TRANSFER EFFECTS AFTER WORKING MEMORY TRAINING IN POST-STROKE APHASIA

LILLA ZAKARIÁS

A cumulative dissertation
submitted to the Faculty of Human Sciences
of the University of Potsdam
in fulfillment of the requirements for the degree of Doctor of
Philosophy (Dr. phil.) in Cognitive Sciences

May 2018

Supervisors: Prof. Dr. Isabell Wartenburger (University of Potsdam)
Dr. Christos Salis (Newcastle University)

Internal reviewer: Prof. Dr. Isabell Wartenburger (University of Potsdam)

External reviewer: Prof. Dr. Nadine Martin (Temple University)

Date of defense: 5 November 2018, Potsdam

https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-423600
https://doi.org/10.25932/publishup-42360

Table of contents

Acknowledgements

Firstly, I would like to thank my supervisor, Prof. Isabell Wartenburger for providing me with the opportunity to work on this topic. I am profoundly grateful to her for her continuous encouragement and support. Without her guidance, valuable suggestions, and persistent help my research would have not been possible.

I am very grateful to Dr. Christos Salis, my co-supervisor for his motivation, enthusiasm, immense knowledge, and invaluable help. I have benefitted a lot from his knowledge and feedback. It has been a great pleasure working with him.

I would like to thank my husband, Attila, for his patience, optimism, and unconditional faith in my success that has also given me hope, as well as his helpful insights on my research. *Köszönöm a kitartó és mérhetetlen támogatást. Köszönöm, hogy mindvégig mellettem álltál. Hálámat nem tudom szavakba önteni.*

My sincere thanks go to Dr. Ágnes Lukács and Dr. Klára Marton. I am grateful that I could be a member of their research groups during my master and early PhD studies. Their work greatly inspired me to start focusing on the topic that became the topic of my PhD.

I thank the students, Vanessa Dawn Löffler, Thomas Hinsdorf, and Leonie Lampe, and the student research assistants, Susanne Pelke and Helene Gerda Killmer for their help with data collection and study preparation. My thanks also go to my former colleagues, Tünde Buják and Éva Mentuszné Jánosi, and Zsuzsanna Barta for their help running the study.

I owe my special thanks to the participants, K.K., B.L., B.B., R.D., V.O., and Z.A., for their supportive attitude and willingness to participate in my studies, and the constant inspiration they provided. It was a great pleasure working with them.

Heartfelt thanks go to my family and friends for their support throughout my life. *Drága szüleim, köszönöm, hogy bátorságotokkal, kitartásotokkal, és kemény munkátokkal lehetővé tettétek, hogy azzal foglalkozhassak, amit igazán szeretek.*

Finally, I am grateful and thankful to members of the Neurocognition of Language Group and of the Department of Linguistics, and all the persons who have given me support, valuable advice, and suggestions for the completion of my research.

List of figures

List of tables

SYNOPSIS

1 General introduction

Although aphasia has been traditionally defined as an acquired language disorder due to brain damage, it has been shown that individuals with aphasia (IWA) also often present with concomitant cognitive impairments, such as impairments of attention, short-term memory, working memory (WM), and executive functions. These observations have tremendous implications for the study and treatment of aphasia. For instance, although studies in the fields of cognitive neuropsychology, psycholinguistics, and cognitive neuroscience revealed a complex relationship between WM and language, and proposed diverse underlying mechanisms to explain the relationship, they agree on the clinical relevance of their findings: It is important to assess WM, and the potential of WM treatments in aphasia rehabilitation. More specifically, research led to the promising hypothesis that treatments of WM would lead to improvements in language functioning as well. To date, however, it is unclear how much IWA can benefit from treatments of WM, and whether improvements in WM generalize beyond the treatment task. This thesis aimed at answering these questions in a series of studies investigating whether (1) WM can be improved through WM training in post-stroke aphasia, and (2) WM training has beneficial effects on language processing, such as spoken sentence (and syntactic) comprehension, and functional communication. Beyond improving the evidence-base of treatments in post-stroke aphasia, the thesis also aimed at elucidating underlying mechanisms of the association between WM and language processing in aphasia.

In *Chapter 1* I first provide a description of models of WM adopted in the thesis and describe its relevant sub-processes. Then, following a brief discussion of previous research on WM interventions in healthy populations and introduction of terms and concepts critical to the thesis, I consider the relationship between WM and language deficits in aphasia. This leads to the discussion of research findings on WM treatments in aphasia, with a special focus on the effects of treatments on spoken sentence comprehension and the ecological validity of treatments. In this section I also describe how previous research led to the formulation of the research questions addressed in this thesis. In the following chapters I describe the research questions and hypotheses of the studies included in the thesis (*Chapter 2*) and summarize the methods and main findings of these studies (*Chapter 3*). In *Chapter 4* I interpret our findings, draw conclusions and make recommendations for future research. *Chapter 5, 6,* and *7* encompass the studies included in the thesis (Study 1, 2, and 3, respectively).

1.1 Theories of working memory

Working memory (WM) is defined as a limited capacity system, which temporarily maintains and stores information, supports human thought processes by providing an interface between perception, long-term memory, and action (Baddeley, 2012). There are many different theoretical concepts of what WM is composed of and how its functions are implemented. A widely accepted model of WM, that has motivated many empirical studies on aphasia (Christensen, Wright, & Ratiu, 2018; Eom & Sung, 2016; Francis, Clark, & Humphreys, 2003; Just & Carpenter, 1992; Rönnberg et al., 1996; Sung et al., 2009; Thothathiri & Mauro, 2018) is Baddeley's WM theory (Baddeley, 2003, 2012, for the original version of the model, see Baddeley & Hitch, 1974). According to Baddeley (2003, 2012), WM is a limited capacity system with two modality-specific components: The *phonological loop*, responsible for temporary storage of verbal material, and maintaining phonological items within a short-term store via verbal rehearsal, and the *visuo-spatial sketchpad*, responsible for temporary storage of visuo-spatial information. The model also contains a modality-general component, the *central executive*, which is an attention-control system manipulating information in the two modality-specific subsystems, controlling encoding and retrieval strategies, and retrieving information from long-term memory. The fourth component of the model, the *episodic buffer* is also considered to be modality-general, and its role lies in binding and integrating information from the phonological loop, visuo-spatial sketchpad, central executive, and long-term memory.

Baddeley (2012) also emphasizes differences between the multicomponent model of WM and earlier models of *short-term memory* (STM). The main differences include the multicomponent character of WM, its emphasis on combined processing and storage, and its importance in facilitating a range of cognitive skills and activities, such as reasoning and learning. Other theories also argue against a unitary character of STM (see R. C. Martin & He, 2004; R. C. Martin & Romani, 1994 for a multicomponent model of verbal STM). Despite certain differences in the characterization of STM, it is generally acknowledged that STM is responsible for the temporary maintenance and retrieval of information, whereas WM is generally viewed as a system of which components actively work together to manipulate information for executing a particular goal or plan (Cowan, 2008).

Baddeley's model (2003, 2012) gained popularity in aphasiology, because two components of the model, the phonological loop (typically measured by simple span tasks) and the central executive (typically measured by complex span tasks) have been successfully linked to a number of language processes, such as syntactic comprehension (e.g., Boyle, Lindell, & Kidd, 2013), as well as to language development, such as vocabulary acquisition (Avons, Wragg, Cupples, & Lovegrove, 1998; Majerus, Poncelet, Greffe, & Van der Linden, 2006) and reading and writing acquisition (Alloway et al., 2005; de Jong & Olson, 2004) in healthy populations. In addition, the phonological loop, suggested to maintain verbal information, has served as a good basis for investigating the role of WM in language processing in language-impaired populations.

Another WM model particularly relevant to the thesis is that by Miyake and colleagues (2000). Miyake et al. (2000) break down the central executive by distinguishing between three functions, each concerned with the manipulation of temporarily maintained information in WM: (1) updating and monitoring WM representations; (2) inhibition of dominant or prepotent responses; and (3) shifting attentional control between tasks and mental sets. These functions have also been considered under the umbrella term *executive functions* (Miyake, Friedman, et al., 2000). In my thesis, I focus on the first two processes (i.e., updating and monitoring WM representations, and inhibition of dominant responses), and use the term *interference control* to refer to different inhibition-related functions, such as the ability to suppress dominant, automatic responses and resolve interference between conflicting information, and the ability to resist memory intrusions from information that was previously relevant to the task but has since become irrelevant (Friedman & Miyake, 2004; Novick, Trueswell, & Thompson-Schill, 2005; Novick, Trueswell, & Thompson-Schill, 2010). An advantage of Miyake et al.'s model is that it allows for specifically testing executive mechanisms of WM, and their contributions to other higher-level cognitive functions, such as language processing. Because such executive mechanisms come into play when an individual is involved in a complex, novel activity, researchers' have suggested that they may serve a mediating role in everyday communication (Ramsberger, 2000). Motivated by this assumption, some studies in aphasiology focused on the role of executive mechanisms in everyday communication and conversation (Frankel, Penn, & Ormond-Brown, 2007; Miyake, Emerson, & Friedman, 2000; Purdy, 2002; Ramsberger, 2005).

In the last two decades, following the seminal work of Baddeley (2003) and Miyake et al. (2000), the concept of WM has undergone substantial change and various mechanisms have been proposed to account for performance patterns observed in WM tasks (for the most influential theories, see Barrouillet & Camos, 2001; Cowan, 2008; Engle, 2002; McElree, 2001; Oberauer & Lewandowsky, 2011; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012). Although these newer models substantially refined our understanding of the cognitive construct of WM, the thesis draws largely on the core construct of WM laid out in the initial work of Baddeley and Miyake and colleagues, and how training of these core cognitive abilities affects language abilities, such as spoken sentence comprehension and everyday communication in aphasia.

Because the literature of WM training in healthy populations has largely motivated theoretical as well as methodological aspects of the studies included in the thesis (e.g., terminology used, theoretical framework adopted, task selection), I begin with a brief description of this literature.

1.2 Previous research on working memory training in healthy individuals

WM training has become an increasingly popular research topic in the last two decades in the fields of experimental psychology and cognitive neuroscience. The central question of this research domain is whether training WM improves skills or abilities that critically depend on WM processes but are not targeted during training, that is, whether improvements in WM *transfer* to untrained abilities. Given several staunch claims that WM contributes to a range of higher-level cognitive abilities, including reading comprehension (Daneman & Carpenter, 1980), language acquisition (Baddeley, 2003), non-verbal problem solving (Logie, Gilhooly, & Wynn, 1994), and a number of domain-specific reasoning skills (Kane et al., 2004), WM training could theoretically yield wide performance gains on a number of abilities. If so, then WM training could be used to remediate cognitive decline and changes in healthy aging as well as in pathological populations, such as individuals with neurodegenerative disorders and brain-injury.

A number of studies have been set out to answer this question, and excitingly, many of them have reported transfer effects from WM training to a range of higher-level cognitive abilities, such as attention (Smith et al., 2009), executive functions (Dahlin, Neely, Larsson, Bäckman, & Nyberg, 2008), intelligence (e.g., Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Jaeggi, Buschkuehl, Jonides, & Shah, 2011), and language processing (Chein & Morrison, 2010; Hussey et al., 2017; Loosli, Buschkuehl, Perrig, & Jaeggi, 2012; Novick, Hussey, Teubner-Rhodes, Harbison, & Bunting, 2014).

Two forms of transfer have been distinguished in the literature (see also Figure 1): 1) *near transfer* referring to improvements on tasks that are similar to the training task but are not practiced during training (e.g., untrained WM tasks) and 2) *far transfer* referring to improvements on different cognitive abilities (e.g., fluid intelligence, language processing). Note that some researchers also distinguish between two forms of near transfer, namely task-specific transfer (referring to improvements on tasks that are structurally similar to the training task but differ in other aspects, for example, stimuli) and process-specific transfer (referring to improvements on tasks that are structurally different from the training task but measure the same construct). For an illustration of levels and examples of transfer see Figure 1.

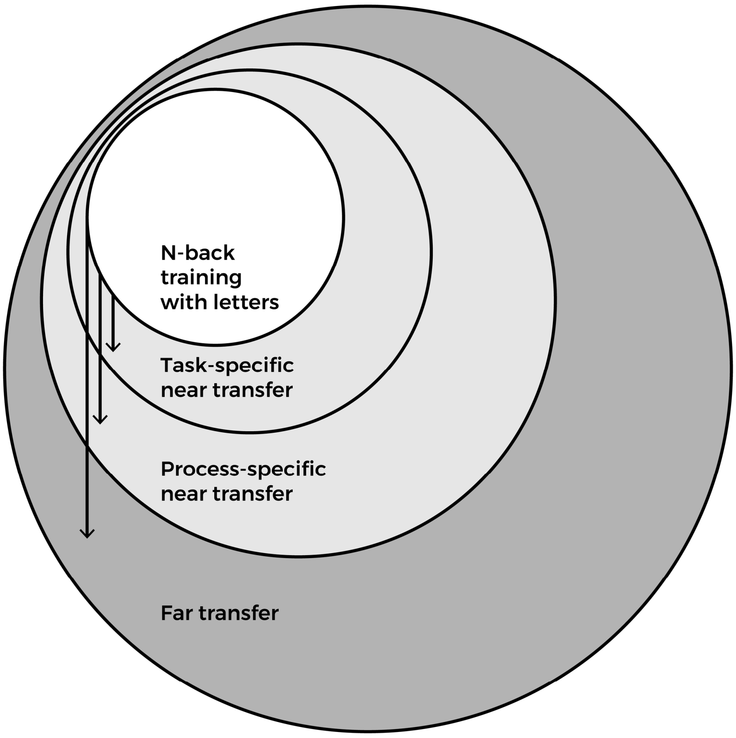

Figure 1 Illustration of the different levels of transfer after WM training
Task-specific transfer (a form of near transfer) refers to improvements on tasks that are structurally similar to the training task but differ in other aspects, for example, stimuli (e.g., *n*-back with digits). Process-specific transfer (another form of near transfer) refers to improvements on tasks that are structurally different from the training task but measure the same construct (e.g., updating tasks). Far transfer refers to improvements on different cognitive domains (e.g., fluid intelligence, language processing).

Neuroimaging studies have suggested that the criterion for the emergence of transfer effects is that the training task and the *outcome measures* (i.e., tasks on which near or far transfer is expected) share underlying cognitive processes and neural networks (Dahlin et al., 2008). Hereinafter, I refer to this criterion definition as the principle of ‘overlapping cognitive and neural mechanisms’, a widespread tenet of research on transfer (Dahlin et al., 2008; Hussey & Novick, 2012; Shipstead, Redick, & Engle, 2010, 2012). For example, Novick and colleagues (2014) showed that training healthy young adults on WM improved syntactic processing, as reflected in increased accuracy and decreased reaction times in a syntactic ambiguity resolution task (i.e., processing garden-path sentences). The authors argued that their outcome measure – the syntactic ambiguity resolution task – and their WM training task rely on shared cognitive mechanisms: Both require updating of WM representations and interference control to resolve conflicts between competing representations. In addition, the two tasks are supported by common neural networks/substrates, such as the left inferior frontal gyrus (LIFG; Novick et al., 2014). According to the authors, the study complied with the principle of 'overlapping cognitive and neural mechanisms', thus they were able to detect improvements on syntactic ambiguity resolution after WM training.

WM training in healthy young and elderly populations typically consists of an intensive (i.e., five times a week), minimum one-month long practice with one or more computerized WM tasks. The majority of WM training studies have used a modified version of the *n*-back task as a training task (Hussey et al., 2017; Jaeggi et al., 2008, 2011; Jaeggi, Buschkuehl, Shah, & Jonides, 2014; Lilienthal, Tamez, Shelton, Myerson, & Hale, 2013; Novick et al., 2014; Schweizer, Hampshire, & Dalgleish, 2011). The classic *n*-back task is a complex WM task involving several cognitive processes (i.e., encoding incoming stimuli, monitoring, maintaining, and updating WM representations, establishing and maintaining bindings between memory contents and their temporal context, as well as resolving interference between WM representations, Kane, Conway, Miura, & Colflesh, 2007) and being supported by several brain regions, including the dorsolateral prefrontal cortex (BA 9 and 46), the ventrolateral prefrontal cortex (BA 45 and 47), the anterior part of the frontal lobe (BA 10), the bilateral medial premotor cortex (BA 6 and 8), and the bilateral posterior parietal cortex (BA 7 and 40) (for review see Owen, McMillan, Laird, & Bullmore, 2005). Therefore, in line with the principle of 'overlapping cognitive and neural mechanisms', broad and large transfer effects on various cognitive domains have been expected after *n*-back training.

The classic *n*-back task is a computerized experimental task, in which participants are presented with a stream of stimuli (e.g., letters, numbers, or written words) and are asked to press the button when the stimulus they see or hear is the same as the one *n* (*n* = 1, *n* = 2, etc.) items before (Kirchner, 1958).

Task difficulty increases with *n*. The *n*-back training task is usually *adaptive*, that is, the difficulty level is continuously adjusted according to the participants' performance, ensuring that they always practise at an optimal level of difficulty.

Although results have consistently shown that WM training improves performance on the trained tasks (i.e., training effects), results regarding near and far transfer effects (i.e., the effectiveness of WM training) in healthy populations are mixed (Guye, Röcke, Mérillat, von Bastian, & Martin, 2016). Recent meta-analyses indicate at most moderate near transfer effects and small far transfer effects after adaptive WM training in healthy populations (Karbach & Verhaeghen, 2014; Kelly et al., 2014), with great inter-individual differences within and across studies (e.g., Kühn & Lindenberger, 2016).

Related to the great inter-individual differences detected in these studies, another central question of WM training research is what factors influence training and transfer effects. *Motivation* or interest in the training task seems to be one such factor. For example, Jaeggi et al. (2014) investigating university students found that participants' interest in the training task was related to the dropout from the study (i.e., participants with stable interest levels over the period of training completed the study more likely than those with gradually decreasing interest levels during training). Relating motivation more directly to training outcome, another study (Jaeggi et al., 2011) including healthy children suggested that the lack of improvement on the training task (and consequently, on the outcome measures) was related to the lack of interest in the training task. Another factor moderating training and transfer effects is the level of *initial cognitive abilities* (i.e., participants' cognitive abilities when entering the study), although this idea has received little attention in WM training studies. For example, Au et al. (2015) conducting a meta-analysis to identify factors that may influence transfer effects after *n*-back training in healthy adults between the ages of 18-50 years found a marginally significant negative relationship between initial *n*-back performance (i.e., WM ability) and far transfer to fluid intelligence. The authors concluded that participants who start with more "room" to improve in WM gain the most from WM training. Another study investigating the effect of strategy training on WM span reported similar results: initial WM span was negatively associated with improvements in WM span and reading ability, suggesting that those with low span benefitted more from strategy training (Turley-Ames & Whitfield, 2003). These findings, however, may be related to the sensitivity of the outcome measures. For example, in the study by Turley-Ames and Whitfield, training-related WM improvements were measured with a span task. While solving math problems, participants were required to remember a set of 2 to 6 unrelated words for spoken recall. Some of the participants' performance in the task may have been close to ceiling already at the beginning of training, thus the extent of improvement in WM may have been underestimated for these participants.

Although this thesis is concerned with WM-related changes at the cognitive-behavioral level, it is important to note that some neuroimaging studies have reported beneficial effects of WM training also on brain structure (e.g., increases in gray matter volume relevant for the trained function, Draganski et al., 2004; Mårtensson et al., 2012; Wenger et al., 2017). More importantly, based on these findings, some researchers have suggested that structural plasticity (i.e., long-lasting alterations in the brain's chemistry, gray matter, and structural connectivity in support of behavior, Kühn & Lindenberger, 2016) may be a prerequisite of training-related changes at the cognitive-behavioral level (i.e., transfer effects). If confirmed by direct tests in WM training studies, this theory may hold important implications for populations with brain damage, such as IWA.

1.3 Working memory and language processing in aphasia

Although aphasia has been traditionally defined as an acquired language disorder due to brain damage (most often stroke)(Basso, 2003; Goodglass, 1993; Hallowell & Chapey, 2008), it has been shown in several studies that IWA also often present with concomitant cognitive WM impairments (N. Martin & Ayala, 2004; Murray, Salis, Martin, & Dralle, 2018). Moreover, WM and language functions seem to interact in aphasia: WM deficits can negatively influence the individuals' lexical-semantic processing (N. Martin, Kohen, Kalinyak-Fliszar, Soveri, & Laine, 2012; Novick, Kan, Trueswell, & Thompson-Schill, 2009; Robinson, Blair, & Cipolotti, 1998), spoken sentence comprehension (Caplan, Michaud, & Hufford, 2013; Just & Carpenter, 1992; Sung et al., 2009; Wright, Downey, Gravier, Love, & Shapiro, 2007), reading (Caspari, Parkinson, LaPointe, & Katz, 1998), and other aspects of verbal communication (Frankel et al., 2007; Fridriksson, Nettles, Davis, Morrow, & Montgomery, 2006; Keil & Kaszniak, 2002; Luna, 2011; Penn, Frankel, Watermeyer, & Russell, 2010; Ramsberger, 2005).

