*	*Randomisation*	*Sampling of behaviour*	*Blinding of people involved in the intervention*	*Blinding of assessor(s)*	*Interrater agreement*	*Treatment adherence*	***Internal validity subscale (max. 14)***	*Baseline characteristics*	*Setting*	*Dependent variable (target behavior)*	*Independent variable (therapy/intervention)*	*Raw data record*	*Data analysis*	*Replication*	*Generalisation*	***External validity subscale (max. 16)***	***Total score (max. 30)***
Mayer and Murray (2002)	0	0	1	0	0	2	0	**3**	2	0	2	2	1	0	0	0	**7**	**10**
Murray et al. (2006)	0	0	0	0	0	0	0	**0**	2	1	1	2	0	1	0	1	**8**	**8**
Peach (1987)	0	0	0	0	0	0	0	**0**	1	0	2	0	2	0	0	1	**6**	**6**
Peach et al. (2017)	0	0	0	0	0	0	0	**0**	0	0	1	1	0	0	2	1	**5**	**5**
Salis (2012)	0	0	2	0	1	0	0	**3**	2	0	2	2	0	1	0	1	**8**	**11**
Salis et al. (2017)	0	0	0	0	1	0	2	**3**	1	1	2	0	0	1	2	1	**8**	**11**
Vallat et al. (2005)	0	0	0	0	0	0	0	**0**	2	0	1	2	0	1	0	1	**7**	**7**

Vallat-Azouvi et al. (2014)	0	0	0	0	1	0	0	**1**	2	0	1	2	0	1	0	1	**7**	**8**
Zakariás et al. (2018)	0	0	2	0	2	2	2	**8**	1	0	2	2	0	1	1	1	**8**	**16**

Table 4. Outcome measures and their results, indicating whether an improvement was based on statistics (IS) or descriptive analysis (DS)

Study	*N*	*Near transfer*	*Far transfer*
Berthier et al. (2014)	3	**STM:** PALPA word[IS(1)] and non-word[IS(3)] repetition, digit span, word pair repetition[IS(2,3)], word triplet repetition[IS(1,2,3)], sentence repetition[IS(1,2 3)]	**Aphasia severity:** WAB AQ[IS(1,2,3)] **Connected speech:** WAB picture description[DS(1,2,3)]
Eom and Sung (2016)	6	**STM/WM:** Digit span forward[IS] and backward[IS] (WAIS), word span forward[IS] and backward (Sung, 2011), sentence repetition[IS]	**Spoken sentence comprehension:** Sentence picture matching[IS] (Sung, 2015) **Aphasia severity:** WAB AQ[IS] **General language assessment:** WAB fluency, comprehension, repetition[IS], and naming, BNT
Francis et al. (2003)	1	**STM/WM:** Digit span forward and backward[IS] (WAIS-3), letter span, sentence repetition[IS], Recognition Memory Test Words[IS] (Warrington, 1984)	**Spoken sentence comprehension:** RTT[IS], TROG[IS], Reversible Sentence Comprehension Test (Byng & Black, 1999) **Control tasks:** PALPA written synonym judgment, PALPA non-word repetition, BNT
Harris et al. (2014)	2	**STM:** Non-word span[IS(1,2)], word span[IS(1,2)], semantic[IS(1)] and rhyming span[IS(1,2)], sentence repetition (Hanten & Martin, 2000)[IS(2)] **Cognitive assessment:** Birmingham Cognitive Screen (Humphreys et al., 2012)	**Spoken sentence comprehension:** Anomaly detection[IS(1)] (Hanten & Martin, 2000), PALPA spoken sentence picture matching[IS(1)]
Kalinyak-Fliszar et al. (2011)	1	**STM:** TALSA phoneme discrimination[DS], rhyme pair judgment[DS], synonym triplet judgment[DS], rhyming triplet judgment[DS], lexical comprehension, category judgment–pictures[DS], picture naming[DS], word and non-word repetition[DS], word pair repetition[DS], sentence repetition[DS], digit span–pointing, digit span–repetition, word and non-word span[DS], repetition span for words[DS], probe memory span	**Spoken sentence comprehension:** TALSA sentence comprehension[DS]

Table 4. (Continued)

Study	*N*	*Near transfer*	*Far transfer*
Koenig-Bruhin and Studer-Eichenberger (2007)	1	**STM:** Sentence repetition (probe task)[IS], digit span–pointing with auditory[DS] and visual presentation, word span–pointing with auditory[DS] and visual presentation, digit matching span[DS], rhyme and category recognition[DS]	**Connected speech:** Picture description[DS]
Lee and Sohlberg (2013)	4	**Cognitive assessment, WM:** Conners' Continuous Performance Test-Second Edition[DS(1,3,4)], TEA map search[DS(2)], elevator counting[DS(2)], elevator counting with distraction[DS(2,4)], visual elevator accuracy[DS(4)] and timing[DS(1,2,4)], elevator counting with reversal[DS(1,2)], telephone search[DS(1,2,3,4)], telephone search with counting[DS(1,2,3,4)], lottery[DS(2,4)]	**Reading comprehension:** Maze reading (probe task)[DS(1,2)], Reading Comprehension Battery for Aphasia-Second Edition[DS(1,3,4)], GORT-4[DS(1,2,4)]
Majerus et al. (2005)	1	**STM:** Digit span–repetition[IS], word span–repetition[IS], non-word span–repetition[IS], word repetition, non-word repetition, rhyme judgment[IS], minimal pair discrimination	
Mayer and Murray (2002)	1	**WM:** Listening span–spoken[DS] (Tompkins et al., 1994)	**Reading speed and comprehension:** Reading passages (probe task)[DS(increased rate)], GORT-3[DS]
Murray et al. (2006)	1	**Attention, STM/WM:** Retell directions to a local apartment complex[DS], Retell directions to get to a waterfall in a park in California[DS], Make phone call to get directions from Bloomington to Philadelphia[DS], Provide score and other numerical details while watching a videotaped football game[DS], Listening span–spoken[DS], digit span forward and backward–repetition (WMS-R), visual tapping forward and backward (WMS-R) **Cognitive assessment:** TEA [DS(Map search, Telephone search while counting)]	**Auditory comprehension and oral expression:** Paragraph listening (probe task)[IS(faster RTs)], Test of Language Competence-Expanded (Wiig & Secord, 1989) **General language assessment:** Aphasia Diagnostic Profiles (Helm-Estabrooks, 1992) **Everyday life functioning:** APT-2 Attention Questionnaire, CETI