Impairments of both the storage (N. Martin, Minkina, Kohen, & Kalinyak-Fliszar, 2018; N. Martin & Saffran, 1997; R. C. Martin, 2007; R. C. Martin & He, 2004; R. C. Martin & Romani, 1994; Minkina, Rosenberg, Kalinyak-Fliszar, & Martin, 2017; Thothathiri & Mauro, 2018) and the executive components of WM (Allen, Martin, & Martin, 2012; Fedorenko, 2014; Haarmann, Just, & Carpenter, 1997; Just & Carpenter, 1992; Novick et al., 2009, 2005; Tan & Martin, 2018; Thompson-Schill, Bedny, & Goldberg, 2005) have been shown to interfere with language processing in aphasia. The components and mechanisms of WM linked to language processing have varied in the literature according to changes in WM models during the last decades (see Caplan & Waters, 2013 for a review).

A major quest of aphasiology is to understand the nature of relationship between WM and language processing. Attempts to integrate results into a common theory linking WM and language processing are organized around two questions: (1) Is WM, and in particular, verbal STM (a subcomponent of WM) an inherent part of the language system, or are WM and language clearly separable (i.e., do WM mechanisms *support* language processing)? And (2) is WM necessary to language processing *per se,* or is its involvement only optional (i.e., is WM involved in language processing only under certain circumstances)? To highlight the first issue, as an example, I describe a model suggesting a complex interaction between verbal STM and word processing. To address the second, and as a focus of the thesis, I discuss theories focusing on the role of WM in spoken sentence comprehension. Finally, I describe theories suggesting a more general role of WM in both language comprehension and production. The theories described below unanimously suggest a complex relationship between WM and language, but they differ in the mechanisms proposed to underlie this relationship.

Focusing on single word retrieval and comprehension in IWA, Martin and Saffran (1997) suggest a close relationship between verbal STM and language processing. Their model (1997) is framed from the perspective of Dell's two-step interactive model of word retrieval (Dell, 1986; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; see also Minkina et al., 2017), which distinguishes between two retrieval steps (i.e., word retrieval and phonological retrieval), each achieved by activation spreading within the network. Martin and colleagues (1997, 2018) posit that verbal STM is encompassed in the temporary linguistic activation process underlying word retrieval, being responsible for sustaining the activation level of selected linguistic information of a target word until the word is produced. With regard to multiple word tasks such as verbal span tasks, verbal STM is responsible for maintaining linguistic representations of words over the time period needed to accomplish the tasks (N. Martin et al., 2018). The model posits that verbal STM is a cognitive requirement of language processing, therefore, it cannot be separated from language (N. Martin et al., 2018; but for alternative views see Majerus, Attout, Artielle, & Van der Kaa, 2015). More importantly, it also posits that impairments of word processing and verbal STM are a consequence of a single processing deficit determining the strength of activation of a word and the maintenance of its activation. Depending on the severity of the impairment, difficulties can emerge either in single word processing tasks such as repetition and naming, or multiple-word processing tasks such as verbal span, sentence repetition, or functional communication situations (for more detail on the Severity Continuum Hypothesis, see N. Martin & Gupta, 2004; N. Martin & Saffran, 1997).

With respect to spoken sentence comprehension, the main question that has been considered in the literature is what aspects of sentence comprehension are supported by components of the WM system[1]. Just and Carpenter (1992) suggested that WM is involved in parsing and interpretation (i.e., construction of

---

[1] A number of studies have investigated the relationship between WM and spoken sentence comprehension in aphasia (R. C. Martin & He, 2004; R. C. Martin & Romani, 1994; Thothathiri & Mauro, 2018; Wright, Downey, Gravier, Love, & Shapiro, 2007; see R. C. Martin, 2007 for a review). For example, within the framework of the multicomponent model of verbal STM (R. C. Martin & Romani, 1994), Martin and He (2004) suggested that semantic STM is involved in spoken sentence comprehension when several lexical-semantic representations need to be maintained prior to their integration (for a recent publication see Tan & Martin, 2018). Note that in this section I discuss only those theories that clearly relate WM to syntactic comprehension.

the syntactic structure of a sentence and the use of this structure to determine sentence meaning, respectively) and is mainly involved in processing syntactically complex sentences, such as object-relative clauses (Haarmann et al., 1997; Just & Carpenter, 1992; Papagno, Cecchetto, Reati, & Bello, 2007). Just and Carpenter (1992) argued that the same pool of WM resources tapped by WM tasks (e.g., verbal span) is also used in sentence processing (see the Capacity Theory, Just & Capenter, 1992). By contrast, Caplan et al. (2013) suggested that WM is not involved in the online, automatic processing of syntactic information (i.e., parsing and interpretation), but is engaged in the later stage of sentence comprehension (post-interpretive or expanded comprehension), namely the revision of the previously encountered, inaccurately interpreted information, and the use of the product of the comprehension to perform a task (e.g., in a picture-matching task keeping sentence meaning in mind while analyzing and interpreting the visual scenes and comparing them to the meaning of the sentence) (Gvion & Friedmann, 2012). Concerning the nature of the relationship between WM and language processing, both theories suggest that WM and spoken sentence comprehension are related but separable processes. However, while Just and Carpenter (1992) argue that WM is necessary for understanding complex syntactic structures, Caplan et al. (2013) suggest that WM plays a role only in situations when reanalysis of the first interpretation of a sentence is needed.

Another line of research suggests that interference control, defined as a domain-general function of WM supports several aspects of language comprehension as well as production. Interference control has a multifaceted role in language processing, such as detecting and resolving conflicts between competing linguistic representations of any kind (i.e., phonological, lexical, semantic, syntactic and/or referential, Biegler, Crowther, & Martin, 2008; Schnur, Schwartz, Brecher, & Hodgson, 2006; see Novick et al., 2010 for a review) and relatedly, inhibiting dominant representations in order to recover the relevant alternative (e.g., processing sentences with lexical, semantic, or syntactic ambiguity, Hamilton & Martin, 2007; Novick et al., 2009; Robinson et al., 1998; Vuong & Martin, 2011, 2015), selecting between competing sources of information (e.g., word order and syntactic cues in passive sentence processing, e.g., Thompson-Schill et al., 2005), and removing no longer relevant linguistic information from WM (Novick et al., 2009; Robinson et al., 1998; Thompson-Schill et al., 2005). A recent study investigating the relationship between IWA's interference control, verbal STM, and sentence comprehension abilities indicates a specific role of interference control and semantic STM specifically with regard to resolving syntactic and semantic interference during sentence processing, respectively (Tan & Martin, 2018). These theories suggest that deficits in interference control necessarily affect language comprehension and production in aphasia by impairing individuals' ability to resolve conflicting language representations (but see Thothathiri & Mauro, 2018).

In summary, the nature of the relationship between WM and language processing is not fully understood; more importantly, it seems that the description of such association depends on the specific language process researchers attempt to link to WM. However, the different lines of research point to the same direction in concluding that (1) WM is involved in several aspects of language processing, (2) there is a complex relationship between WM and language processing, therefore, (3) assessments and treatments of WM deficits may have a critical importance in aphasia rehabilitation. The studies in the thesis contribute further to this crucial debate by focusing on a limited number of preselected language functions, namely spoken sentence (and syntactic) comprehension, functional communication, and the effects of WM treatments on these functions. The ultimate goal of this thesis is to improve our understanding of the underlying mechanisms of the association between these functions.

1.4 Previous research on working memory treatments in aphasia

I will use the term *transfer* when referring to improvements after WM treatment throughout the thesis, despite that the term generalization is more widely used in the aphasiological treatment literature. Generalization is generally used to refer to improvements after language treatment on unpracticed items or language tasks (response and stimulus generalization, respectively, Harnish, Schwen Blackett, Zezinka, Lundine, & Pan, 2018; Webster, Whitworth, & Morris, 2015), that is, improvements emerging *within* the language domain. The term transfer refers to improvements after cognitive treatment (e.g., WM training) on different cognitive domains (e.g., general intelligence), that is, improvements emerging *across* cognitive domains. Because we aim at investigating the effects of WM training on language abilities (e.g., spoken sentence comprehension), I use the term transfer (near and far) throughout the thesis.

A detailed description of WM treatment studies (participants' cognitive-linguistic profile, target cognitive construct, treatment procedure, schedule and duration as well as outcome measures used) can be found in Chapter 7. Although studies of WM treatments in aphasia are scarce, the results so far are promising. Overall, studies have shown that IWA improved in the treatment tasks, suggesting that WM can be enhanced with practice in aphasia (e.g., Eom & Sung, 2016; Francis et al., 2003; Harris, Olson, & Humphreys, 2014; Kalinyak-Fliszar, Kohen, & Martin, 2011; Lee & Sohlberg, 2013; Mayer & Murray, 2002; Peach, 1987; Salis, 2012; Vallat et al., 2005). Improvements have also been reported on WM tasks that were not practiced during treatment (here considered near transfer effects) in the majority of studies (Berthier et al., 2014; Eom & Sung, 2016; Francis et al., 2003; Harris et al., 2014; Kalinyak-Fliszar et al., 2011; Lee & Sohlberg, 2013; Majerus et al., 2005; Murray, Keeton, & Karcher, 2006; Salis, 2012; Vallat et al., 2005; Vallat-Azouvi, Pradat-Diehl, & Azouvi, 2014). However, results of WM treatments on language processing (e.g., spoken sentence comprehension, spoken discourse, and reading) as well as other everyday functions (here considered far transfer effects) have been highly variable across studies. In addition, hypotheses about the underlying mechanisms of potential improvements in language and everyday functioning have been lacking in many studies. Clearly formulated hypotheses about underlying mechanisms of transfer effects are particularly important for making sense of the mixed pattern of transfer detected across studies and factors contributing to individual differences in training outcomes.

In the following section I summarize findings of WM treatment studies that focused on spoken sentence comprehension, which is the primary interest to this thesis.

1.4.1 Effects of working memory treatments on spoken sentence comprehension

Recent WM treatment studies that aimed to improve spoken sentence comprehension in aphasia have revealed promising findings (Eom & Sung, 2016; Francis et al., 2013; Harris et al., 2014; Salis, 2012). However, substantial variations in participant characteristics, treatment tasks, intensity and duration of treatment, as well as outcome measures used to detect potential improvements make the generalization of the results to the population level of aphasia difficult.

The majority of these studies targeted auditory-verbal STM and used repetition and/or recognition tasks involving words, non-words, or sentences as training tasks (Francis et al., 2003; Harris et al., 2014; Kalinyak-Fliszar et al., 2011; Salis, 2012). Recently, interest turned to investigating the effects of treatments of executive components of WM (Eom & Sung, 2016; Paek & Murray, 2015; Vallat-Azouvi et al., 2014). These studies mainly targeted the *updating* component of WM, using tasks such as *n*-back tasks with pictures and/or written words (Paek & Murray, 2015; Vallat-Azouvi et al., 2014), complex span tasks (Vallat-Azouvi et al., 2014), and sentence reconstruction tasks (Eom & Sung, 2016).

Although the majority of studies have reported improvements on spoken sentence comprehension (Eom & Sung, 2016; Francis et al., 2003; Harris et al., 2014; Kalinyak-Fliszar et al., 2011; Salis, 2012; Vallat et al., 2005)[2], only two studies have used outcome measures that allowed to investigate transfer effects on specific syntactic structures (Eom & Sung, 2016; Francis et al., 2003). These tests were a sentence picture-matching test by Sung (2015) in Eom and Sung (2016), and the Reversible Sentence Comprehension Test by Byng and Black (1999) in Francis et al. (2003). Interestingly, using a repetition-based WM treatment protocol, Eom and Sung (2016) detected improvements in the comprehension of treated syntactic structures after treatment, such as active sentences with two- and three- argument verbs, and passive sentences with two argument verbs. By contrast, Francis et al. (2003) did not detect significant

---

[2] See Table 4 in Chapter 6 (p. 83) for further details regarding methods and main findings of these studies.

improvements in the comprehension of reversible active sentences with two-argument verbs after a sentence-repetition treatment. Note that Eom and Sung (2016) used a pre-post intervention design where assessments on the outcome measures are conducted on one occasion prior to treatment and on one occasion after the completion of treatment. The use of a pre-post intervention design without introducing a control group (i.e., lack of proper experimental control) in Eom and Sung's study undermines the internal validity of their findings (Backman, Harris, Chisholm, & Monette, 1997).

One methodological concern related to WM treatment studies in aphasia is that only one study (Salis, Hwang, Howard, & Lallini, 2017) has aimed to replicate their previous positive findings (Salis, 2012). Salis and colleagues (2017) delivered a treatment in five IWA, however, unlike in their previous study (Salis, 2012), they did not detect improvements on spoken sentence comprehension or untrained WM tasks.

In summary, results of WM treatment studies investigating transfer effects on spoken sentence comprehension are promising, however, these effects should be replicated in future studies to draw conclusions about the external validity and generalizability of findings. In addition, underlying mechanisms behind transfer effects need to be better understood by studies focusing on the specificity of effects on syntactic comprehension. To our knowledge, among the studies focusing on the executive components of WM, none has looked at the effects of treatments of interference control, an executive component of WM proposed to be associated with spoken sentence comprehension (e.g., Allen et al., 2012; del Río et al., 2011; Fedorenko, 2014; Hsu, Jaeggi, & Novick, 2017; Novick et al., 2005, 2010; Thompson-Schill et al., 2005; Ye & Zhou, 2009). Therefore, future research in aphasia needs to be complemented with such studies.

1.4.2 Ecological validity of working memory treatments and maintenance of improvements in aphasia

The *ecological validity* of an intervention refers to the extent to which improvements in the intervention (e.g., WM treatment) transfer to real-life situations (e.g., everyday conversation abilities in IWA). WM deficits can affect many aspects of everyday life (e.g., conversation, organizing daily life activities, reading) as well as rehabilitation outcomes (e.g., returning and staying in work), according to studies conducted with stroke survivors (Balasooriya-Smeekens, Bateman, Mant, & De Simoni, 2016) and people with brain injury (Vallat-Azouvi, Pradat-Diehl, & Azouvi, 2012). Thus, it is of great interest to test for potential changes in everyday life functioning after WM treatment in aphasia. Note that treatment effects on everyday life functioning are also considered a form of far transfer effects according to the definition I adopt in the thesis.

In aphasia, only five studies have investigated the effects of WM treatment on everyday life functioning, and these have revealed mixed findings. Studies investigating transfer effects on everyday communication have mostly used self- and/or spouse-reported communication questionnaires (Murray et al., 2006; Peach, Nathan, & Beck, 2017; Salis et al., 2017; Vallat et al., 2005), such as the Communicative Effectiveness Index (Lomas et al., 1989). Among these studies, only two demonstrated improvements in this domain (Vallat et al., 2005; Salis et al., 2017). Results of questionnaires assessing everyday life problems related to WM and attention deficits were also variable across studies. Two studies reported improvements in WM and attention (Vallat et al., 2005; Peach et al., 2017, respectively) after treatment, whereas two studies did not report any changes in these domains (Murray et al., 2006; Vallat-Azouvi et al., 2014).

Finally, it is unresolved whether transfer effects last beyond the treatment period, and if so, for how long. To date, only two studies have tested the long-term effects of WM treatment in aphasia by retesting individuals some time after treatment completion (Kalinyak-Fliszar et al., 2011; Paek & Murray, 2015). Kalinyak-Fliszar et al. (2011) delivering a repetition with delay paradigm as treatment for one participant found that improvements in word and non-word repetition were maintained at 3-month follow-up. Paek and Murray (2015) conducting a single case study found that improvements in spoken discourse were maintained at 6-week follow-up after WM treatment including updating tasks.

Taken together, WM treatments in aphasia often lack ecological validity (e.g., Harris et al., 2014; Salis, 2012). The effect of WM treatments on everyday life functioning such as functional communication needs to be investigated in future studies to demonstrate clinically significant improvements. Studies should also include specific follow-up testing to demonstrate that improvements after WM treatment are maintained.

1.4.3 The role of individual differences in working memory training and transfer: Previous findings and their relevance to aphasia

Based on inconsistent results with regard to transfer effects across and within studies, the question arises as to what individual factors if any can contribute to the effectiveness of WM training (i.e., transfer effects). Studies in healthy populations suggest that such factors can include, among others, individual differences in motivation (Au et al., 2015; Jaeggi et al., 2014; Katz, Jaeggi, Buschkuehl, Stegman, & Shah, 2014) and initial cognitive abilities (e.g., Au et al., 2015; Morrison & Chein, 2011; Turley-Ames & Whitfield, 2003).

Studies beyond aphasia research suggest that motivation plays a substantial role in improvements on the training task, and consequently, outcome measures assessing near and far transfer effects (see also the section on WM training in healthy populations). For example, the lack of improvement in the training task as well as dropout from the study were related to the lack and instability of interest in the training task throughout the training, respectively (Jaeggi et al., 2011, Jaeggi et al., 2014). In aphasia, Brady and colleagues (2016) reported higher dropout for groups receiving high-intensity compared to low-intensity speech and language therapy. Based on these results, one can assume that motivation plays an important role in treatment outcomes and, especially, dropout from treatment in high-intensity treatments (i.e., five times a week), such as treatments of WM (for recommendations for dose and intensity of treatment in healthy participants, see Shipstead et al., 2012). Nonetheless, to my knowledge, no study has incorporated measures of motivation in WM treatment studies in aphasia.

Individual differences in initial cognitive abilities have received little attention in WM training studies both in healthy populations and IWA. As described above, studies in healthy participants suggest that those who start with poorer WM (i.e., more room to improve) benefit the most from WM training, both in terms of training (Jaeggi et al., 2011) and transfer effects (Au et al., 2015; Turley-Ames & Whitfield, 2003). In aphasia, although initial cognitive profile was taken into account in the majority of studies when selecting participants (e.g., Francis et al., 2003; Harris et al., 2014; Kalinyak-Fliszar et al., 2011; Koenig-Bruhin & Studer-Eichenberger, 2007; Mayer & Murray, 2002; Murray et al., 2006; Salis, 2012; Vallat et al., 2005), the potential role of WM abilities in *transfer effects* has not been investigated in any studies. Similarly, no study has specifically tested whether initial language abilities are associated with the extent of transfer to these abilities (e.g., sentence repetition, spoken sentence comprehension). Based on the authors' reports, we know that the majority of participants involved in WM treatment studies presented with moderate to mild aphasia (for the only study involving a participant with severe aphasia, see Salis, 2012). However, it is unclear whether initial language abilities have an influence on how much IWA benefit from WM treatment.

Therefore, in our studies we explored the potential role of motivation in WM treatment studies in aphasia, and made an initial attempt to elucidate cognitive and linguistic factors that may influence transfer effects.

1.5 The use of the *n*-back task in aphasia

N-back is a complex WM task involving multiple processes, such as encoding incoming stimuli, monitoring, maintaining, and updating WM representations, establishing and maintaining bindings between memory contents and their temporal context, as well as resolving interference between WM representations (Kane et al., 2007). In a typical *n*-back task, participants are presented with a continuous stream of items (e.g., numbers, letters, pictures) and are instructed to judge whether an item matches a previous one that was presented n items (e.g., n=1, n=2) before. WM load in the task increases with n. Some studies investigating the effect of "lures" (distractor items at n-1 or n+1 position) in the task suggest that lures additionally challenge attention control over familiarity-based interference (Hussey et al., 2017; Kane et al., 2007; Novick et al., 2014).

N-back has strong face validity as a WM task (Kane et al., 2007; Wright & Fergadiotis, 2012), but the fact that it shows only modest to weak correlations with typical WM tasks, such as the complex span suggests that its construct validity is not well established (for data on construct validity collected in healthy participants, see (DeDe, Ricca, Knilans, & Trubl, 2014; Kane et al., 2007). Results regarding its test-retest reliability are mixed, varying across *n*-s: while some studies have reported good reliability for 1-back (e.g., DeDe et al., 2014) and poor reliability for 2-back (DeDe et al., 2014; Hockey & Geffen, 2004) in young and old adults (Hockey & Geffen, 2004; DeDe et al., 2014, respectively), others have reported excellent reliability for their *n*-back in *aphasia* (Mayer & Murray, 2012). Note that the only study investigating test-retest reliability for *n*-back in aphasia pooled the data in the different conditions of the task (Mayer & Murray, 2012), therefore, conclusions about the task's reliability cannot be drawn in this population.

Several studies have measured working memory in IWA using *n*-back tasks, because this task does not require verbal responses (i.e., more language-free in nature; Christensen & Wright, 2010; Mayer & Murray, 2012) and the task structure is easy to convey and the administration is simple and in most cases automatized. In aphasia, most *n*-back studies have used pictures as stimuli (e.g., Christensen & Wright, 2010; DeDe et al., 2014; Harnish et al., 2018; Mayer & Murray, 2012; Zakariás, Keresztes, Demeter, & Lukács, 2013; Zakariás et al., 2013). Studies using *n*-back have reported significantly worse performance in IWA than neurologically intact control participants, suggesting that IWA demonstrate WM deficits (but see Friedmann & Gvion, 2003 for intact performance in the task). In addition, some studies have shown that IWA are significantly more affected by increasing WM load than healthy participants, as demonstrated by steeper performance decrements across levels of *n* in the aphasic than in the control group (Mayer & Murray, 2012; Zakariás et al., 2013; but see Christensen et al., 2018 for parallel performance patterns across n-s in the two groups). Concerning the stimulus type, IWA perform better on tasks involving nameable stimuli, such as objects and fruits than non-nameable stimuli, such as faces and computer-generated tones (Christensen & Wright, 2010; Mayer & Murray, 2012; Zakariás et al., 2013). Interestingly, a recent study has shown that a group with 14 IWA with different types and severities of aphasia did not demonstrate difficulties in spatial *n*-back, suggesting that WM deficits in aphasia are not necessarily present in the spatial domain (Christensen et al., 2018).

Wright and colleagues (2007) investigated the relationship between WM and spoken sentence comprehension in nine individuals with different types and severities of aphasia. Using an *n*-back task including five-word sentences with different syntactic structures as stimuli, they found a significant relationship between performance in 2-back and the comprehension of non-canonical sentences, such as object relative clauses and passive sentences (Wright et al., 2007). The authors concluded that separate WM systems are involved in language comprehension and their involvements depend on the type of linguistic information (i.e., phonological, syntactic, and semantic) being processed. Also suggesting an association between updating WM representations and spoken sentence comprehension, Szöllősi, Lukács, and Zakariás (2015) reported a marginally significant correlation between performance in *n*-back with letters and the comprehension of grammatical structures as measured by the Test for the Reception of Grammar (Bishop, 1989; Hungarian adaptation: Lukács, Győri, & Rózsa, 2012).

In summary, the fact that *n*-back relies on binary responses and does not require spoken output, but only button presses makes it very suitable for IWA who find repetition and verbal output in general difficult. In addition, its administration is simple and in most cases automatized, which may enhance research validity and treatment fidelity (i.e., the reliability of the administration of an intervention) in studies using *n*-back as a training task in aphasia.

1.6 Rationale for the studies included in the thesis

As discussed in the previous sections, treatments focusing on WM without targeting language processing may be beneficial to language functioning (i.e., far transfer effects). Accumulating evidence for the positive effects of WM treatments on spoken sentence comprehension are promising, however, further investigation is warranted given methodological issues in prior research. In addition, of particular interest to aphasia, it is poorly understood what underlying mechanisms lead to improvements on spoken sentence comprehension and, relatedly, what aspects of spoken sentence comprehension (i.e., online processing of syntactic comprehension, post-interpretative comprehension) can benefit from WM treatments. Identifying such mechanisms may help us understand more about the relationship between WM and spoken sentence comprehension, as well as the spoken sentence comprehension deficit in this population.

Furthermore, although it has been previously suggested that WM-related functions are engaged in everyday communication (hereinafter referred to as functional communication) in aphasia, we know little about the effects of WM treatments in this domain. The few studies that have included outcome measures on functional communication revealed inconclusive findings. This calls for future investigations into WM treatments and their effects on functional communication. Improving communication is the main goal of linguistic rehabilitation for most IWA, thus understanding the effects of WM treatments on this domain is critical in this population.

It is also important that future research focuses on interference control, an ability that has been often overlooked in studies of WM treatments in aphasia. The fact that interference control has been shown to support both spoken sentence comprehension and functional communication underscores the importance of incorporating interference control into WM treatments and investigating transfer effects after such treatments in aphasia.

The aforementioned ambitious goals can only be achieved by a systematic investigation of transfer effects after WM treatment. Such investigations may involve a careful choice of outcome measures that are sensitive enough to detect potential improvements (near and far transfer effects) following WM treatment, as well as a formulation of clear hypotheses about the underlying mechanisms of such improvements. In addition, there is also a clear need for investigating the patterns of training and transfer effects following WM treatment (i.e., inter- and intra-individual differences in training and transfer effects), which has not been frequently done in WM treatment studies including IWA. Understanding these patterns may help understand the underlying mechanisms behind and principles of transfer effects in aphasia, and thus, the relationship between WM and domains of far transfer, such as spoken sentence comprehension and functional communication.

The studies constituting this thesis used novel manipulations and experimental designs to systematically investigate near and far transfer effects after WM treatment in aphasia (Study 1 and 2). In addition, it includes a first systematic review in the field, investigating the methodological quality of studies testing transfer effects following WM treatment in aphasia (Study 3).

2 Aims, research questions, and hypotheses

As discussed above in the general introduction, previous research suggests that WM can be enhanced with practice in aphasia. However, the extent to which improvements in WM lead to improvements on untrained WM tasks, language (e.g., spoken sentence comprehension, functional communication), and other everyday functions (i.e., near and far transfer effects) is poorly understood.

In **Study 1** (Chapter 5), our aim was to investigate whether a complex WM training focusing on the storage as well as the executive components of WM (i.e., updating WM representations and interference control) leads to improvements on untrained tasks and abilities (i.e., transfer effects) in IWA. Questions of Study 1 were:

(1) Can WM be improved through an adaptive *n*-back training in IWA?
(2) Does the training lead to near transfer effects on WM measured by tasks not practiced during training?
(3) Does the training lead to far transfer effects on spoken sentence comprehension?

To answer these questions, we conducted a pre-post case-controls study (Crawford & Garthwaite, 2005) with three Hungarian-speaking individuals with post-stroke aphasia and a matched control group with five IWA. Participants in the training group practiced an adaptive *n*-back task with letters modified for use with aphasia. 'Adaptivity' involved adjusting the task's difficulty level according to the participants' performance, ensuring that they always practiced at an optimal level of difficulty. Control participants received no training (see Jaeggi et al., 2008; Novick et al., 2014) but underwent the same assessments in the same timeframe as participants in the training group. Outcome measures included two experimental WM tasks measuring near transfer effects, and a standardized spoken sentence comprehension test measuring far transfer effects. A control task putting some demands on WM (oral naming, see N. Martin

& Saffran, 1997) was also included to test that possible improvements on the outcome measures were specifically related to the training. We expected trained participants to improve on the training task as well as on outcome measures of WM and spoken sentence comprehension. We expected no improvement or at least less improvement on the control task following training compared to outcome measures assessing near and far transfer effects.

In **Study 2** (Chapter 6), based on the results of Study 1, our main objective was to systematically investigate patterns and potential domains of transfer after *n*-back training in IWA, and to test for the ecological validity of potential findings. Questions of Study 2 were:

(1) Does WM training lead to near transfer effects on cognitive domains targeted by the training but measured by untrained tasks in IWA?
(2) Does WM training lead to far transfer effects on spoken sentence comprehension, functional communication, and everyday memory in IWA?
(3) Are training and transfer effects maintained over time (i.e., at 4-6 weeks follow-up)?
(4) Do motivational factors play a role in IWAs' WM training performance?

To this end, we conducted a multiple-baseline study with three German-speaking individuals with chronic post-stroke aphasia. Participants practiced two adaptive *n*-back tasks (one with pictures and one with spoken words). To systematically investigate potential transfer effects (i.e., the levels of transfer), we chose a set of outcome measures ranging from the training task (*n*-back) to far transfer (e.g., spoken sentence comprehension, functional communication). To extend the ecological validity of findings, we included a broad set of outcome measures, such as tests of functional communication, and everyday memory. Changes in functional communication were assessed by the Amsterdam-Nijmegen Everyday Language Test (ANELT; Blomert, 1992), for the first time in WM treatment studies in aphasia. To assess the specificity of transfer effects and to better understand the underlying mechanisms of transfer on spoken sentence comprehension, we included an outcome measure testing specific syntactic structures that have been proposed to involve WM processes (e.g., non-canonical structures with varying complexity; Caplan et al., 2013; Just & Carpenter, 1992; Haarmann et al., 1997; Thompson-Schill et al., 2005; Wright et al., 2007). A control task (oral word reading) was also included to test that potential improvements on the outcome measures were specifically related to the training. To capture the effects of motivational factors on training performance across time, we monitored participants' motivation on a daily basis by using a self-report questionnaire. We expected participants to improve on the training tasks as well as on outcome measures of memory, spoken sentence comprehension, and functional communication. We expected no improvement or at least less improvement on the control task following training compared to outcome measures assessing near and far transfer effects. In addition, we hypothesized that stable and generally high levels of interest (a factor of motivation) would be associated with greater improvement in the training task.

In **Study 3** (Chapter 7), we aimed to identify and describe WM treatments in stroke aphasia through a systematic review of relevant literature, as well as to appraise the methodological quality of studies describing these treatments and identify abilities that can benefit from WM training (i.e., domains of near and far transfer). Questions of Study 3 were:

(1) What treatments of WM have been used in stroke aphasia?
(2) What is the internal and external validity of WM treatments in stroke aphasia?
(3) Which cognitive, linguistic, and everyday functions can benefit from WM treatments in stroke aphasia?

To answer these questions, we conducted a systematic search of 13 relevant electronic databases, ending in 2016 December. For each included study, we extracted information about the study method, participants' characteristics, treatment procedure and setting, outcome measures, and treatment outcome. To assess the methodological quality of included studies, we used the Risk of Bias in N-of-1 Trials (RoBiNT, Tate et al., 2015) quantitative scale, which is designed to evaluate the internal and external validity of studies applying single-case experimental designs.

3 Summary of methods and main findings

Main findings of Study 1 and 2 are summarized in Table 1. In **Study 1**, three Hungarian-speaking IWA with moderate spoken sentence comprehension deficit and verbal STM deficit practiced an adaptive WM task (*n*-back with letters) three to four times a week for a month (altogether for 13 sessions). Their performance was assessed before and after the training on two experimental WM tasks assessing near transfer effects and a standardized spoken sentence comprehension test assessing far transfer effects. For all outcome measures, each individual's pretest–posttest differences were compared to the mean pretest–posttest difference of the control group (Crawford & Garthwaite, 2005). To our knowledge, this is the first WM treatment study in aphasia using such a case-controls design. This is particularly important because case-controls designs considerably increase the internal validity of any treatment effects found, and thus it is highly recommended in single-case and case series studies, such as in intervention studies with IWA (Crawford & Garthwaite, 2005). Overall, we detected a mixed pattern of training and transfer effects across participants. One participant improved in the training task as well as on an untrained WM task (i.e., near transfer effects) and in spoken sentence comprehension (i.e., far transfer effects). Another participant showed a tendency for improvement on the training task and significantly improved in spoken sentence comprehension but did not show improvements on untrained WM tasks. The third participant did not show improvement in the training task but did show increases in performance both in sentence comprehension and in untrained WM tasks. Participants did not show significant performance changes in the control task, suggesting that effects detected in the outcome measures were specific to the training. Taken together, these results suggest that WM can be improved through an intensive and adaptive *n*-back training in aphasia, and these improvements can lead to improvements in spoken sentence comprehension in some individuals.

In **Study 2**, three German-speaking IWA with moderate spoken sentence comprehension deficit and verbal WM deficit practiced two computerized WM tasks (*n*-back with pictures and *n*-back with spoken words) three-four times a week for four-five weeks (altogether 16 sessions), in a counterbalanced order. Their performance was assessed before and after the training on WM tasks (i.e., near transfer effects), as well as on tests of spoken sentence comprehension, functional communication, and everyday memory (i.e., far transfer effects). Similar to Study 1, we detected a mixed pattern of training and transfer effects across participants. One participant improved on both training tasks, as well as in a WM task, in tests of spoken sentence comprehension and in functional communication. She also showed a tendency for improvement in speech-related everyday memory activities. Another participant improved only in one training task (*n*-back with pictures) and showed improvements on a WM task and tests of spoken sentence comprehension. She also showed a statistical tendency for improvement in speech-related everyday memory activities. The third participant also improved only in one training task (i.e., *n*-back with pictures), and showed improvements on tests of spoken sentence comprehension, functional communication, and everyday memory (i.e., learning new things), but did not show improvements on untrained WM tasks. Participants' performance remained stable on the control task, suggesting that improvements on the outcome measures were related to the training. Motivation questionnaires showed a moderate to high interest in the training tasks for two participants, which remained stable throughout the training. The third participant, however, showed a considerable fluctuation in all factors of motivation that were tested in the study. In addition, motivation was linked to training performance in two participants: factors of motivation, such as perceived competence and effort showed a relationship with performance in the second training for one participant and the first training for another participant, respectively. In summary, results of Study 2, similar to Study 1, suggest that WM can be improved through intensive and adaptive *n*-back training in aphasia, and these improvements can transfer to spoken sentence comprehension and functional communication in some individuals.

Table 1 ***Overview of main findings of Study 1 and 2***

	Case	Improvements in the training tasks	Near transfer observed in	Far transfer observed in
Study 1	K.K.	*N*-back with letters	1-back and 2-back with letters	Spoken sentence comprehension (TROG)
	B.L.	*N*-back with letters	–	Spoken sentence comprehension (TROG)
	B.B.	–	1-back with letters	Spoken sentence comprehension (TROG)
Study 2	V.O.	*N*-back with spoken words *N*-back with pictures	2-back and 3-back with letters	Spoken sentence comprehension (canonical structures) Functional communication (ANELT Understandability, %CIUs) Everyday memory (speech-related activities)
	R.D.	*N*-back with pictures	2-back and 3-back with letters	Spoken sentence comprehension (number-marked OVSs and non-canonical structures) Everyday memory (speech-related activities)
	Z.A.	*N*-back with pictures	–	Spoken sentence comprehension (TROG) Functional communication (ANELT number of words, CIUs) Everyday memory (learning new things)

Note. TROG = Test for the Reception of Grammar; ANELT = Amsterdam-Nijmegen Everyday Language Test; %CIUs = percent of correct information units; OVS = object-verb-subject sentences; CIUs = number of correct information units.

In **Study 3**, we conducted a systematic review to describe studies of WM treatment in stroke aphasia, appraise the internal and external validity of these treatments, and identify potential domains of transfer, that is, abilities that can benefit from WM treatments. The systematic search and inclusion/exclusion procedure yielded 17 studies (mainly single case or case-series designs) with overall 37 participants for inclusion. Treatment in nine studies targeted verbal short-term memory, whereas treatment in eight studies focused on attentional and executive components of WM. 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. Seventy-three percent of the individuals presented with mild or moderate aphasia; only 8% had severe aphasia. Internal validity was poor across studies, with a score between 0-8 (mean = 1.6) on a 0–14 scale. Similarly, external validity scores, although higher than for internal validity, were still low, ranging from 4-12 (mean = 7.9) on a 0-16 scale. Because of the methodological limitations of the studies included in the review, any conclusions regarding the beneficial effects of WM treatments (see below) have to be viewed with caution. Ninety-four percent of the studies (16/17) that investigated effects on WM reported improvements in this domain. Seventy-seven percent of the studies (7/9) 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. The two studies that assessed effects on reading comprehension reported promising results: Higher accuracy and increased reading rate in paragraph reading tasks after treatment. Two studies reported nominal improvements in spoken discourse, using a picture description task. Among the four studies using a communication questionnaire, only two was successful in demonstrating improvements in some of their participants. Results in measures of cognitive functioning in everyday life (i.e., WM and attention questionnaires) were variable across studies. Two studies reported improvements in WM and attention, whereas two studies did not report any changes in these domains. Taken together, methodological limitations make it difficult, at present, to draw firm conclusions about the effectiveness of WM treatments in stroke aphasia. Results in terms of WM, spoken sentence comprehension, and reading are promising, but further studies with more rigorous methodology and stronger experimental control are needed to determine the beneficial effects of this type of intervention. Future studies need to include outcome measures of memory functioning in everyday life, communication, and psychosocial functioning to demonstrate clinically significant improvements after WM treatments in aphasia.

4 Conclusions and future directions

The aim of this thesis was to improve our understanding of the relationship of WM and language abilities in aphasia. Our specific goals were to investigate whether WM can be improved through an adaptive *n*-back training in aphasia (Study 1 and 2), and to test whether WM improvements lead to near transfer on unpracticed WM tasks (Study 1, 2, and 3) and far transfer on spoken sentence comprehension (Study 1, 2, and 3), functional communication and other everyday abilities (Study 2 and 3) in IWA. Furthermore, we also aimed at appraising the internal and external validity of existing WM treatments in IWA and identifying cognitive and linguistic abilities, and other everyday functions that can benefit from WM treatments in this population (Study 3). To address our goals, we conducted two empirical studies of WM training involving six IWA in total, and a systematic review of available STM/WM treatments in chronic post-stroke aphasia.

4.1 Patterns of transfer after working memory training

We detected a mixed pattern of training and transfer effects across individuals: five participants out of the six significantly improved in the *n*-back training. Our most important finding is that all of them also improved significantly in spoken sentence comprehension (i.e., far transfer effects). In addition, we also found far transfer to functional communication (in two participants out of three in Study 2) and everyday memory functioning (in all three participants in Study 2), and near transfer to unpracticed *n*-back tasks (in four participants out of six). The lack of improvement on a structurally different WM task measuring the hypothesized same cognitive construct as the *n*-back (i.e., running span) suggests that near transfer effects were task-specific rather than process-specific. Taken together, we detected both near far and transfer effects in our studies, but the effects varied across participants.

Unlike a chain of transfer reported in earlier WM training studies in healthy participants (i.e., participants showing near transfer effects show improvements in the training, and participants showing far transfer effects show near transfer effects, see Jaeggi et al., 2014; Waris, Soveri, & Laine, 2015), transfer effects in our studies did not follow a clear pattern in some participants. More specifically, one participant showed near and far transfer effects (i.e., improvements on unpracticed WM and language tasks, respectively) without showing improvement on the training task, and two participants showed far transfer effects without showing near transfer effects. In addition, these two participants showed only a modest improvement in the training. Note that these inconsistent patterns seemingly contradict the principle of 'overlapping cognitive and neural mechanisms'. Based on this principle, one would predict a chain of transfer where training effects are a prerequisite of the emergence of near transfer effects, and training *and* near transfer effects are a prerequisite of the emergence of far transfer effects.

However, the inconsistent patterns observed in our studies are in line with the observation that WM training-induced language improvements do not always occur in the presence of improvements on the trained WM task in aphasia (Minkina et al., 2017). Identifying the underlying mechanisms behind these inconsistent patterns is crucial for understanding the cognitive mechanisms of transfer effects and interpreting findings of WM treatment studies in IWA. In the studies presented in the thesis, one such mechanism can be related to the fact that our training tasks were complex WM tasks involving several cognitive mechanisms. It may be possible that improvements detected in spoken sentence comprehension and functional communication after training were not primarily induced by improvements in updating WM representations, which was measured by the training tasks and the WM outcome measures, but by improvements of other aspects of WM-functioning (e.g., resolving interference and/or attention functions). Improvements of such non-linguistic cognitive functions after training may have gone undetected in our pre-post assessments. Future studies could include broader pre-post assessments of interference control, attention, and verbal STM to detect improvements that may lead to transfer to spoken sentence comprehension and functional communication.

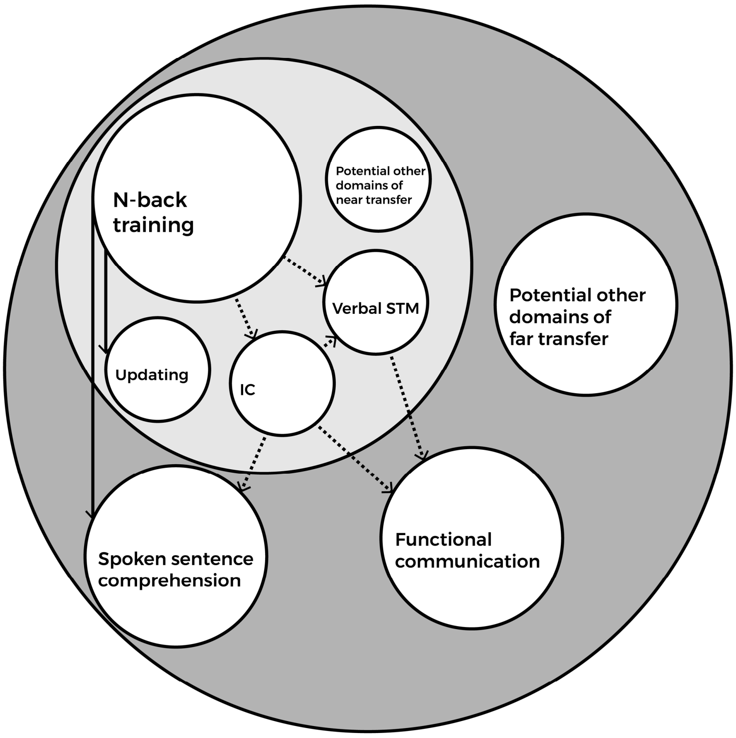

Figure 2 An illustration of levels and possible "routes" of transfer, proposed based on results of Study 1 and 2
This figure is a schematic illustration of the complexity of potential mechanisms underlying transfer effects following WM training. The aim of this illustration is to inspire future research trying to understand these mechanisms and to provide guidelines for designing new interventions. Continuous arrows represent the expected (and measured) routes of transfer from the training task (*n*-back) through updating tasks (i.e., near transfer) to spoken sentence comprehension and functional communication (i.e., far transfer). Dashed arrows represent additional possible routes of observed transfer in some individuals. These speculative routes are based on theories suggesting functional and neuronal associations between components of WM – interference control (IC), verbal STM – and language processing. Improvements in IC due to *n*-back training may have transferred to spoken sentence comprehension (Hussey & Novick, 2012; Thompson-Schill et al., 2005) and functional communication (Penn et al., 2010). Alternatively, improvements in the training tasks may have transferred to verbal STM, and in turn, functional communication (N. Martin et al., 2018). Note that a relationship is also suggested between IC and verbal STM (N. Martin et al., 2012; R. C. Martin, 2007), allowing for transfer effects also between these components. The speculated associations highlight the need for including broad baseline assessments of linguistic and non-linguistic cognitive abilities in future WM treatment studies in aphasia.