Table 4. (Continued)

Study	*N*	*Near transfer*	*Far transfer*
Peach (1987)	1	**STM:** Sentence repetition (probe task)[DS]	**General language assessment:** BDAE [DS(Naming, Repetition)]

Peach et al. (2017)	4	**Attention and WM:** TEA map search[DS(1)], elevator counting, elevator counting with distraction[DS(1,3,4)], visual elevator accuracy[DS(1)] and timing[DS(1)], elevator counting with reversal[DS(1)], telephone search, telephone search with counting[DS(1)], lottery, Paced Auditory Serial Addition Test (PASAT)[DS(1,4)] **Executive functions:** Stroop Test[DS(1,4)]	**Aphasia severity:** WAB AQ[DS(1,4)] **Language:** Discourse Comprehension Test, Object and Action Naming Battery[DS(1)] (Druks & Masterson, 2000), American Speech-Language Hearing Association Quality of Communication Life Scale **Everyday life functioning:** Rating Scale of Attention Behaviors[DS(1,2)]
Salis (2012)	1	**STM/WM:** PALPA digit listening span[DS], digit span forward[DS] and backward–repetition (WMS-R)	**Spoken sentence comprehension:** TT, TROG[IS] **Control tasks:** PALPA minimal pairs, PALPA spoken noun comprehension
Salis et al. (2017)	5	**STM/WM:** Word span (probe task), digit span forward and backward–repetition (WMS-R), digit span forward–pointing, PALPLA digit matching listening span, word span forward–repetition, visual tap forward and backward (WMS-R)	**Spoken sentence comprehension:** TROG, TT **Psychosocial functioning:** CETI[IS(3)], Communication Outcome After Stroke (COAST)
Vallat et al. (2005)	1	**STM/WM:** Digit span forward[IS] and backward, visuo-spatial span forward and backward, letter span, word span, Brown-Peterson paradigm (consonant recall with or without interference)[IS], arithmetic problem solving task[IS]	**Spoken sentence comprehension:** Text oral comprehension task (based on Ducarne de Ribeaucourt, 1989)[IS] **Everyday life functioning:** Verbal communication questionnaire[IS] (Darringrand, 2000), Working memory questionnaire[IS]

Table 4. (Continued)

Study	*N*	*Near transfer*	*Far transfer*
Vallat-Azouvi et al. (2014)	1	**Attention, STM/WM:** Digit span forward[IS] and backward, visuo-spatial span forward and backward, Brown-Peterson paradigm (consonant recall with or without interference)[IS], *n*-back, TAP divided attention and mental flexibility, arithmetic problem solving	**Spoken sentence comprehension:** Text oral comprehension task (Ducarne de Ribeaucourt, 1989), TT **Everyday life functioning:** Working Memory Questionnaire (Vallat-Azouvi et al., 2012) **Control tasks:** Rey Complex Figure Recall, TAP visual reaction times and long-term verbal memory
Zakariás et al. (2018)	3	**WM:** *N*-back with pictures[IS(1,2)], *n*-back with computer-generated tones	**Spoken sentence comprehension:** TROG[IS(1,3)] **Control task:** BNT

Note. IS and DS in superscripts indicate an improvement in the task based on statistics and descriptive analysis, respectively; numbers in superscripts refer to the participants; PALPA = Psycholinguistic Assessments of Language Processing in Aphasia (Kay, Coltheart, Lesser, 1992); WAB = Western Aphasia Battery (Kertesz, 1982); K-WAIS = Korean version of the Wechsler Adult Intelligence Scale (Yeom, Park, Oh, & Lee, 1992); K-WAB = Koran version of the Western Aphasia Battery (Kim & Na, 2001); K-BNT = Korean version of the Boston Naming Test (Kim & Na, 1997); TT = Revised Token Test (McNeil & Prescott, 1978); BNT = Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1983); WAIS-3 = Wechsler Adult Intelligence Scale-Third Edition (Wechsler, 1997); TALSA = Temple Assessment of Language and Short-term Memory in Aphasia (Martin, Kohen & Kalinyak-Fliszar, 2010); TEA = Test of Everyday Attention (Robertson, Nimmo-Smith, Ward, & Ridgeway, 1994); GORT-4 = Gray Oral Reading Test-Fourth Edition (Wiederholt, & Bryant, 2001); GORT-3 = Gray Oral Reading Test-Third Edition (Wiederholt, & Bryant, 1992); WMS-R = Wechsler Memory Scale-Revised (Wechsler, 1987); CETI = Communicative Effectiveness Index (Lomas, Pickard, Bester, Elbard, Zoghaib, & Finlayson, 1989); BDAE = Boston Diagnostic Aphasia Examination Goodglass & Kaplan, 1983); TROG = Test for the Reception of Grammar (Bishop, 1989); TAP = Tests of Attentional Performance (Zimmermann & Fimm, 2002); TROG-H = Hungarian version of the Test for the Reception of Grammar (Lukács, Győri, & Rózsa, 2012).