4.2 The role of individual differences in transfer effects

Overall, our results are in line with previous studies detecting large individual differences in training and transfer effects in healthy participants (Jaeggi et al., 2014) and IWA (Lee & Sohlberg, 2013; Peach et al., 2017). To shed light on potential factors influencing transfer effects, we investigated the relationship between initial WM and language abilities, and improvement on spoken sentence comprehension after training. Initial spoken sentence comprehension was associated with the extent of transfer to spoken sentence comprehension, suggesting that the more severe the participants' spoken sentence comprehension deficit was at the beginning of training, the more it improved after training. Contrary to our expectations, we did not find a relationship between initial WM and the extent of transfer to spoken sentence comprehension (Fedorenko, 2014). This may be due to methodological properties of our analyses: The small sample size and the fact there was only one task available (*n*-back) to calculate measures of initial WM (and cognitive abilities in general). Calculating multiple WM measures using various WM tasks (e.g., simple and complex span tasks) would potentially increase the chances of identifying factors that modulate transfer effects (for a related issue, see the mono-method bias; Coolican, 2014). Future studies could extend their investigation of the role of initial cognitive abilities, such as WM (Au et al., 2015; Harnish et al., 2018; Morrison & Chein, 2011), interference control and attention (Geranmayeh, Brownsett, & Wise, 2014; Morrison & Chein, 2010) in transfer effects in IWA by including a broader assessment of these functions.

Future studies could also include tests of training-related structural changes in the brain to further our understanding of underlying neural mechanisms of transfer effects in aphasia. These studies could also test the hypothesis that transfer effects occur in the presence of structural changes in the brain (for the potential time course of training-induced behavioral and structural changes, see Kühn & Lindenberger, 2016; Lindenberger, Wenger, & Lövdén, 2017). A clinically highly relevant prediction derived from this hypothesis is that structural plasticity in stroke aphasia (see Lukic et al., 2017; Xing et al., 2016 for evidence of structural plasticity in stroke aphasia) makes individuals particularly responsive to WM training, resulting in greater training gains and transfer effects in this population compared to healthy individuals. This needs to be tested in future studies.

4.3 Effects of working memory training on syntactic comprehension

With respect to the effects of WM training in spoken sentence comprehension and in particular, syntactic comprehension, we detected improvements in (1) non-canonical number marked (object-verb-subject) sentences[3], (2) non-canonical sentences including varying syntactic structures, such as case marked and number marked object-verb-subject sentences and right-branching[4] and center-embedding object relative clauses, and (3) canonical sentences including case marked and number marked subject-verb-object sentences and right-branching and center-embedding subject relative clauses[5] in some individuals. The fact that improvements were not specific to certain syntactic structures (i.e., non-canonical complex sentences) suggests that the involvement of WM in spoken sentence comprehension is not syntax-specific as suggested by Just and Carpenter (1992), but WM may play a more general role in the later stage of spoken sentence comprehension (i.e., expanded comprehension) by providing extra resources for the revision of previously encountered inaccurately interpreted information (e.g., Caplan et al., 2013). Because of the small number of participants in our studies, this conclusion has to be viewed with caution and needs to be tested in future studies.

4.4 Long-term maintenance of improvements

Results of Study 2 show that improvements in spoken sentence comprehension were maintained at 6-week follow-up in one participant. Improvements that were present only at follow-up but not at the posttest (i.e., “sleeper effects”; Jaeggi et al., 2014) for another participant are difficult to interpret (for similar results in other populations, see Holmes, Gathercole, & Dunning, 2009; Van der Molen, Van Luit, Van der Molen, Klugkist, & Jongmans,

---

[3] E.g., Die Tanten sucht das Kind
The aunts$_{\text{ACC}}$ are looking for the child$_{\text{NOM}}$
“The child is looking for the aunts”

[4] E.g., Das ist der dicke Vater, den der Sohn sucht
That is the fat father who$_{\text{ACC}}$ the child$_{\text{NOM}}$ is looking for
“That is the fat father the child is looking for”

[5] E.g., Der Vater, der den Sohn sucht, ist dick
The father, who$_{\text{NOM}}$ the child$_{\text{ACC}}$ is looking for, is fat
“The father, who is looking for the child, is fat”

2010). Results of Study 2 also suggest that gains achieved in the training tasks (*n*-back with pictures and *n*-back with spoken words) were not consistently maintained until 6-weeks after posttest. Note that variable performance between posttest and follow-up may also be due to methodological features of the study design and task procedure: The small number of target items (n = 15) included in the outcome measures used to detect maintenance of improvements in the *n*-back tasks could have lead to low sensitivity and poor reliability in the task. Therefore, results regarding maintenance of improvements in WM need to be viewed with caution.

4.5 Feasibility and psychometric properties of *n*-back in aphasia

Results of Study 1 and 2 showed that participants performed better on the visual version (*n*-back with letters and pictures) than on the auditory version (*n*-back with computer-generated tones and spoken words) of the task. The fact that participants performed neither at ceiling nor at floor (cf. DeDe et al., 2014) suggests that visual 1-back and 2-back are likely to be feasible for use with moderate to mild aphasia. However, we recommend optimizing task parameters, such as timing (e.g., interstimulus interval, ISI) and stimulus type on a study-by-study basis.

In Study 1, 1-back and 2-back with letters, and 1-back with computer-generated tones showed acceptable test-retest reliability (r = .71, r = .75, and r = .77, respectively), whereas 2-back with computer-generated tones showed poor test-retest reliability (r = .28). Our results are not consistent with findings of Meyer and Murray (2012) reporting excellent test-retest reliability for their *n*-back tasks in aphasia. Variable test-retest reliability may be due to the different measures calculated in the tasks (i.e., measures calculated separately for 1-back and 2-back vs. measures calculated for the pooled data of the two conditions, see also DeDe et al., 2014). Here we argue that it may also be due to differences in task procedures and stimulus materials. Higher task demand elicits higher degrees of intra-individual differences in attention tasks (see Villard, 2016). Given the fact that Meyer and Murray (2012) applied a longer ISI (i.e., 1600ms) in their task than we did (i.e., 1000ms), it may be possible that Meyer and

Murray's *n*-back task was less demanding than the one in Study 1. Longer ISI compared to shorter ISI in WM tasks can allow for more successful updating (and removal of no longer relevant items from WM, see Oberauer et al., 2012) and consequently, make the task less demanding. A similar logic can be applied to explain poor test-retest reliability in 2-back with computer-generated tones as well. As non-nameable stimuli have been shown to be significantly more demanding for IWA than nameable stimuli (Christensen & Wright, 2010; Mayer & Murray, 2012; Zakariás et al., 2013), higher task demand due to the type of stimuli could have led to less stable performance in this task. Future studies should focus on systematically investigating the role of stimuli (nameable vs. non-nameable) and timing (short vs. long ISI) to ensure feasibility and reliability of the task in individuals with a wide range of severities of aphasia.

4.6 Methodological quality of existing working memory treatments

In Study 3, we investigated the methodological quality of existing WM treatments in post-stroke IWA and we found that internal and external validity were poor across the included 17 studies. One of the most important features regarding internal validity of studies relates to the design (Tate et al., 2015). Almost half of the included studies, although they were case studies, instead of using a single-case methodology, rather, employed a pre-post intervention design. 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 et al., 1997; Tate et al., 2015). In contrast to pre-post intervention design, single-case methodology allows for participants to serve as their own controls by the systematic manipulation of the intervention and assessments. Thus, single-case methodology should be used in future studies evaluating treatments effects at the individual level.

Such important methodological limitations highlighted in Study 3 make it difficult to draw conclusions about the beneficial effects of WM treatments in stroke aphasia. Overall, participants showed improvements in the treatment tasks, suggesting that WM functions can be improved in stroke aphasia. Improvements were also noted on unpracticed WM tasks in the majority of studies (i.e., near transfer). Around three quarters 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 communication; improvements in these domains were reported in some but not all studies. Studies rarely included patient-reported outcome measures of everyday life and psychosocial functioning (e.g., questionnaires assessing everyday life problems related to deficits of WM). When they did, results were inconclusive. As we noted before, due to low methodological validity of the reviewed studies, any conclusions regarding the positive effects of treatments have to be viewed with caution.

To our knowledge, this is the first series of studies specifically incorporating aspects of inhibition-related functions, such as interference control in WM treatment in IWA. The individual differences in treatment outcomes call for future research to clarify how far these results are generalizable to the population level of IWA. Future studies are needed to identify a few mechanisms that may generalize to at least a subpopulation of IWA as well as to investigate baseline non-linguistic cognitive and language abilities that may play a role in transfer effects and the maintenance of such effects. These may require larger yet homogenous samples. Taken together, our results suggest that WM can be improved through adaptive *n*-back training in aphasia, and these improvements can transfer to spoken sentence comprehension and functional communication in some individuals.

ORIGINAL JOURNAL ARTICLES

5 Study 1: Positive effects of a computerized working memory and executive function training on sentence comprehension in aphasia[6]

Abstract

Aphasia, the language disorder following brain damage, is frequently accompanied by deficits of working memory (WM) and executive functions (EFs). Recent studies suggest that WM, together with certain EFs, can play a role in sentence comprehension in individuals with aphasia (IWA), and that WM can be enhanced with intensive practise. Our aim was to investigate whether a combined WM and EF training improves the understanding of spoken sentences in IWA. We used a pre-posttest case control design. Three individuals with chronic aphasia practised an adaptive training task (a modified *n*-back task) three to four times a week for a month. Their performance was assessed before and after the training on outcome measures related to WM and spoken sentence comprehension. One participant showed significant improvement on the training task, another one showed a tendency for improvement, and both of them improved significantly in spoken sentence comprehension. The third participant did not improve on the training task, however, she showed improvement on one measure of spoken sentence comprehension. Compared to controls, two individuals improved at least in one condition of the WM outcome measures. Thus, our results suggest that a combined WM and EF training can be beneficial for IWA.

---

[6] This chapter has been published as:
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

5.1 Introduction

Verbal working memory (WM) is a complex cognitive construct, referring to processes that underlie the temporary maintenance and manipulation of linguistic information (Engle, 2002; Jaeggi et al., 2014; N. Martin et al., 2012). According to prominent models of WM (e.g., Baddeley, 2003; Miyake, Friedman, et al., 2000), these processes involve various executive functions (EFs), such as shifting between tasks or mental sets, updating and monitoring WM representations, and inhibiting prepotent responses (Miyake et al., 2000).

Aphasia can result from damage to different brain regions often overlapping with the regions associated with WM and EFs (Alexander, 2006; Freedman, Alexander, & Naeser, 1984). Consequently, many studies have shown that aphasia is frequently accompanied by WM and EF deficits (e.g., Helm-Estabrooks & Albert, 1991; Nickels, Howard, & Best, 1997; Purdy, 2002; Zakariás et al., 2013). Moreover, impairments of WM and EFs can negatively influence various language processes in aphasia, including lexical-semantic processing (N. Martin et al., 2012; Novick et al., 2009; Robinson et al., 1998), syntactic processing (Haarmann et al., 1997; Novick et al., 2009), communication (Frankel et al., 2007; Ramsberger, 2005), as well as the recovery pattern in aphasia (Green et al., 2010; Lambon Ralph, Snell, Fillingham, Conroy, & Sage, 2010; Penn et al., 2010).

5.1.1 Working memory, executive functions, and sentence comprehension in aphasia

A number of studies suggest that there is a relationship between WM and/or EF impairments and sentence comprehension deficits found in aphasia (Caspari et al., 1998; Haarmann et al., 1997; Sung et al., 2009). These studies included individuals with different types of aphasia, such as non-fluent Broca's aphasia (Ivanova, Dragoy, Kuptsova, Ulicheva, & Laurinavichyute, 2015), dynamic aphasia, a form of transcortical motor aphasia (Novick et al., 2009, Robinson et al., 1998), and fluent conduction aphasia (Gvion & Friedmann, 2012; R. C. Martin & He, 2004; R. C. Martin, Shelton, & Yaffee, 1994).

In general, WM, implicated in maintaining linguistic information over a short period of time, has been shown to play a role in processing complex sentences. For instance, in semantically complex sentences, WM is required to simultaneously maintain several individual word meanings (e.g., Martin & He, 2004). In syntactically complex sentences, such as passives (e.g., *'The rat was hit by the dog'*) or object relative clauses (e.g., *'It was the rat that the dog hit'*), WM supports the initial assignment of the preferred structure (i.e., processing who-does-what-to-whom, e.g., Haarmann et al., 1997) or supports the reanalyses of previously encountered, inaccurately interpreted information (Caplan & Waters, 2013; R. C. Martin, 1993).

Further, EFs, and interference control in particular (Friedman & Miyake, 2004), have also been shown to play a role in sentence processing in aphasia (Novick et al., 2009). These terms, although referring to largely overlapping cognitive constructs, are somewhat inconsistently used in the literature. Based on Friedman and Miyake's model (2004), we use the term interference control to refer to the ability to resolve interference from irrelevant, distracting information, and more specifically to resist memory intrusions from information that was previously presented but is irrelevant to the current task.

In sentences with linguistic ambiguities (e.g., garden-path sentences with syntactic ambiguities, such as '*Put the apple on the napkin into the box'*), interference control is necessary to suppress the context-inappropriate meaning of the ambiguous information (e.g., Novick et al., 2010; Thompson-Schill et al., 2005). In non-canonical sentences, such as passives and object relative clauses (see examples above), interference control is needed to inhibit the dominant interpretation of word order to build the correct syntactic structure (Thompson-Schill et al., 2005).

After reviewing functional magnetic resonance imaging (MRI) studies, Geranmayeh and colleagues (2014) concluded that EFs mediate top-down processes acting on impaired domain specific language functions, such as syntax and semantics. Moreover, top-down cognitive control is linked to the left inferio-frontal gyrus (LIFG), a brain area that has also been related to auditory sentence comprehension in neurolinguistic studies (Ben-Shachar, Palti, & Grodzinsky, 2004; Friederici, 2002), and is often impaired in aphasia (Alexander, 2006).[7]

---

[7] There has been a longstanding debate about whether LIFG serves language-specific or domain-general functions (Ben-Shachar, Hendler, Kahn, Ben-Bashat, & Grodzinsky, 2003; Novick,

Thus, the interaction among WM, EFs, and sentence comprehension is maintained across cognitive and structural levels. Although the exact relationship between WM, EFs, and sentence comprehension is still not clear (for a review, see Caplan & Waters, 2013), the studies reviewed above highlight the potential importance of WM and EFs in sentence comprehension in aphasia.

5.1.2 Working memory and executive function training effects on sentence comprehension in aphasia

So far, no studies have specifically addressed the effects of EF training on sentence comprehension in aphasia. A few studies have investigated the effects of WM training on sentence comprehension, but the results are mixed. These studies measured the training effects on WM tasks that were similar to the training task (i.e., near transfer effects) and on sentence comprehension tasks that differed substantially from the training task (i.e., a form of far transfer effects).

Salis (2012) treated one participant with non-fluent transcortical motor aphasia using a WM training task with noun-stimuli. The effects of the training were measured with forward and backward digit spans (i.e., near transfer tasks), the Test for the Reception of Grammar (TROG, Bishop, 1989) and the Token Test (McNeil & Prescott, 1978) (i.e., far transfer tasks). Following treatment, both WM performance and sentence comprehension (as measured by the TROG) improved. Although these results suggest that WM training may yield significant transfer effects on syntactic processing, this study did not include a control group to exclude test-retest effects.

Another cognitive training study by Murray and colleagues (2006) examined the effects of attention training on auditory comprehension and untrained attention tasks. One participant with fluent conduction aphasia practised tasks of the Attentional Training Program II (ATP-II; Sohlberg, Johnson, Raskin, & Mateer, 2001), focusing on several aspects of attention involving WM

Trueswell, & Thompson-Schill, 2005, respectively). In an elegant fMRI study designed to resolve this controversy, Fedorenko, Duncan, and Kanwisher (2012) found that this area exhibits a high degree of functional heterogeneity: domain-general and language-specific regions lie side by side and, accordingly, one region is engaged in language processing whereas the other is broadly engaged in domain-general processes independently of task and content.

components. Although the participant improved on the training tasks, only nominal changes were observed on the untreated attention tasks measuring near transfer effects. With regards to the far transfer effects, improvement was only found in response speed, one measure of auditory comprehension. This finding contradicts the outcomes of Salis (2012). Murray and colleagues concluded that attention training can enhance specific attention abilities but cannot produce broader changes in other cognitive and language functions. However, the authors noted in their conclusions that ATP-II has been designed to remediate cognitive deficits associated with traumatic brain injury, thus, it contains verbal tasks that have not been adapted for individuals with aphasia (IWA). Alternatively, a lack of transfer effect in Murray and colleagues' study might also be related to the relatively rare training sessions (once a week, lasting 60 minutes).

Recently, Harris and colleagues (2014) treated two IWA, and used repetition and recognition tasks as training tasks. One participant (DS) showed semantic short-term memory impairment together with a pronounced sentence comprehension deficit, whereas the other participant (AK) had phonological short-term memory impairment and only a mild deficit in sentence comprehension at the beginning of the training. After the training, DS showed transfer effects on semantically anomalous sentence judgements and sentence picture matching. AK did not show any transfer effects, but, notably, his initial performance on sentence comprehension had almost reached ceiling before the treatment.

5.1.3 Unresolved issues

These studies raised several issues that motivated the present study. Although previous results suggest that WM can indeed be enhanced in aphasia (Harris et al., 2014; Salis, 2012), whether this enhancement leads to substantial improvement in sentence comprehension is still unclear. As shown above, some researchers reported far transfer effects of WM training on sentence comprehension (e.g., Harris et al., 2014; Salis, 2012), while others did not detect any effects of training on untreated processes (Murray et al., 2006). As the majority of these studies were either single-case studies (Francis et al., 2003; Salis, 2012) or included maximum two participants (Harris et al., 2014) and used varying training schedules (from one 90-min session per week over 10 weeks (Harris et al., 2014) to two 30-min sessions per week over 13 weeks (Salis, 2012)), it is difficult to generalize the findings.

More importantly, despite the suggested link between EFs and sentence comprehension, no study has examined the direct effects of EF training on sentence comprehension in aphasia. Last but not least, our study was motivated by the need to include a control group in aphasia training studies (Nickels, 2002). The inclusion of control participants in training studies of neuropsychological populations is a potential means to control for test-retest effects, and the effect of spontaneous recovery.

5.1.4 The present study

Our aim was to investigate whether a training focusing on WM (maintaining and updating WM representations) and EF processes (specifically interference control) leads to improvement on the same processes measured by tasks not practised during the training sessions (i.e., near transfer effects), and in spoken sentence comprehension (i.e., far transfer effects) in IWA. Based on a study by Novick et al. (2014), we designed an *n*-back task with ‘lures’ for the training of IWA. The *n*-back task is a widely used, complex task involving multiple processes (e.g., encoding incoming stimuli, monitoring, maintaining, and updating WM representations, establishing and maintaining bindings between memory contents and their temporal context). With the inclusion of lures (distractor items) we were able to target interference skills (e.g., Kane et al., 2007; Novick et al, 2014), and potentially recruit brain areas known to be involved in spoken sentence processing (e.g., LIFG, see Friederici, 2002, Thompson-Schill et al., 2005). As previous research has shown, the amount of transfer following training depends on the extent of the overlap between cognitive and neural resources shared by the trained and the untrained tasks (Dahlin et al., 2008; Novick et al., 2014). Therefore, it was expected that training with an *n*-back task that includes lures would result in transfer to spoken sentence comprehension in IWA.

Several other factors influenced our task selection as well. First, simple *n*-back tasks (without lures) have already been shown to improve working memory performance in healthy young adults (Jaeggi et al., 2008), in pediatric populations (Jaeggi et al., 2011), in elderly individuals (Li et al., 2008), and in populations with neuropsychological disorders (e.g., traumatic brain injury, Cicerone, 2002). Second, the *n*-back task can be easily designed to be adaptive, which seems to be a critical factor in producing transfer effects (Morrison & Chein, 2011). In adaptive training tasks, the level of difficulty is continuously adjusted according to the participant's performance. This allows participants to practise the task at a level that is sufficiently demanding (Shipstead et al., 2012) but not too difficult. Third, the existing data on the optimal training schedule in healthy populations (roughly 20 sessions, each lasting 30-60 minutes, 5 times a week, Shipstead et al., 2012) provided us with a guideline in designing the present study.

Briefly, in the present study, IWA were trained on an *n*-back task with lures and their performance was assessed before (pretest) and after (posttest) the training on outcome measures related to WM and EFs, and sentence comprehension. To examine possible test-retest effects, a control group of IWA was assessed on the same outcome measures without participating in the training.

Our choice of participants was primarily motivated by previous findings in the literature suggesting that harnessing cognitive resources through WM or EF trainings might bear a higher potential (both in terms of effectiveness and treatment value) in the presence of marked sentence comprehension deficits. However, we needed to select participants who, despite their sentence comprehension deficits, were able to follow task instructions. Accordingly, for the training group, we recruited two individuals with transcortical motor aphasia and one individual with anomic aphasia. For the control group, we recruited individuals with similar types and severity level of aphasia.

Our aim in this study was to answer the following questions: (1) Can WM and EFs be enhanced through training in IWA? If yes, (2) does the training lead to near transfer effects on WM and EFs? (3) Does the training lead to far transfer effects on sentence comprehension as well? Apart from an improvement on the training task, we expected trained participants to show both near and far transfer effects.

5.2 Methods

5.2.1 Participants

Table 2 summarizes the characteristics – including lesion location – of all participants. Three participants with chronic aphasia took part in the training phase of the study. B.B. and B.L. were classified as having transcortical motor aphasia and K.K. as having anomic aphasia by using the Western Aphasia Battery (Kertész, 1982; Hungarian adaptation: Osmánné Sági, 1991). The control group consisted of five IWA who participated only in the pretest and the posttest, but not in the training sessions. One of them was classified as having Broca's aphasia, two as having transcortical motor, and two as having anomic aphasia (Kertész, 1982; Hungarian adaptation: Osmánné Sági, 1991)[8]. All participants had a single left hemisphere infarct, confirmed by CT or structural MRI (see Table 2). All of them spoke Hungarian as their native language and were right-handed. They have been recruited at two rehabilitation centres in Budapest, Hungary: the Flór Ferenc Hospital and a non-profit organization supporting aphasia rehabilitation ([in Hungarian] Újrabeszélők Egyesülete [Association for People Speaking Again]). The results of a neurological assessment indicated intact visual acuity for all participants. All IWA also reported hearing within normal limits. Participants in the training and control groups were matched in age, education, intelligence, and time post-onset.

---

[8] Participants peforming the training had a WM deficit measured with a digit span task (Racsmány, Lukács, Németh, & Pléh, 2005): K.K. and B.B. had a span of three, and B.L. had a span of four.

Table 2 *Background description of the participants*

									WAB Subtests' Scores				
									SS		Comp	Rep	Nam
Case	Age	Gender	Education	Aetiology	Lesion	TPO (months)	Aphasia Type (WAB)	WAB AQ	IC	Flu			
K.K.	57	M	11	CVA	Left fronto-temporal haemorrhage	12 m	Anomic	71.3	7	5	8.65	8.4	6.6
B.B.	63	F	11	CVA	Infarct of the left MCA	12 m	TMA	62.5	5	4	7.25	9.6	5.4
B.L.	64	M	10	CVA	Infarct of the left MCA	8 m	TMA	57	4	3	7.5	10	4
Control													
B.T.	67	M	17	CVA	Left fronto-temporo-parietal infarct	180 m	Broca	62.7	5	4	8.45	7.9	6
T.GY.	73	M	11	CVA	Left fronto-temporo-occipital infarct	20 m	TMA	72.5	7	4	8.7	9	7.5
M.S.	53	F	10	CVA	Left fronto-parieto-occipital infarct	8 m	TMA	68.1	7	4	8.3	8.28	6.45
CS.R.	35	F	13	TBI	Left fronto-temporo-parietal contusion	240 m	Anomic	96.3	10	10	9.75	9.4	9
H.I.	45	F	11	CVA	Infarct of the left MCA	96 m	Anomic	79.5	9	7	9.15	7.4	7.2

Note. CVA = cerebrovascular accident; TBI = traumatic brain injury; MCA = middle cerebral artery; TMA = transcortical motor aphasia; WAB = Western Aphasia Battery; SS = spontaneous speech; IC = information content; Flu = fluency; Comp = comprehension; Rep = repetition; Nam = naming.

5.2.2 General design and procedures

We used a combined *pre/post-test case control design*. In the first three sessions, each lasting approximately 90 minutes, all participants underwent an initial assessment of their language skills and intelligence using the Western Aphasia Battery (Kertész, 1982; Hungarian adaptation: Osmánné Sági, 1991) and the Raven Progressive Matrices (Raven, Raven, & Court, 1962) (Session 1) and they completed two *n*-back tasks and two language tasks as outcome measures (Session 2 and 3 – Pretest sessions). Participants were offered breaks between tasks. Following the pretest, the three IWA in the training group participated in 13 training sessions over a period of four weeks (Session 4-16). There were three-to-four training sessions per week, lasting approximately 20 minutes each. One day after the completion of the training, participants completed two posttest sessions identical to the two pretest sessions (Session 17-18).

Control participants underwent the same initial assessments and pre- and posttest sessions as participants in the training group. The timeframe between pretest and posttest sessions was the same for the two groups (i.e., 4-5 weeks). Unlike the training group, control participants received no training during this interval (see Jaeggi et al., 2008; Novick et al., 2014). Three experimenters (two trained speech and language therapists and one trained nurse) conducted the study. They all received the same instructions regarding the training procedures. Sessions were distributed among experimenters according to their availability based on their working schedule. Each session was conducted by one experimenter. All computerized tasks were run by Presentation® software (Version 14.1) on an IBM T40p ThinkPad®.

5.2.2.1 The training task

We created an *n*-back task with 'lures' to target both WM and interference control. Based on the classic *n*-back paradigm that focuses on updating WM (Logan, 1994), participants were exposed to a stream of letters and were asked to press a button when a letter was the same as the one appearing *n* trials prior to this presentation (see Figure 3A). In addition, lures were incorporated into the task; letters that were the same as the one presented *n-1* or *n+1* (but not *n*) trial before (see Figure 3B and 3C). These trials, evoking interference between the representation of the target and that of a lure, required interference control (Kane et al., 2007; Novick et al., 2014).

Letters were presented sequentially on a computer screen at a rate of three seconds (stimulus length: 1000 ms; interstimulus interval: 2000 ms) per trial. In each trial, the stimulus was sampled from a pool of eight letters: B, F, K, H, L, S, C, and N. Participants responded manually by pressing the SPACE bar of the computer's keyboard. No responses were required for non-target items. One training session comprised eight blocks consisting of 16+5*(*n*-1) trials including 5 targets, resulting in a daily training time of ~20 minutes.

The level of difficulty of an upcoming block was set adaptively based on performance on the previous block. If a given threshold – described below – was reached at the end of a block, then difficulty level for the upcoming block increased by one, if the threshold was not reached for four consecutive blocks, the difficulty level decreased by one. Increase in difficulty level meant advancing through three lure levels at each value of *n* (i.e., no lures, *n+1* lures only, both *n+1* and *n-1* lures). Once participants completed three lure levels at a given *n*, there was an increase in *n* by one. For example, difficulty level 1 was a one-back block without lures, difficulty level 2 was a one-back block with *n+1* lures, difficulty level 3 was a two-back block without lures, difficulty level 4 was a two-back block with *n+1* lures, difficulty level 5 was a two-back block with *n+1* and *n-1* lures, difficulty level 6 was a three-back block without lures, and so on.

The threshold to reach was defined based on three measures: hit rates (proportion of responses to targets), false alarm rates for non-targets (proportion of responses to non-targets), and false alarm rates for lures (proportion of responses to lures), when lures were present in the block. The threshold was defined as having a hit rate above or equal to 80%, a false alarm rate for non-targets below 30%, and a false alarm rate for lures (when lures were present in the block) below 20% (or only one false alarm for lures). The number of lures in blocks including lures was always a random integer between three and five. The order of target and lure trials was pseudorandomised in a way that they did not overlap with each other, and that a lure or a target trial never preceded or followed immediately a target item.

To maximize motivation and compliance with the training instructions, participants received two types of feedback after each block. Following the work of Jaeggi et al. (2011), one type of feedback was provided based on participants' average hit rate, false alarm rate for non-targets, and false alarm rate for lures. A second type of feedback was provided based on the pattern of participants' errors. When false alarm rate for non-targets was higher than 50%, they were given the feedback "Caution: you might be pressing the button too often." When false alarm rate for non-targets was below 50%, but false alarm rate for lures was above 60%, the feedback was "Caution: there are some tricky trials that might lure you into pressing the button." If hit rates were below 40%, the feedback was "Caution: you're pressing the button quite rarely. Don't be afraid to press it more often." The feedback was always presented both visually and verbally to the participants.

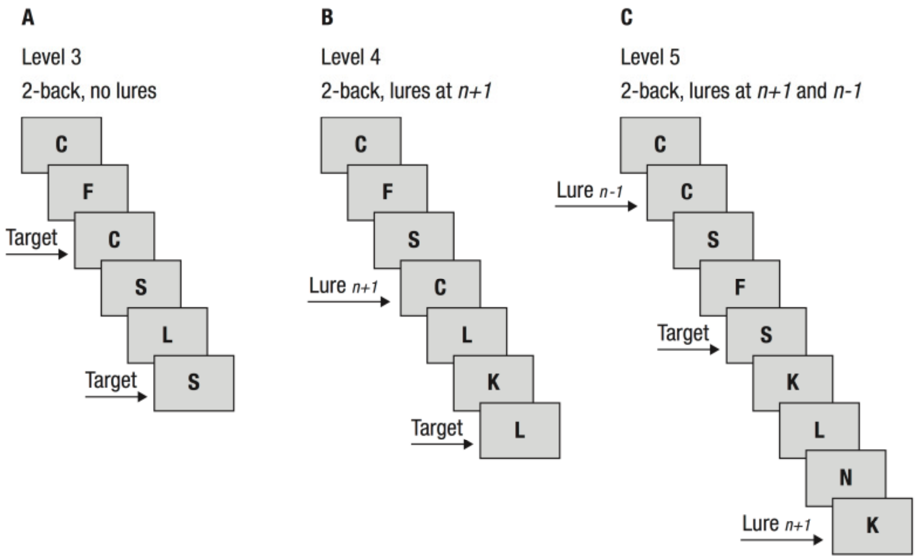

Figure 3 The *n*-back task with "lures" used as the training task, illustrated here with three levels of difficulty comprising three lure levels within the 2-back level
Participants had to perform three lure levels before n increased. (A) Difficulty level 3: 2-back with no lures. (B) Difficulty level 4: 2-back with lures at *n+1* position. (C) Difficulty level 5: 2-back with lures at *n+1* and *n-1* position. Note that at the 1-back level there could be no lures at the *n-1* position, hence there are only two difficulty levels before level 3: 1-back with no lures, and 1-back with lures at the *n+1* position.

As mentioned above, the 13 training sessions spread across four weeks lasted each 20 minutes. The choice of twenty-minute-long sessions was motivated by earlier results (Jaeggi et al., 2008; 2011) showing that sessions this long were successful in producing training and transfer effects in healthy adults and children, without exhausting participants.

5.2.2.2 Outcome measures used at pretest and posttest

To assess near transfer effects on WM systematically, we used two computerized experimental tasks that differed from the training task in varying degrees (for similar approach, see Waris et al., 2015). One task was a visual *n*-back task with letters. The structure and the stimuli of this task were identical to the training task but they differed in timing. The second task was an auditory *n*-back task with computer-generated tones. The structure of this task was the same as that of the training task but the stimuli were different. Neither of these tasks included lures. To assess far transfer effects on spoken sentence comprehension, we used the Hungarian version of the TROG (TROG-H; Lukács et al., 2012). To control for non-specific effects of training (Nickels, 2002; Vallat et al., 2005), we used a naming task as a non-target measure. This task was less demanding on WM than sentence comprehension and participants were not trained on it. Hence, we expected no improvement or at least less improvement on this task following the training sessions compared to outcome measures assessing near and far transfer effects. In the following section, we present details about the outcome measures.

Near transfer 1: Visual n-back task with letters. This task was very similar to the training task, allowing for detecting the smallest possible transfer effects. A minor change was the duration of the interstimulus interval (1000 ms compared to 2000 ms in the training task). Participants were exposed to a stream of letters. One letter appeared on each trial and participants had to respond by pressing the SPACE bar on the keyboard when the stimulus presented was the same as the one presented n trials before. We varied n within subjects, and all participants performed the *n*-back task first with $n = 1$, and then with $n = 2$. In both conditions, the task consisted of three blocks, with 48 trials (including 15 targets) in total in the 1-back condition and 63 trials (also including 15 targets) in total in the 2-back condition. For each trial, a letter was sampled from the same pool of letters as in the training task (i.e., B, F, K, H, L, S, C, and N). Sampling was pseudorandomised, so that 11-15 trials required a hit response in both conditions from each participant. In each trial, the letter was presented in the middle of the screen for 1000 ms, and trials were separated by a 1000 ms-long interstimulus interval.

Near transfer 2: Auditory n-back task with computer-generated tones. The task was used to assess near transfer effects across modalities (auditory versus visual modality). This task was originally designed for a previous study (Zakariás et al., 2013). Participants were exposed to a stream of tones. One tone was presented at each trial and participants had to respond when the stimulus presented was identical to the one appearing n trials before. We varied n within subjects, and all participants performed the *n*-back task first with $n = 1$, and then with $n = 2$. In both conditions, the task consisted of 5 blocks of 30 trials. Blocks were separated by self-paced resting periods. The first blocks in both conditions were used as practice blocks. The results show data from Blocks 2–5 in both conditions. For each trial, a sound was sampled from a pool of eight pure frequency sounds generated by Presentation® software (Version 14.1) (ca. half sounds starting from the standard musical note A5: 440 Hz, 490 Hz, 540 Hz, 590 Hz, 640 Hz, 690 Hz, 740 Hz, and 790 Hz). Sampling was fully randomised so that the chance of sampling a sound that was presented n trials before was one to four in each trial. Sounds were presented for 300 ms at a constant volume followed by a silent period of 1500 ms (1000 ms response window plus 500 ms intertrial interval). In the practice blocks, all trials were followed by a 1000 ms feedback trial if the participant pressed the response button. No feedback was provided in Blocks 2–5.

Far transfer: TROG-H. Spoken sentence comprehension was examined by administering the Hungarian version of the TROG (TROG-H, Lukács et al., 2012). The test comprises 80 multiple-choice items, organized into 20 blocks, each testing a different grammatical structure. For each item, an array of four coloured pictures is presented and the task is to select the picture matching the word, phrase or sentence read aloud by the experimenter. The grammatical complexity of the test increases as the test progresses. For each item, there are three – either lexical or grammatical – distractors. Performance on TROG-H was scored by giving one point for each correct response. Each participant completed the entire test, lasting between 25-35 minutes, depending on the severity of the comprehension deficit of the participants.

Non-target measure: Boston Naming Test. It is a confrontation naming task widely used in aphasia diagnostics. It consists of 60 black and white line-drawings (Kaplan, Goodglass, & Weintraub, 1983). Participants were presented with one picture at a time, and were asked to name the object they saw. To minimize test-retest effects, the experimenter was instructed to give neither phonemic, nor semantic cues to elicit naming following the completion of a trial. The experimenter recorded all responses; a positive score was given for every correct answer given within the first 20 seconds following the presentation of a picture, resulting in a maximum score of 60 points. Participants completed this task between 20-35 minutes, depending on the severity of their naming difficulties.

5.3 Results

5.3.1 Improvement on the training task

Individual and group level performances on the training task are shown in Figure 4. To analyse performance at the individual level (Fig 3.A), correlations between number of training sessions and mean difficulty level at a session were calculated using Pearson correlation coefficient. According to this, K.K. showed a significant increase in performance ($r = .701$, $p < .01$), B.L. showed a tendency for increase ($r = .501$, $p = .08$), and B.B. did not show statistically significant improvement ($r = .220$, ns.). Despite these considerable differences in individual performance, Friedman's ANOVA indicated a significant change in training task performance at the group level ($\chi^2(12) = 21.25$, $p < .05$, Fig 3.B).

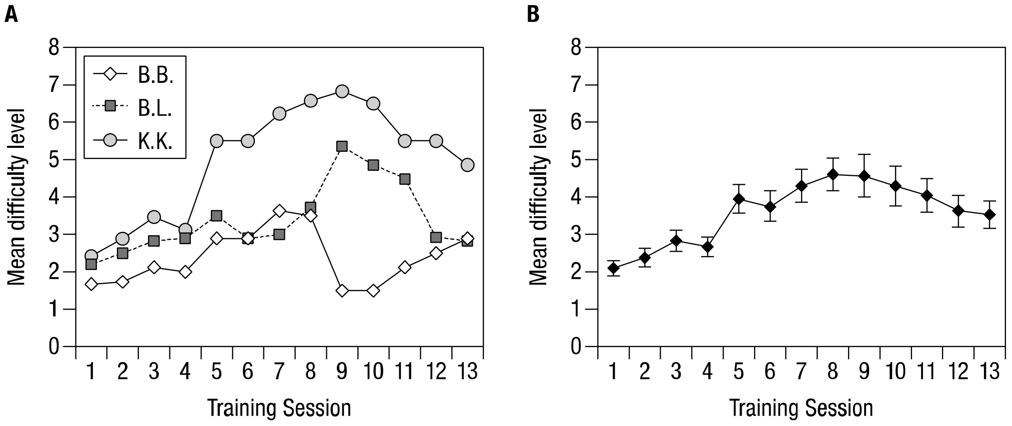

Figure 4 Performance on the training task during the thirteen sessions of training
(A) Performance shown for each participant. K.K. improved significantly ($p < .05$) across sessions, B.L. showed a level of tendency of improvement ($p < .1$), whereas B.B. did not show significant improvement. Error bars show standard errors. (B) Performance shown at group level. There was a significant change in performance across training sessions ($p < .05$).

5.3.2 Improvement on the outcome measures

We used the Revised Standardized Difference Test (RSDT; Crawford & Garthwaite, 2005; Crawford, Garthwaite, & Porter, 2010) to compare each trained participant's data separately to the data of all controls. Because RSDT allows for comparing a single case even to a modest-size control group, or non-normally distributed group data, it is extremely useful in studies dealing with neuropsychological populations. For all outcome measures, K.K.'s, B.B.'s, and B.L.'s individual pretest-posttest differences were compared to the mean pretest-posttest difference of the control group. The results are presented in Table 3.

5.3.2.1 Near transfer

Scores for the visual and auditory *n*-back tasks were calculated as hit rates minus false alarm rates. These scores showed that K.K. improved significantly in the visual 1-back ($t = 2.219$, $p = .045$) and showed a tendency for an increase in the visual 2-back as well ($t = 1.778$, $p = .075$). B.B. improved significantly in the visual 1-back ($t = 2.539$, $p = .032$), and B.L. did not show significant improvement in either *n*-back task.

Table 3 ***Comparison of each case to the average of control group for all outcome measures***

		Case's score		Control group's score (*n*=5)			Significance test		Estimated effect size (z_{dcc})
Case	Task (transfer)	Pre	Post	Pre Mean (*SD*)	Post Mean (*SD*)	Correlation between pre- and post	*t*	*p*	Point (95% CI)
K.K.	Visual 1-back (near)	.58	1	.85 (.15)	.81 (.26)	.71	2.22	**.045***	-3.26 (-6.04 to -1.21)
	Visual 2-back (near)	.17	.51	.47 (.18)	.47 (.26)	.75	1.78	**.075'**	-2.58 (-4.91 to -0.80)
	Auditory 1-back (near)	.89	.83	.51 (.32)	.52 (.35)	.77	0.32	.382	0.45 (-0.82 to 1.81)
	Auditory 2-back (near)	.31	.46	.23 (.25)	.42 (.21)	.28	0.07	.474	0.09 (-0.81 to 0.99)
	TROG-H (far)	.76	.85	.76 (.11)	.75 (.09)	.99	4.07	**.008****	-7.34 (-13.23 to -3.21)
	BNT (non-target)	.57	.62	.64 (.22)	.63 (.22)	.98	0.97	.193	-1.41 (-2.84 to -0.24)
B.B.	Visual 1-back (near)	.31	.63	.85 (.15)	.81 (.26)	.71	2.54	**.032***	-3.81 (-7.58 to -1.01)
	Visual 2-back (near)	-.04	-.15	.47 (.18)	.47 (.26)	.75	0.40	.357	-0.55 (-3.17 to 1.92)
	Auditory 1-back (near)	-.04	.23	.51 (.32)	.52 (.35)	.77	0.90	.209	-1.27 (-3.09 to 0.26)
	Auditory 2-back (near)	.18	.16	.23 (.25)	.42 (.21)	.28	0.64	.280	0.83 (-0.26 to 2.10)
	TROG-H (far)	.60	.64	.76 (.11)	.75 (.09)	.99	1.44	.111	-2.13 (-4.45 to -0.20)
	BNT (non-target)	.45	.55	.64 (.22)	.63 (.22)	.98	1.67	**.083'**	-2.52 (-4.78 to -0.80)
B.L.	Visual 1-back (near)	.94	.75	.85 (.15)	.81 (.26)	.71	0.82	.229	1.14 (0.05 to 2.41)
	Visual 2-back (near)	.30	.10	.47 (.18)	.47 (.26)	.75	0.58	.297	0.81 (-0.58 to 2.38)
	Auditory 1-back (near)	.70	.74	.51 (.32)	.52 (.35)	.77	0.04	.484	-0.06 (-1.09 to 0.97)
	Auditory 2-back (near)	.26	.32	.23 (.25)	.42 (.21)	.28	0.37	.365	0.48 (-0.44 to 1.46)
	TROG-H (far)	.56	.75	.76 (.11)	.75 (.09)	.99	6.04	**.002****	-13.12 (-23.58 to -5.86)
	BNT (non-target)	.23	.20	.64 (.22)	.63 (.22)	.98	0.27	0.40	0.38 (-1.62 to 2.44)

Note. Each case's pre-post difference is compared to the mean pre-post difference in the control group. *N*-back scores are corrected scores: hit rates minus false alarm rates. Scores of TROG-H and BNT are the proportion of correct responses. TROG-H = Hungarian version of the Test for the Reception of Grammar; BNT = Boston Naming Test; CI = confidence interval. Significance tests and effect sizes are calculated with the Revised Standardized Difference Test (RSDT, Crawford & Garthwaite, 2005).

*Significant improvement in the task ($p < .05$, one-tailed).

**Significant improvement in the task ($p < .01$, one-tailed).

'Tendency for an increase in the task ($p < .1$, one-tailed).

5.3.2.2 Far transfer

Two scores were calculated for the TROG-H: First, proportion of correct responses (Lambon Ralph, Graham, Ellis, & Hodges, 1998; Stothard & Hulme, 1992) showed that K.K. and B.L. improved significantly in sentence comprehension (t = 4.070, p = .007 and, t = 6.035, p = .001, respectively), whereas B.B. did not (t = 1.44, ns.). Second, the number of blocks in which the participant correctly responded to all items (i.e., raw scores, Bishop & Adams, 1990; Bishop & Hsu, 2015) showed that K.K. improved at the level of tendency (t = 1.632, p = .089), whereas B.L. and B.B. improved significantly (t = 2,465, p = .035 and t = 2,953, p = .021, respectively).

5.3.2.3 Non-target measure

Results on BNT showed that K.K.'s and B.L.'s performance remained stable on the task (t = 0.97, ns. and t = 1.27, ns., respectively), but B.B. showed a tendency for an increase in naming (t = 1.67, p = 0.083).

5.4 Discussion

The aims of the current study were to investigate whether WM and EFs can be enhanced in aphasia, and to see whether this enhancement transfers to unpractised WM tasks and to spoken sentence comprehension. To our best knowledge, this is the first study using an adaptive *n*-back task for training in IWA. This task targeted several processes of WM and EFs, such as maintaining and updating WM representations, as well as interference control.

Before we further interpret our data, it is important to note the differences between our group-level results and the individual outcomes. For instance, although the group-level statistics showed a significant training effect on the adaptive *n*-back task with 'lures', at the individual level only one participant (K.K.) showed statistically significant improvement. This participant showed improved performance on some of our near and far transfer measures, namely on WM tasks that were similar to the training task but were not practised during training, as well as on the sentence comprehension task (supposedly requiring

WM and EFs). However, the other two participants showed a less consistent pattern. B.L. showed a tendency for improvement on the training task and significant improvement on the TROG-H without a significant change on any of the near transfer measures, whereas B.B. did not improve on the training task, but showed increased performance in one of the measures computed from the TROG-H and one of the near transfer tasks. Note that the pattern of results on the near transfer tasks may be due to the low sensitivity of our measures (see the section 'Limitations of the study').

Our most important finding is that two of the trained participants (K.K. and B.L.) improved significantly in spoken sentence comprehension measured by the TROG-H without a change in performance in the non-target language task (Boston Naming Test). This finding is in line with outcomes of studies that reported enhanced sentence comprehension following WM training (Harris et al., 2014; Salis, 2012). From a theoretical point of view, these results provide converging evidence for a specific relationship among sentence comprehension, WM, and EFs, as suggested by previous research (Caplan & Waters, 2013; Haarmann et al., 1997; Martin & He, 2004; Novick et al., 2005). Given the large individual differences though, future research is needed to clarify how far these results are generalizable to the population of IWA.

Because our study lacks information on exact lesion locations (i.e., we did not have structural MRI data for all participants, and the resolution of the available MRI images was not sufficient for this purpose), we can only speculate on the neural mechanisms of the language improvements found after training. According to a recent review on functional MRI studies on the changes in brain activations following language treatment in aphasia (Geranmayeh et al., 2014), intact domain-general networks linked to the LIFG and serving top-down interference control may influence language recovery (in stroke-related aphasia). Thus, an individual's ability to activate domain-general networks in response to impaired language task performance can play an important role in recovery over time (Geranmayeh et al., 2014). Future studies could assess interference control abilities together with structural MRI to be able to make conclusions about the exact relationship between interference control, structural integrity, and training outcomes. Nevertheless, our results are consistent with findings suggesting that improvement in domain-general interference control can be beneficial also in the improvement of language-specific functions, such as spoken sentence comprehension.

5.4.1 Limitations of the study

First, we did not use a multiple baseline design that would have allowed for taking into account variability within individuals in the outcome measures (Nickels, 2002). Collecting data on baseline measures across multiple sessions seems to be especially important in studies using experimental tasks with individuals with neuropsychological impairments. In a recent study, we suggested that experimental tasks can be more appropriate to tap process-specific deficits in WM and EFs compared to standardized tests (Zakariás et al., 2013) and, therefore, could be more sensitive to changes occurring due to interventions. However, the mixed pattern of results on some of our near transfer tasks also suggest that there is a need for more practice on experimental outcome measures prior to training. Alternatively, a multiple baseline design may be necessary when testing neuropsychological populations, in order to decrease variability within individuals.

A further concern is related to the use of the BNT as a non-target measure. Although naming requires less WM than sentence comprehension, a growing literature suggests that it requires interference control (e.g., Biegler et al., 2008; Geranmayeh, Brownsett, et al., 2014; Robinson et al., 1998; Schnur et al., 2006), and that deficits in interference control may account for certain types of naming errors in aphasia (e.g., perseveration). Although none of the participants improved significantly on the BNT, the relationship between interference control and naming may have contributed to the tendency-level performance increase on the BNT for participant B.B.

5.4.2 Suggestions for future research

Using the TROG-H as an outcome measure enabled us to detect general improvement in sentence comprehension. However, the inclusion of various syntactic structures makes it difficult to point to the exact underlying mechanisms of transfer on sentence comprehension. Future studies can address the specificity of such transfer effects by developing experimental outcome measures that test improvement on specific syntactic structures, such as relative clauses or garden-path sentences that are assumed to involve WM or EFs (Novick et al., 2009, 2013, Thompson-Schill et al., 2005).

Future research is also needed to explore how dose of training (i.e., session length, frequency, training duration, and the total number of training sessions) could be optimized for training effects. Although our results show that the training schedule (about four-week-long, with a total of 13 sessions, lasting approximately 20 minutes a day) used in this study was appropriate to produce training effects at the group level, modifications may be needed to show more robust effects at the individual level.

To achieve more robust effects, future studies may need to systematically investigate other factors known to affect performance during training. Potential targets for such investigations are motivational factors, as they have been suggested to be closely related to training efficiency and to account for individual differences in the responsiveness to training (Jaeggi et al., 2011; Katz et al., 2014). In fact, motivational factors might also explain the decrease in training performance, observed in our data, starting between the seventh and the ninth session for all participants. Only studies with higher statistical power could determine whether this inverted U-shape pattern is a significant and reliable one. Developing a more systematic feedback scheme could also increase the training and transfer effects as feedback has been shown to play a key role in different fields of learning (e.g., for a review on motor learning, see Maas et al., 2008) as well as maintaining motivation during training (Jaeggi et al., 2011). A more comprehensive feedback scheme could, for instance, include a trial-by-trial feedback on hits, false alarms, and misses.

In sum, mapping out the role of dose, motivation, and feedback in producing training effects could lead to a better understanding of learning and transfer mechanisms in neuropsychological populations, as well as to the development of more efficient intervention methods.

In conclusion, our results suggest that working memory and executive functions could be enhanced through computerized training in a sample of individuals with chronic aphasia and this enhancement may have led to improvement in spoken sentence comprehension.

Acknowledgements

We thank K.K., B.L., and B.B. for participating in the study, Éva Jánosi Mentuszné, Tünde Buják and Zsuzsanna Barta for their help in running the study. During part of this study, Lilla Zakariás was supported by the Collaboration of Aphasia Trialists (COST IS1208).

6 Study 2: Transfer effects on spoken sentence comprehension and functional communication after working memory training in stroke aphasia[9]

Abstract

Recent treatment protocols have been successful in improving working memory (WM) in individuals with aphasia. However, the evidence to date is small and the extent to which improvements in trained tasks of WM transfer to untrained memory tasks, spoken sentence comprehension, and functional communication is yet poorly understood. To address these issues, we conducted a multiple baseline study with three German-speaking individuals with chronic post-stroke aphasia. Participants practised two computerised WM tasks (*n*-back with pictures and *n*-back with spoken words) four times a week for a month, targeting two WM processes: updating WM representations and resolving interference. All participants showed improvement on at least one measure of spoken sentence comprehension and everyday memory activities. Two of them showed improvement also on measures of WM and functional communication. Our results suggest that WM can be improved through computerized training in chronic aphasia and this can transfer to spoken sentence comprehension and functional communication in some individuals.

---

[9] This chapter has been published as:
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

6.1 Introduction

Individuals with aphasia (IWA) may present with concomitant cognitive deficits including deficits of short-term memory, working memory (WM)[10] (e.g., Friedmann & Gvion, 2003; Mayer, Mitchinson, & Murray, 2017; Nickels et al., 1997; Sung et al., 2009) and executive functions (e.g., Helm-Estabrooks & Albert, 1991; Nicholas, Hunsaker, & Guarino, 2017; Purdy, 2002; Zakariás et al., 2013). WM is a complex cognitive construct referring to processes that support the temporary maintenance *and* manipulation of information (Baddeley, 2012; Engle, 2002; Martin et al., 2012). Manipulation in WM involves various processes, such as shifting attentional control between tasks or mental sets, updating and monitoring WM representations, inhibiting prepotent responses, and resolving different types of interference (Friedman & Miyake, 2004; Miyake et al., 2000). Such processes have been considered under the umbrella term executive functions (e.g., Miyake et al., 2000).

There is strong evidence suggesting that WM impairments can negatively influence various language processes in aphasia, such as lexical-semantic processing (Martin et al., 2012; Novick et al., 2009; Robinson et al., 1998), sentence comprehension (Novick et al., 2009; Sung et al., 2009; Wright et al., 2007), spoken discourse and functional communication (Frankel et al., 2007; Fridriksson et al., 2006; Keil & Kaszniak, 2002; Luna, 2011; Penn et al., 2010; Ramsberger, 2005), and reading (Caspari et al., 1998). Spontaneous recovery (Sharp, Turkheimer, Bose, Scott, & Wise, 2010) and responsiveness to language treatment have also been shown to relate to WM skills in aphasia (Brownsett et al., 2014; Lambon Ralph et al., 2010).

With such strong links between WM and aphasia, researchers began to devise experimental treatments that heavily rely on WM, hypothesizing transfer of treatment effects to language functions. In these studies, treatments of WM included one or more WM tasks practised intensively, and treatment effects were

---

[10] Short-term memory and WM are related constructs. It is generally acknowledged that short-term memory is responsible for the temporary maintenance and retrieval of information (Caplan & Waters, 2013), whereas WM is generally viewed as the combination of multiple components working together and actively manipulating information in short-term memory (Cowan, 2008). There is a multitude of theoretical accounts describing the relationship between short-term memory and WM. In the present paper we adopt the view that short-term memory is a component of WM (Baddeley, 2012; Cowan, 2008).

measured on components of WM (i.e., near transfer) and language (i.e., far transfer), including spoken sentence comprehension (Eom & Sung, 2016; Francis et al., 2003; Harris et al., 2014; Salis, 2012; Salis et al., 2017; Zakariás, Keresztes, Marton, & Wartenburger, 2018), reading comprehension (Coelho, 2005; Mayer & Murray, 2002; Sinotte & Coelho, 2007), and spoken discourse (Paek & Murray, 2015; Peach et al., 2017). In the next section we discuss in detail treatment studies of WM and spoken sentence comprehension in people with non-progressive aphasia, which is the focus of the present paper.

6.1.1 Working memory treatments and sentence comprehension

Recent WM treatment studies that aimed to improve spoken sentence comprehension in aphasia reveal mixed findings, possibly due to substantial variations in participant characteristics, treatment tasks, intensity and duration of treatment, as well as variations in the domains and patterns of transfer detected. For example, Paek and Murray (2015) described a patient with mild anomic aphasia and semantic short-term memory deficit. The treatment included various tasks aiming to improve components of WM (i.e., updating, phonological loop) as well as semantic processing (see Table 4). The intervention was delivered remotely (teletherapy) consisting of 20 hourly sessions distributed over four weeks. Although the authors reported improvements in all training tasks, they observed near transfer effects only in one measure of short-term memory (identity span). With respect to far transfer, no substantial change was observed in spoken sentence comprehension. However, greater improvements were found in spoken discourse as measured by story-telling tasks. Additionally, improvements in short-term memory and spoken discourse were maintained at 6-week follow up.

Eom and Sung (2016) conducted a group study with six participants presenting with different types and severity of aphasia (see Table 4). They used a repetition-based treatment, incorporating sentences with varying length and syntactic complexity. The treatment combined repetition of sentences after auditory presentation, reconstruction of sentences by using word cards, and reading sentences aloud. Trained structures included active sentences with two- and three-argument verbs, passive sentences, conjoined sentences, and centre-embedded sentences with a subject-relative clause. Twelve sessions were administered over a month (three hourly sessions a week). With respect to the outcome, participants improved in the repetition of treated and untreated sentences, as well as in WM measured by digit and word span tasks. More importantly, they improved in the comprehension of treated syntactic structures (see Table 4).

Zakariás et al. (2018) used a computerised adaptive training approach (e.g., Jaeggi et al., 2008; Novick et al., 2014) utilising an *n*-back task with letters. *N*-back targets components of WM, such as updating (Cohen et al., 1997) and interference control (Kane et al., 2007; Novick et al., 2014). The adaptive training task involved adjusting the difficulty level according to the participants' performance, ensuring that they always practised at an optimal level of difficulty. Training was delivered three to four times a week for a month (a total of 13 20-min sessions) to three Hungarian-speaking IWA (see Table 4). The authors detected a mixed pattern of training and transfer effects. One participant improved in the training task as well as untrained WM tasks and spoken sentence comprehension. Another participant improved in the training task and spoken sentence comprehension but did not show improvement in other measures of WM. The third participant did not show improvement in the training task but did show increases in performance, both in sentence comprehension and untrained WM. Zakariás and colleagues (2018) argued that individual differences in motivation as well as in cognitive abilities, such as interference control at the beginning of training could have influenced treatment outcome and transfer effects.

To replicate previous positive findings based on one IWA (Salis, 2012), Salis and colleagues (2017) delivered a training involving a recognition memory task (matching listening span) in five participants (for more information, see Table 4). The authors hypothesised far transfer to spoken sentence comprehension and improvements on psychosocial functioning as well as other communication skills after training. Participants received 27-30 treatment sessions. With respect to short-term memory (near transfer), changes were found only in one outcome measure (i.e., digit matching listening span). None of the changes observed in spoken sentence comprehension was statistically significant (see Table 4). As for the psychological measures of communication, a statistically significant increase was observed only in case of one participant.

In summary, although previous results suggest that components of WM indeed can be flexibly improved with training, the extent of transfer to untrained abilities and its boundary conditions are not well understood. There have been variations in the domains (i.e., WM and/or language abilities) and patterns of transfer detected after training: some researchers reported substantial effects on WM (e.g., Eom & Sung, 2016, but for null effects, see Salis et al., 2017), spoken sentence comprehension (e.g., Eom & Sung, 2016; Salis, 2012; Zakariás et al., 2018, for null effects, see Paek & Murray, 2015), and spoken discourse (Paek & Murray, 2015), whereas others did not find any effects on untreated processes after training (Salis et al., 2017). Although the role of WM in syntactic comprehension has drawn much attention in the last decades (e.g., Caplan et al., 2013; Caplan & Waters, 2013 for a review; Fedorenko, 2014; Haarmann et al., 1997), only Eom and Sung (2016) has investigated the specificity of transfer effects on syntactic comprehension. The inconsistent pattern of transfer can be observed also across participants: for example, in Zakariás et al. (2018), some participants showed near but not far transfer effects, while others showed the opposite pattern. In addition, there is limited knowledge as to which participants – with respect to type and severity of aphasia or degree of impairment in certain linguistic and WM processes – can benefit from training. Although some researchers suggested that training WM might bear a higher potential for IWA with moderate or severe sentence comprehension deficits (e.g., Salis, 2012; Zakariás et al., 2018), Eom and Sung (2016) concluded that WM treatments might be more beneficial for people with relatively preserved comprehension abilities. Based on observations that IWA with WM spans of zero performed at chance on the sentence comprehension tasks, whereas participants with WM spans of 1 or 2 showed normal performance on the tasks, Caplan et al. (2013) suggested that there is a minimal WM capacity (span above 1) that is needed to perform normally in sentence comprehension. These findings also suggest that WM treatments might bear a higher potential for IWA demonstrating with severe WM impairments.

In summary, potential training and transfer effects following WM training in aphasia warrant further systematic study to refine our understanding of the nature and the underlying mechanisms of transfer of WM training to different levels of linguistic processing.

6.1.2 Extending the ecological validity of working memory trainings in aphasia: Motivation, functional communication, and everyday memory

Besides resolving the issues discussed above, the present study aimed to extend the investigation to motivation and two relevant domains of target in aphasia. Research from other literature domains, beyond aphasia, suggests that motivation plays a substantial role in the effectiveness of WM training (Jaeggi et al., 2011; Jaeggi et al., 2014; Katz et al., 2014). Studies using *n*-back tasks for training in healthy children (Jaeggi et al., 2011) and healthy young adults (Jaeggi et al., 2014) suggest that motivational factors, such as interest in or engagement with the training activity mediates improvement in the training task, and, in turn, transfer to other untrained abilities (Lindeløv et al., 2016). Yet, motivation is an overlooked aspect of training, and to our knowledge no study has yet incorporated measures of motivation in WM treatment studies in aphasia.

For most IWA, the important goal of linguistic rehabilitation is improvement in functional communication, that is, the individual's ability to understand and convey information in everyday life situations (Blomert, Kean, Koster, & Schokker, 1994; Lind, Kristoffersen, Moen, & Simonsen, 2009). Therefore, such improvements are seen as the gold standard for demonstrating the effectiveness of any intervention. Despite its importance in aphasia rehabilitation and the suggested link between WM and functional communication (Frankel et al., 2007; Fridriksson et al., 2006; Keil & Kaszniak, 2002; Luna, 2011; Penn et al., 2010; Ramsberger, 2005), only very few studies have investigated transfer effects after WM training on functional communication (Murray et al., 2006; Salis et al., 2017).

Table 4 *Summary of WM treatments including outcome measures of spoken sentence comprehension, spoken discourse, and verbal communication in individuals with aphasia*

Studies	Participant(s)	Treatment procedures	Outcomes on language
Francis et al. (2003)	N = 1 (mild chronic aphasia)	Sentence repetition	⇑ TROG, TT; – active reversible sentences
Salis (2012)	N = 1 (severe TMA)	Matching listening span with nouns	⇑ TROG; – TT
Harris et al. (2013)	N = 2 (Broca's aphasia [DS], mild aphasia [AK])	Repetition and recognition tasks with words and non-words	⇑ for DS in semantically anomalous sentence judgements and sentence-picture matching (PALPA 55)
Paek and Murray (2015)	N = 1 (mild anomic aphasia)	N-back with pictures/written words, updating with pictures/written words, reading span involving grammaticality judgments and category naming, naming with spaced retrieval, opposite sentence training, reconstitution of words from oral spelling	– RTT; ⇑ %CIUs, CIUs/min in story-telling
Zakariás et al. (2018)	N = 3 (moderate chronic anomic [KK] and TMA [BL, BB])	Adaptive n-back with letters	⇑ for KK and BL in the TROG-H
Eom and Sung (2016)	N = 6 (Broca's, anomic, and Wernicke aphasia)	Repetition-based treatment protocol (active sentences with two- and three-argument verbs, passive sentences, conjoined sentences, and centre-embedded sentences with a subject-relative clause)	⇑ for five participants in sentence picture matching (Sung, 2015) including active sentences with two-argument verbs, active sentences with three-argument verbs, and passive counterparts of active sentences with two-argument verbs
Salis et al. (2017)	N = 5 (moderate chronic aphasia)	Matching listening span with nouns	– TROG, TT; ⇑ in the CETI for one participant

Note. ⇑ = improvement in the task; – = no change in the task; TROG = Test for the Reception of Grammar; TT = Token test; TMA = transcortical motor aphasia; %CIUs = percent of correct information units; CIUs/min = correct information units per minute; RTT = Revised Token test; TROG-H = Hungarian version of the Test for the Reception of Grammar; CETI = Communication Effectiveness Index; PALPLA = Psycholinguistic Assessments of Language Processing in Aphasia.

Although aspects of memory functioning in everyday life activities, such as difficulty in remembering appointments or recognizing familiar faces have been observed after stroke (e.g., Stewart, Sunderland, & Sluman, 1996; Wilson, Cockburn, Baddeley, & Hiorns, 1989), studies have provided limited or incomplete information about participants' aphasia. For instance, the presence and the number of IWA in some stroke studies are unclear (e.g., Barker-Collo, Feigin, Parag, Lawes, & Senior, 2010), or the diagnostic method to identify aphasia is not described (e.g., Duffin et al., 2012). Thus, knowledge about the extent of everyday memory problems, recovery of everyday memory, and its improvement in response to treatment in participants presenting with aphasia is limited (for the only study see Vallat-Azouvi et al., 2014).

6.1.3 The present study

In the present study, we used the *n*-back task for the training. *N*-back is a complex WM task involving multiple processes, such as encoding incoming stimuli, monitoring, maintaining, and updating WM representations, establishing and maintaining bindings between memory contents and their temporal context, as well as resolving interference between WM representations (Kane et al., 2007). In a typical *n*-back task, participants are presented with a continuous stream of items and are instructed to judge whether an item matches a previous one that was presented *n* items (e.g., n = 1, n = 2) before. Although the task commonly used to investigate WM in language-impaired populations, results regarding its reliability in aphasia are mixed with some studies showing excellent test-retest reliability (Mayer & Murray, 2012), whereas others showing only acceptable test-retest reliability (Zakariás et al., 2018). Varying test-retest reliabilities are likely due to differences in task stimulus materials, procedures, and the measures used to describe performance, as well as participants' cognitive and linguistic profile (cf., DeDe et al., 2014). Despite such challenges, certain properties of the task enhance research validity and treatment fidelity (i.e., the reliability of the administration of an intervention) in studies using *n*-back as a training task in aphasia. These properties include, among others, that the task does not require speech response, or that the task structure is easy to convey and the administration is simple and in most cases automatized.

The present study was motivated by the need to strengthen and extend the evidence base of WM treatments in aphasia and also improve our knowledge as to why inconsistent patterns of transfer were reported in previous studies. Our main objective was to systematically investigate patterns and potential domains of transfer after WM training. To this end, we chose a set of outcome measures that allowed for a systematic investigation of potential transfer effects, ranging from the training task (*n*-back) to very far transfer (functional communication). To assess the specificity of transfer effects and to better understand the underlying mechanisms of transfer on sentence comprehension, our outcome measures included specific syntactic structures that have been proposed to involve WM processes (e.g., non-canonical structures with varying complexity; Caplan et al., 2013; Haarmann et al., 1997). In addition, we aimed to extend earlier reports of WM training related transfer effects in aphasia by extending the ecological validity of our findings. Therefore, we included a set of far transfer tasks that covered a broad range of WM-relevant language and everyday functions, such as spoken sentence comprehension, functional communication, and everyday memory. To capture the effects of motivational factors on training performance across time, we monitored participants' motivation on a daily basis. In summary, the research questions in this study are:

(1) Does WM training transfer to cognitive domains targeted by the training but measured by untrained tasks (i.e., near transfer) in IWA?
(2) Does WM training transfer to spoken sentence comprehension, functional communication, and everyday memory (i.e., far transfer) in IWA?
(3) Are training and transfer effects maintained over time (i.e., at 4-6 weeks follow up)?
(4) Do motivational factors play a role in IWAs' WM training performance?

Our design followed an earlier report by Zakariás and colleagues (2018), that suggested that intensive *n*-back training can lead to improvements on untrained WM domains and spoken sentence comprehension (i.e., near and far transfer, respectively) in aphasia. We expected that IWA improving on the training tasks will improve on all outcome measures, but not on the non-targeted control measure (oral word reading). In addition, we hypothesized that stable and generally high interest levels (i.e., a factor of motivation) would be associated with greater improvement in the training task.

6.2 Methods

6.2.1 Participants

Three IWA participated in the study. Participants were included based on the following criteria: (1) aphasia as a result of left hemisphere stroke, (2) at least eight months post-onset, (3) German as the native language, (4) self-reported pre-stroke right-handedness, (4) moderate to severe impairment in sentence comprehension together with good single word comprehension (based on the Aachen Aphasia Test, AAT; Huber, 1983), (5) a score of three items or below in a verbal WM task (i.e., listening span, developed based on Tompkins, Bloise, Timko, & Baumgaertner, 1994)[11] and a score of five items or below in a computerised visuo-spatial WM task (Corsi block tapping). Exclusion criteria were: (1) bilateral lesions, (2) additional neurological or psychiatric disorder, and (3) participation in speech and language therapy during the time of study. Participants were recruited through the aphasia database of the University of Potsdam.

Participant 1 (R.D.) was a 39-year-old female six years post-onset. She worked as a beautician at the time of her stroke. Prior to the study, she had received individual speech and language therapy, which was suspended during the present study (altogether for four months). Participant 2 (V.O.) was a 77-year-old female 25 years post-onset. She had studied German literature and history, then had worked as a teacher, and later as a television editor. At the time of the study she was retired, was living with her husband and was not participating in any therapy. Participant 3 (Z.A.) was a 51-year-old female 15 years post-onset. Her right hand and arm were still non-functional at the time of the study. Before the stroke, she had worked as a trained nurse. She was not

---

[11] The procedure of the listening task followed that of Tompkins et al., 1994. Stimuli were modified to make the task suitable for use with participants with aphasia. For stimuli and procedure of the task see Appendix – Table A1.

participating in speech and language therapy but received physiotherapy once a week during the present study. The study was approved by the local research ethics committee of the University of Potsdam. The participants provided informed voluntary consent during the initial meetings. There was no dropout. Participants' biographical information and initial scores on various assessments are shown in Table 5.

Table 5 ***Background description of the participants***

	R.D.	V.O.	Z.A.
Gender	F	F	F
Age (years)	39	77	51
Education (years)	10	12	10
Etiology	CVA	CVA	CVA
Lesion	Infarct of the left MCA	Infarct of the left MCA	Infarct of the left MCA
Time post onset (years)	6	25	15
Aphasia type (AAT profile)	99.3% Broca's, 0.7% anomic	Unclassified	Unclassified
AAT (%)			
Token	60	56	30
Repetition	79.3	72.6	69.3
Written language	90	72.2	36.6
Naming	85.83	80.83	56.6
Comprehension	91.66	88.33	70.83
Spoken words	100	93.3	66.66
Spoken sentences	80	80	73.33
Written words	100	90	73.33
Written sentences	86.66	90	70
TROG-D (%)	77.38	76.19	53.57
Listening span – verbal WM (span)	2	2	1
Corsi block tapping – visuo-spatial WM (span)	5	4	5

Note. CVA = cerebrovascular accident; MCA = middle cerebral artery; AAT = German version of the Aachen Aphasia Test; TROG-D = German version of the Test for the Reception of Grammar; WM = working memory; note that AAT scores were obtained one and two years before the present study (for Z.A., and for R.D. and V.O., respectively). Other assessment data was obtained at the beginning of the study.

General design and procedures

For each participant, a multiple-baseline (with control) experimental design was adopted (see Figure 5 for an overview). Each participant received two blocks of WM training: (A) a visual *n*-back task with pictures; and (B) an auditory *n*-back task with spoken words. Following random assignment of participants to block order, R.D. and Z.A. started with the visual WM training, followed by the auditory WM training. V.O. received the training in the reverse order (auditory WM training, followed by visual WM training).

Participants were assessed before the first training block (i.e., pretest) and after the second training block (i.e., posttest) on several memory and language tasks. Assessments were distributed over six sessions in both test phases. The experimental tasks were administered twice in both test phases. In addition, four to six weeks after completion of the posttest, we conducted one follow-up test session –including a subset of the tasks administered at pre- and posttest – to tap into the time-course of training induced changes and maintenance of potential transfer effects. Experimental tasks were administered once at follow-up. The training blocks consisted of eight sessions each (approximately 25-35 minutes/session, three-four sessions/week), resulting in a four to five-week total training period. After each training session, participants completed a motivation questionnaire assessing their subjective experience related to the training. Altogether, the study consisted of 30 sessions for each participant, lasting approximately 10 weeks (see Figure 5). All computerised tasks were delivered by Presentation® software (Version 18.3) on a Lenovo X201 ThinkPad® (R.D.) or a Lenovo IdeaPad U310 (V.O. and Z.A.).

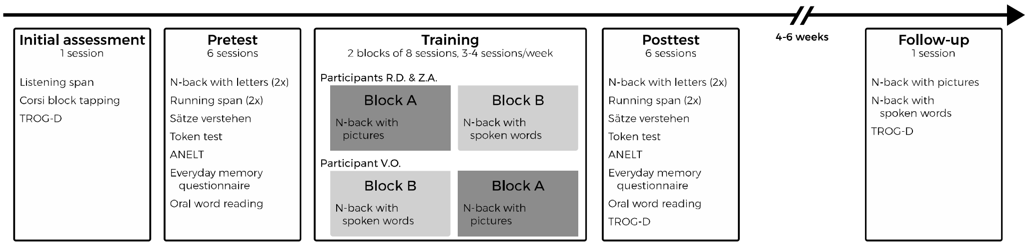

Figure 5 Design and tasks used in the study
Participants were randomly assigned to the order of the training blocks. Initial assessment was used to assess suitability in the present study. Pretest and posttest took 2.5 weeks each. Training blocks took 2-3 weeks each. The study lasted altogether ~10 weeks (30 sessions). Follow-up was conducted 4-6 weeks after completion of the posttest.

6.2.1.1 *Training tasks*

Based on Zakariás et al. (2018), we created two *n*-back tasks with identical design and procedure (one with pictures, one with spoken words). The two *n*-back tasks were chosen to tax verbal WM as well as domain general executive functions (e.g., interference control) (Kane et al., 2007; Redick &

Lindsey, 2013). Since the participants' word comprehension abilities were relatively good at the beginning of the training, we hypothesised that both semantic and phonological processes would be activated, at least to some extent, in both tasks.

Stimuli. Eight stimuli sets, each including eight stimuli belonging to different semantic categories (64 stimuli altogether), were created for the eight blocks in both training tasks (pictures, words). This allowed us to present eight stimuli belonging to different semantic categories in each block. For the *n*-back with pictures, eight pictures from eight semantic categories (animals, furniture, clothes, body parts/tools, vehicles/musical instruments, food, toys/household items, kitchen/home) were taken from the coloured version (Rossion & Pourtois, 2004) of the Snodgrass and Vanderwart (1980) set. When there were fewer than eight items available from the same category, we chose the remaining items from another category (e.g., vehicles and musical instruments, respectively). For the *n*-back with spoken words, eight words from the eight semantic categories (animals, vegetables/drinks, vehicles, furniture/house, musical instruments/toys, tools, clothes, professions) were recorded by a female native German speaker in an acoustically shielded recording studio, at a sampling rate of 44.1 kHz (16 bit, mono). The speaker was instructed to produce the words naturally with normal intonation and speech rate. Auditory recordings were created, edited, denoised, and segmented into single word sound files using Audacity®2.1.2. Words across the blocks were balanced for length (i.e., each block included three 1-syllable and five 2-syllable words) as well as for lexical frequency (i.e., no significant difference between the blocks). Frequency values were obtained from the CLEARPOND database (Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities; Marian, Bartolotti, Chabal, & Shook, 2012). Any two words in a block were checked not to result in a meaningful compound word if presented one after the other by a native German speaker.

Note that the limited number of pictures available in the Snodgrass and Vanderwart (1980) database as well as the limited number of words meeting the criteria in our auditory *n*-back did not allow us to choose items from the same eight categories in both tasks. Due to the category, frequency, and length constraints, 22% of the stimuli overlapped between the two training tasks. Stimuli of the tasks are included in the appendix (Table A2).

Procedure. Participants were exposed to a continuous stream of stimuli (i.e., either pictures or spoken words) and were asked to press a button on the keyboard when the stimulus presented was the same as the one that had been presented in *n* preceding trials (see Figure 6). In addition, “lures” were incorporated into the task; stimuli that were the same as the one presented *n-1* or *n+1* (but not *n*) trials before, requiring participants to resolve the conflict between the representation of the target and that of a highly familiar lure. The tasks were adaptive, that is, the task difficulty was always continuously adapted according to participants’ performance by means of automatic computer algorithms. If a given threshold (described below) was reached at the end of a block, then difficulty level for the upcoming block automatically increased by one, if the threshold was not reached for four consecutive blocks, the difficulty level decreased by one. Increase in difficulty level meant advancing through three lure levels at each value of *n* (i.e., no lures, *n+1* lures only, and both *n+1* and *n-1* lures), then advancing through to the next *n*.

The required threshold was defined based on three measures: (1) hit rates (proportion of responses to targets), (2) false alarm rates for non-targets (proportion of responses to non- targets), and (3) false alarm rates for lures (proportion of responses to lures), when lures were present in the block. The threshold was defined as having a hit rate above or equal to 80%, a false alarm rate for non-targets below 30% (R.D. and V.O.) or 10% (Z.A.)[12], and a false alarm rate for lures (when lures were present in the block) below 10%. In the *n*-back with pictures, stimuli were presented sequentially on a computer screen at a rate of 3 seconds (stimulus length: 1500 ms; interstimulus interval: 1500 ms) per trial. In the *n*-back with spoken words, stimuli were presented at the same rate

---

[12] Because Z.A.’s false alarm rate for non-targets was very high (above 20%) in blocks of the first training session (Training A), keeping the threshold for false alarms at 30% for the whole time of training would have let her advance to the next levels without actually mastering the task (based on trial-by-trial strategy). Therefore, after the first session we changed it from 30% to 10% for her.

(mean stimulus length: 785 ms, range: 445-1180 ms) via a loudspeaker (Speedlink Ellipz Stereo Speakers). Volume was adjusted to each participant's comfort with the volume control on the loudspeaker. Participants responded manually by pressing the SPACE bar on the computer keyboard. No responses were required for non-target items. One training session comprised six to eight blocks consisting of $18 + 5*(n - 1)$ trials including 5 targets, resulting in a daily training time of 25-35 minutes. The number of lures in blocks including lures was always five. The sequence of the stimuli in each block was randomized in both tasks.

Feedback. Recent studies have shown that feedback can impact participants' performance during training as well as keep them engaged with the training regimen (Jaeggi et al., 2011; Katz et al., 2014). To maximise motivation and compliance with the training, participants received three types of feedback during training. The first type of feedback was provided after each block. It was based on participants' hit rate, false alarm rate for non-targets, and false alarm rate for lures, by displaying their average performance in percentage on the screen. The second type of feedback was displayed based on the pattern of participants' errors. When the false alarm rate for non-targets was higher than 50%, they were given the feedback, "Caution: you might be pressing the button too often." When the false alarm rate for non-targets was below 50%, but false alarm rate for lures was above 60%, the feedback was, "Caution: there are some tricky trials that might lure you into pressing the button." If hit rates were below 40%, the feedback was, "Caution: you're pressing the button quite rarely." The third type of feedback was provided after certain trials: after each hit and at 80% of the misses, a message was displayed on the screen ("Good!" and "Missed out!", respectively). The first and second types of feedback were always also read aloud to the participants by the trainer.

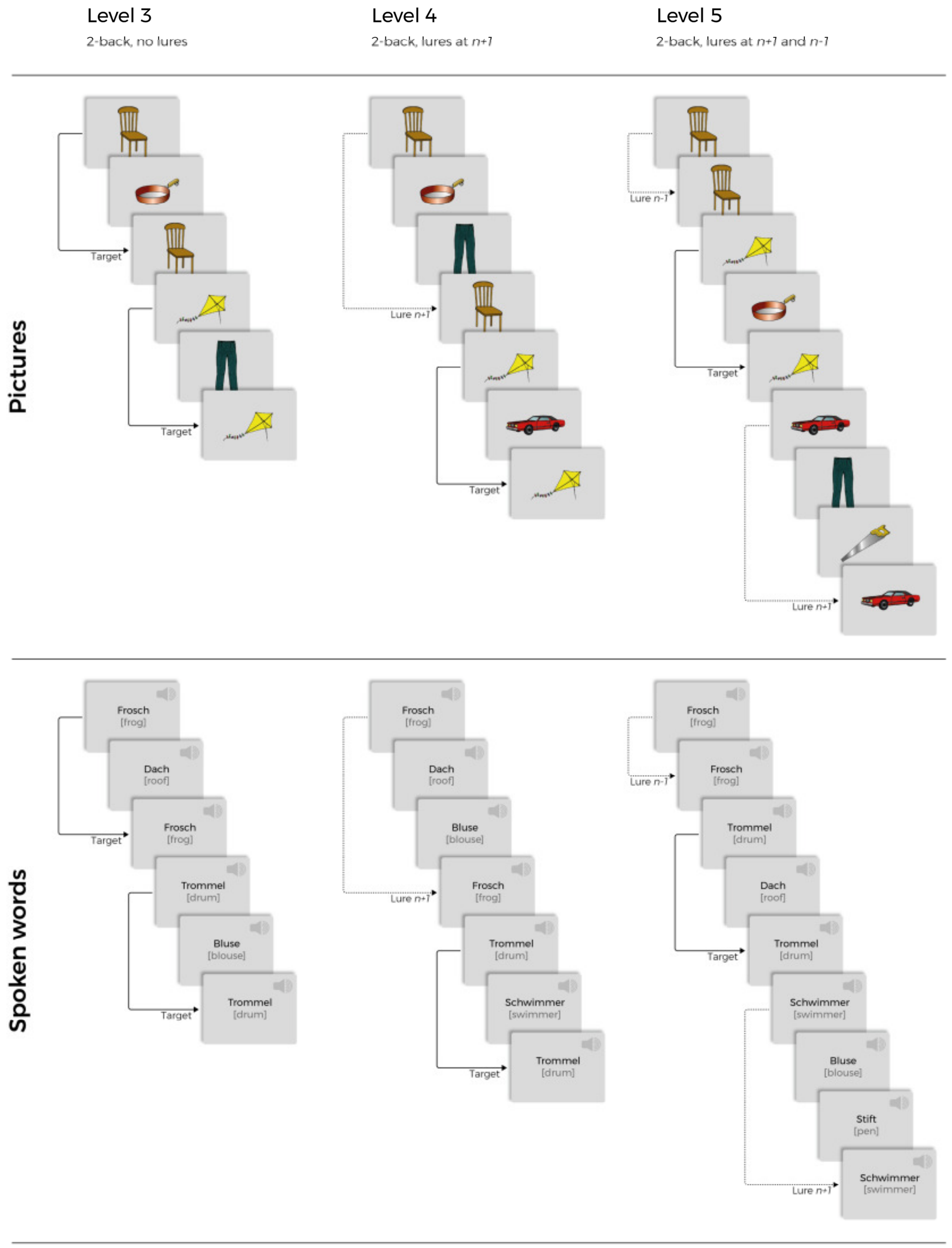

Figure 6 Two *n*-back tasks (pictures, spoken words) with “lures” used as training tasks, illustrated here with three levels of difficulty comprising three lure levels within the 2-back level

Participants had to perform three lure levels before *n* increased. Level 3: 2-back with no lures. Level 4: 2-back with lures at *n+1* position. Level 5: 2-back with lures at *n+1* and *n−1* position. Note that at the 1-back level there could be no lures at the *n−1* position, hence there are only two difficulty levels before level 3: 1-back with no lures, and 1 back with lures at the *n+1* position.

Delivery. The training was delivered in the participants' home in a quiet room. The training to R.D. was delivered by a trained speech-language pathologist (SLP) and a SLP student (Student 1). The training to V.O. and Z.A. was delivered by two SLP students (Student 2 and 3, respectively). All trainers had completed the same 3-hour training session regarding conducting and administering the training tasks (i.e., setting up the computer and the tasks, providing the computer-generated instructions and feedback to the participants, saving data on computer). The trainer was present at all training sessions.

6.2.1.2 Outcome measures

WM 1: N-back with letters (near transfer). This experimental task was used to assess near transfer effects across stimuli. Because the structure of this task was the same as those of the training tasks but the stimuli were different, it allowed us to assess task-specific effects. Participants were exposed to a stream of letters. One letter appeared on each trial and participants had to respond by pressing the SPACE bar on the keyboard when the stimulus presented was the same as the one presented *n* trials before. We varied *n* within subjects, and all participants performed the *n*-back task first with $n = 1$, then with $n = 2$, and finally with $n = 3$. In all conditions, the task consisted of three blocks, with 90 trials (including 15 targets) in total. In addition, a practice block with 15 trials (including 3 targets) was also included with $n = 1$. Data of the practice block was not included in the analysis. Blocks were separated by self-paced resting periods. For each trial, a letter was sampled from the same pool of letters (i.e., B, F, K, H, L, S, C, and N). Sampling was pseudorandomized to always provide exactly five targets in all blocks. In each trial, the letter was presented in the middle of the screen for 1500 ms, and trials were separated by a 1500 ms interstimulus interval (temporal parameters in the task were the same as in the training tasks). No feedback was provided to the participants.

WM 2: Running span (near transfer). This experimental task was used to assess transfer effects on updating (Collette et al., 2007; Pollack, Johnson, & Knaff, 1959). Running span involves similar processes as the *n*-back task, but has a different structure (Collette et al., 2007). Because it was not practised during the training, it also allowed us to separate task-specific from process-specific effects. The running span task was adapted to aphasia as follows: in each trial, participants were presented with a list of digits (one digit at a time), and were asked to respond by pointing the *n* last number of digits (*n*-span) when the list presentation ended. Importantly, participants were informed about *n* (i.e., how many digits they would need to report), but not the list length (i.e., they did not know when the list would end). Lists included two-six random digits (from the set 1-9) presented via computer. Digits appeared in the middle of the screen, one at a time, for 1500 ms. Immediately after each list, digits (separated by commas) together with one blank per to-be-recalled item appeared on the screen. For example, given the list, “6, 2, 4, 7, 5” in a 2-span condition, participants saw “6, 2, 4, _, _”. Participants had to report digits by pointing to the corresponding digits on a sheet of paper, which was positioned in front of them on a desk (i.e., no oral response was required). The experimenter noted down the answers on a scoring sheet and pressed ENTER to advance to the next trial. The task comprised three blocks of five trials (altogether 15 experimental trials), with span increased from 1 to 3 across blocks. Experimental trials were preceded with two probe trials with 1-span length. Probe trials were not included in the analysis. The number of correct trials was calculated in the task (max. 15).

Spoken sentence comprehension 1: TROG-D (far transfer). This standardized test measures the spoken comprehension of grammatical structures (Fox-Boyer, 2013). It comprises 84 multiple-choice items, organized into 21 blocks, each testing a different grammatical structure. The grammatical complexity and hence difficulty of the linguistic material increases with each block. For each item, an array of four coloured pictures is presented and the task is to select the picture matching the word, phrase or sentence read aloud by the experimenter. For each item, there are three – either lexical or grammatical – distractor pictures and one target picture. Each participant completed the entire test in approximately 30 minutes. We calculated and analysed the number of correct responses in the task.

Spoken sentence comprehension 2: Token test (far transfer). This standardized test measures comprehension of spoken commands of increasing length and, in the last subtest, of increasing grammatical complexity (Huber, 1983). Understanding of commands requires pointing to or manipulating with plastic tokens with different sizes, shapes, and colours. This version taken from the AAT (Huber, 1983) consists of five subtests, including 10 sentences in each. The number of correct responses was the dependent variable (max. 50).

Spoken sentence comprehension 3: Sätze verstehen (far transfer). This standardized test measures the comprehension of spoken sentences varying in syntactic complexity, semantic reversibility, and length (Burchert, Lorenz, Schröder, De Bleser, & Stadie, 2011). It consists of 204 sentences and uses a sentence-picture matching paradigm with two or four pictures (irreversible and reversible sentences with two-argument verbs, and relative clauses, respectively). It includes *short* and *long irreversible sentences* (22 sentences each), *case-marked canonical* (SVO) and *non-canonical* (OVS) *reversible sentences* (20 sentences each), *number-marked canonical* (SVO) and *non-canonical* (OVS) *reversible sentences* (20 sentences each), and *right-branching* and *centre-embedded subject* and *object relative clauses* (20 sentences each). Each participant completed the entire test over three sessions. With the inclusion of specific syntactic structures, the test assesses the specificity of transfer effects in terms of underlying mechanisms of transfer on sentence comprehension. The number of correct responses was calculated for each syntactic structure. In addition, aggregate scores in the canonical (i.e., SVOs plus SRCs) and the non-canonical (i.e., OVSs plus ORCs) conditions, as well as a total score (i.e., the number of all correct responses in the task) were calculated.

Functional communication: Amsterdam-Nijmegen Everyday Language Test, ANELT (far transfer). This test measures spoken communicative skills (Brunner & Steiner, 1994). There are two parallel versions (ANELT 1 and 2), each consisting of 10 items involving familiar everyday life situations (e.g., calling a doctor, talking to a cashier). Items are presented verbally to the participant. The experimenter is instructed to avoid conversing with the participant but to act as an interested listener, while the participant answers the items as a brief monologue. The administration of the ANELT is recorded on audiotape for later scoring and it takes 15-20 minutes to administer. The response of the participant for each item is rated on two 5-point scales (0-4), evaluating the *understandability of the message* and the *intelligibility of the utterance (sic)* (scale A and B, respectively). The maximum score for both understandability and intelligibility is 40.

Finally, we performed a quantitative analysis of the data (Nicholas & Brookshire, 1993). Language samples were transcribed and analysed for number of words, number of correct information units (CIUs), the percentage of correct information units (%CIUs), and efficiency (e.g., CIUs/minute, words/minute). A speech and language pathology student previously trained in clinical and experimental linguistics completed the transcription of the speech samples. For information on scoring the scales and analysing word and CIU measures, see the Data analysis section.

Everyday memory questionnaire (EMQ, far transfer). We adapted the everyday memory questionnaire developed by Sunderland, Harris, and Baddeley (1983) to aphasia. Thirty-one examples of memory difficulties were included in the present questionnaire (see the Appendix – Table A3). A close relative or partner of the participants was asked to judge how often a difficulty occurs in the participant's activities of daily living, using a 5-point rating scale (where 0 indicates *never* and 4 indicates *quite often*). Difficulties were grouped under the headings "Speech" (e.g., "Confusing the names of common things or using the wrong names"), "Faces and places" (e.g., "Forgetting where s/he has put something, losing things around the house"), "Actions" (e.g., "Forgetting to do some routine thing that s/he would normally do once or twice in a day"), and "Learning new things" (e.g., "Unable to pick up a new skill such as a game or working some new gadget after s/he has practiced once or twice"). Items followed each other in a fixed order. Ratings for each type of memory difficulty were summed and analysed.

Control task: Oral word reading. As oral word reading potentially does not tax WM majorly, we used it as a control task to test that possible improvements on the outcome measures were specifically related to the WM training. The task consisted of simple words (25 items) and compound words (20 items) with varying length (1-4 syllables) and frequency (low frequent vs. high frequent), as well as one-syllable pseudo-words (15 items). We selected words from Lorenz, Heide, and Burchert (2014) and pseudowords from the subtest of LeMo 2.0 (Stadie, Cholewa, & De Bleser, 2013). Items were printed separately on A4 format paper sheets (font size 44) and presented one at a time in a fixed order. Participants were instructed to read aloud the words, each within a 10 seconds limit. If there was no response within this time limit, the examiner proceeded to the next item. The task took approximately 10 minutes. The participants' responses were audio recorded and later transcribed and scored by two SLP students (one of them previously mastered in clinical and experimental linguistics). The total number of correctly read items was calculated.

6.2.1.3 Motivation questionnaire (MQ)

We developed a self-report motivation questionnaire based on Jaeggi et al. (2011) and McAuley, Duncan, and Tammen (1989). The questionnaire consisted of 10 questions assessing the participants' *interest/enjoyment*, *perceived competence*, and *effort/importance* while performing the training. Participants responded on a 7-point Likert scale from 1 (low degree of approval) to 7 (high degree of approval). Four questions focused on interest/enjoyment (e.g., "How much did you enjoy the activity today?" – 1: *not at all*, 7: *a lot*), three questions on perceived competence (e.g., "How satisfied are you with your performance today?" 1: *not satisfied at all*, 7: *very satisfied*) and three on effort/importance (e.g., "How much effort did you put into this today?" – 1: *nothing at all*, 7: *a lot*, see Appendix – Table A4). Participants completed this questionnaire after each session. Experimenters were instructed to read aloud the questions and note the response of the participant. They were also instructed to explain questions if needed but to avoid influencing the participants' response selection in any way. We calculated the mean score for each factor for each session to capture the changes in motivation over time and possibly relate them to the performance pattern in the training tasks.

Similar to the training sessions, test sessions were conducted by an SLP and by SLP students. The same person(s) for each participant conducted test and training sessions. Importantly, for outcome measures that were obtained by scoring/rating the participant's responses by the experimenter (i.e., that were not computer generated) the responses were also scored by an independent experimenter and tested for inter-rater reliability (for details of this step, see the Data analysis and Results section). All experimenters participated in a 5×2 hour training provided by the first author of the paper regarding the conduction, administration, and scoring of the tasks.

6.2.2 Data analyses

6.2.2.1 Performance change in the training tasks and outcome measures

Individual performances on the training tasks were tested using non-parametric Spearman correlations. Based on Vallat et al. (2005), Fisher's exact test was used to compare performance in the two baselines (to demonstrate stability, p should > .1, two-tailed).

We used Fisher's exact and McNemar's test ($p < .05$, one-tailed) to compare performance on pretest and posttest, by taking the aggregate data obtained on two occasions for both pretest and posttest (note that data was obtained on two occasions only in the experimental tasks). To investigate long-term maintenance of potential effects (i.e., compare performance between posttest and follow up), we used Fisher's exact and McNemar's chi square test. Group level performance on the outcome measures was analysed with Wilcoxon signed rank test ($p <$ 05, one-tailed). The relationship between the improvement in the training task and changes in motivation factors was tested with Spearman correlation (*rho*).

6.2.2.2 Inter-rater reliability

Inter-rater reliability represents the correspondence between raters' scores, thus indicates a measure of reliability for the collected data (Morgan & Morgan, 2008). Inter-rater reliability of the sentence comprehension tests and the running span (i.e., in case of dichotomous data) was examined using proportion scoring agreement, by dividing the number of agreements by the number of agreements plus disagreements (Franklin, Allison, & Gorman, 2014; Morgan & Morgan, 2008). The running span and the sentence comprehension tests were scored on 55% of the samples (range 33-75%) by two experimenters who were both present during the assessment (i.e., the trainer and an independent but not blind assessor). Inter-rater reliability of the oral tasks' measures was determined using an ICC two-way random effects model (ICC(2,k)) (Franklin et al., 2014). The oral tasks (i.e., ANELT and word reading) were audiotaped and transcribed; 100% of the oral word reading, 33% and 17% of the ANELT speech samples (qualitative and the quantitative analysis, respectively) were analysed by two independent experimenters who were also blind to the study phase.

6.3 Results

Inter-rater reliability was excellent for all tasks: mean proportion scoring agreement was 1 for the running span, .98 for the Token, .98 for the Sätze verstehen, and 1 for the TROG-D. ICC(2,k) was .96 for the oral word reading, .74 for the ANELT (Scale A), and ranged between .85-.99 for quantitative measures of the ANELT. All discrepancies were resolved by consensus prior to analysis.

Participants demonstrated unstable baseline in some conditions: R.D. and Z.A. were not stable in the 3-back condition of the *n*-back with letters (Fisher's exact test, p = .042 and p = .035, respectively). V.O. was not stable in the running span (Fisher's exact test, p = .042). To get a more accurate picture of the participants' performance, we took the aggregate data obtained on two occasions for both pretest and posttest in the tasks.

6.3.1 Training tasks

To analyse performance at the individual level (Figure 7), correlations between number of training sessions and mean difficulty level at a session were calculated using Spearman correlation coefficient. V.O. showed a significant increase in performance in both the auditory and the visual training (*rho* = 1, $p < .01$ and *rho* = .786, $p < .05$, respectively), whereas R.D. and Z.A. only improved in the first training comprising the visual *n*-back task (*rho* = .905, $p < .01$ and *rho* = 1, $p < .01$, respectively).

Comparisons between posttest and follow up revealed changes in the participants' performance over time. With respect to the *n*-back with pictures, R.D. showed a significant increase in performance in 2-back (Fisher's exact test, $p = .045$) and V.O. showed a significant decrease in performance in 3-back (Fisher's exact test, $p = .001$). With respect to the *n*-back with spoken words, R.D. showed a tendency for a decrease in performance in the 2-back condition (Fisher's exact test, $p = .085$). In summary, participants consistently showed performance increases during training. However, improvement was not consistently maintained until 6-weeks after posttest.

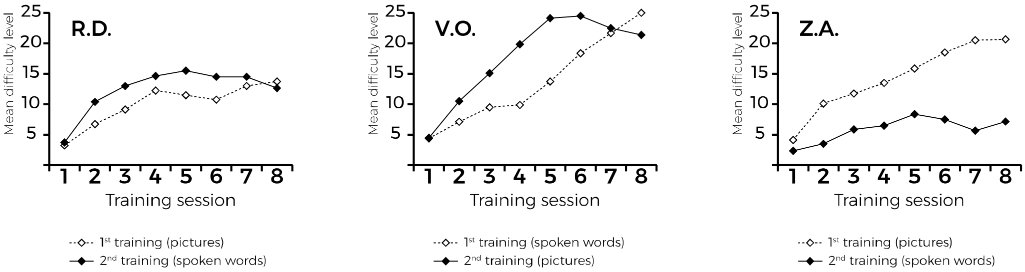

Figure 7 Performance on the training tasks during the 16 sessions of training
V.O. improved significantly ($p < .05$) across sessions in both training tasks, whereas R.D. and Z.A. improved statistically significantly only in the first training comprising the *n*-back with pictures ($p < .05$).

6.3.2 Outcome measures

Overview of the results of the outcome measures is in Table 6.

6.3.2.1 *WM 1: N-back with letters*

Aggregated scores showed that R.D. improved significantly in 2-back and 3-back (Fisher's exact test, $p = 03.$ and $p < .001$, respectively), V.O. improved significantly in 2-back and 3-back (Fisher's exact test, $p = .024$ and $p = .034$, respectively), whereas Z.A. did not improve in any of the conditions. Group level analysis showed no significant improvement in any of the conditions ($p > .05$ for all conditions). Note that in the 1-back condition R.D. and V.O. were close to ceiling already at the beginning of the training.

6.3.2.2 *WM 2: Running span*

Analysis of the number of correct trials showed that none of the participants improved in the running span task (Fisher's exact test, $p > .05$ for all participants). Group level analysis showed a tendency level improvement in the task ($Z = -1.60$, $p = .054$)

6.3.2.3 *Spoken sentence comprehension 1: TROG-D*

Z.A. significantly improved between pretest and posttest (McNemar chi square = 5.281, $p = .011$) and the improvement was maintained also at follow up (comparing posttest and follow up: McNemar chi square, $p > .1$); V.O. showed a tendency level improvement (McNemar chi square = 3.6, $p = .054$) between pretest and follow-up; whereas R.D. did not improve. Group level analysis on total scores showed a tendency level improvement between pretest and follow-up ($Z = -1.60$, $p = .054$) as well as posttest and follow-up ($Z = -1.34$, $p = .09$). Thus, we detected a tendency for improvement on the comprehension of grammatical structures coupled with heterogeneous individual performance patterns.

6.3.2.4 *Spoken sentence comprehension 2: Token test*

Comparing pretest and posttest performance, a tendency towards improvement was found for R.D. and Z.A. (McNemar chi square = 2.37, $p = .061$

and chi square = 2.207, $p = .068$, respectively), whereas no significant change in performance was found for V.O. Group level analysis showed a tendency level improvement in the task ($Z = -1.41$, $p = .07$).

6.3.2.5 Spoken sentence comprehension 3: Sätze verstehen

R.D. significantly improved in the comprehension of number-marked OVS sentences (McNemar test chi square = 7.53, $p < .01$) and non-canonical structures (McNemar chi square = 6.618, $p < .01$); V.O. significantly improved in the comprehension of canonical structures (McNemar chi square = 8.33, $p < .01$) and showed a tendency for increase in the total score (McNemar chi square = 1.75, $p = .09$); whereas Z.A. did not improve in any of the conditions. At group level they showed a tendency for increase in the comprehension of right-branching subject relative clauses (Z = -1.34, $p = .09$) and centre-embedded object relative clauses (Z = -1.60, $p = .054$), and in the total score ($Z = -1.60$, $p = .054$).

6.3.2.6 Functional communication: ANELT

Analysis of the understandability scores (scale A) showed a significant positive change in V.O.'s functional communication ($U = 16.5$, $p < .01$). R.D. and Z.A. also showed an increase in performance but these were not statistically significant. Group level analysis showed a tendency level improvement in the task ($Z = -1.60$, $p = .054$).

Analysis of quantitative measures complemented this picture: Z.A. significantly improved in number of words ($U = 17$, $p < .01$) and CIUs ($U = 19$, $p < .05$), V.O. significantly improved in percentage of CIUs ($U = 19$, $p < .05$) and showed a statistical tendency for improvement in CIUs/min ($U = 25$, $p = .056$), whereas R.D. did not show statistically significant improvement in the task. At group level they showed a tendency level increase in performance according to the CIUs, %CIUs, and CIUs/min ($Z = -1.60$, $p = .054$ for all three measures).

Table 6 ***Improvements on the outcome measures***

	Case							
	R.D.		V.O.		Z.A.		Group	
Outcome measure	Pre-post	FU	Pre-post	FU	Pre-post	FU	Pre-post	FU
N-back with letters								
1-back	⬜		⬜		⬜		⬜	
2-back	🟦		🟦		⬜		⬜	
3-back	🟦		🟦		⬜		⬜	
Running span	⬜		⬜		⬜		🔷	
TROG-D	⬜	⬜	⬜	🔷	🟦	🟩	⬜	🔷
Token test	🔷		⬜		🔷		🔷	
Sätze verstehen								
Short irreversible	⬜		⬜		⬜		⬜	
Long irreversible	⬜		⬜		⬜		⬜	
Case-marked SVO	⬜		⬜		⬜		⬜	
Case-marked OVS	⬜		⬜		⬜		⬜	
Number-marked SVO	⬜		⬜		⬜		⬜	
Number-marked OVS	🟦		⬜		⬜		⬜	
Right-branching SRC	⬜		⬜		⬜		🔷	
Right-branching ORC	⬜		⬜		⬜		⬜	
Centre-embedded SRC	⬜		⬜		⬜		⬜	
Centre-embedded ORC	⬜		⬜		⬜		🔷	
Total	⬜		🔷		⬜		🔷	
Canonical	⬜		🟦		⬜		⬜	
Non-canonical	🟦		⬜		⬜		⬜	
ANELT								
Understandability	⬜		🟦		⬜		🔷	
Number of words	⬜		⬜		🟦		⬜	
CIUs	⬜		⬜		🟦		🔷	
%CIUs	⬜		🟦		⬜		🔷	
CIUs/min	⬜		🔷		⬜		🔷	
Words/min	⬜		⬜		⬜		⬜	
EMQ								
Speech	🔷		🔷		⬛		⬜	
Learning new things	⬜		⬜		🔷		⬜	

Note. FU: follow-up; TROG-D: German version of the Test for the Reception of Grammar; SVO: subject-verb-object; OVS: object-verb-subject; SRC: subject relative clauses; ORC: object-relative clauses; ANELT: Amsterdam-Nijmegen Everyday Language Test; CIUs: correct information units; %CIUs: percent of correct information units; CIUs/min: correct information units per minute; EMQ: Everyday memory questionnaire; dark blue and light blue indicate a statistically significant improvement and a tendency for improvement, respectively; dark grey indicates a statistically significant decrease in performance; light grey cells indicate that data was available, but did not produce statistically significant change; green indicates maintenance of performance at follow-up. Note that performance was close to ceiling already at the beginning of the training in the letter 1-back, EMQ 'Speech', and 'Short irreversible', 'Long irreversible', 'Case-marked SVO', 'Number-marked SVO' for R.D. and V.O., and 'Long irreversible' for Z.A.

6.3.2.7 *Everyday memory questionnaire*

Ratings for each type of memory failure were summed. We only analysed the total score in the section 'Speech' for each participant and the total score in the section 'Learning new things' for Z.A., because in the other sections there was virtually no error reported. Scores in 'Speech' showed a tendency level decrease in memory failures for R.D. and V.O. ($Z = -1.53$, $p = .063$ and $Z = -1.41$, $p = .078$, respectively) but a significant increase in memory failures for Z.A. ($Z = -1.90$, $p = .028$). Scores in ‘Learning new things’ showed a statistically significant decrease in memory failures for Z.A. ($Z = -1.73$, $p = .041$).

6.3.2.8 Control task: Oral word reading

Pre-post comparisons for oral word reading showed that the participants’ performance remained stable on the task (Fisher’s exact test, $p > .05$ for all participants).

6.3.3 Motivation questionnaire

Mean scores were calculated for each motivation factor (i.e., interest/enjoyment, perceived competence, and effort/importance), based on each participant’s ratings that were elicited in each session. Changes in the motivation scores were analysed on a descriptive basis as well as statistically compared to the changes in performance on the training tasks by means of Spearman rank correlation coefficient. Changes in each motivation factor can be seen in Figure 8 for each participant.

R.D. and V.O. reported moderate to high interest in the training tasks; their interest levels remained stable throughout the training. Both participants put great effort into the training tasks over the whole training period. M.N, however, showed a considerable fluctuation in all motivation factors. She reported greatly varying interest levels, with a decreasing tendency in the second training block. In addition, she reported generally lower effort levels than the other two participants during the whole training period.

For Z.A., changes in perceived competence were significantly associated with changes in performance in the second training block ($rho = .89$, $p = .007$), suggesting that she was able to evaluate her performance on the training task. For R.D., changes in effort were significantly associated with changes in performance in the first training block ($rho = -.817$, $p = .025$), suggesting that the more effort she put into the training task, the more she improved. All the other comparisons between changes in motivation factors and in performance on the training tasks were not statistically significant.

Mean interest and perceived competence scores showed a positive correlation (at the level of tendency) both in the first and the second training block (*rho* = .67, *p* = .068 and *rho* = .67, *p* = .097, respectively) for V.O., a positive correlation at the level of tendency in the first training block (*rho* = .66, *p* = .078) and a significant positive correlation in the second training block (*rho* = .96, *p* < .001) for Z.A., and a tendency level negative correlation in the second training block for R.D. (*rho* = -.66, *p* = .073). In addition, mean effort scores showed a significant positive correlation with mean interest scores and a tendency for a positive correlation with mean perceived competence scores (*rho* = .852, *p* = .007 and *rho* = .66, *p* = .076, respectively) in the first training block for R.D.

6.3.4 Relationship between initial cognitive, linguistic abilities, and training outcome

To investigate the potential relationship between initial WM, language comprehension abilities, and improvement on spoken sentence comprehension after training, we performed a Spearman rank correlation on the data of the current study and the data collected in our previous study (Zakariás et al., 2018). This comparison was possible, because some of the WM tasks and the spoken sentence comprehension tests used in the two studies were identical in terms of task design and procedures (i.e., *n*-back with letters), or were standardized in both languages (i.e., TROG). Results of the analysis revealed a relationship between initial spoken sentence comprehension ability and training outcome (*rho* = -.754, *p* = .084), suggesting that the more severe the spoken sentence comprehension deficit was at the beginning of training, the more it improved after training.

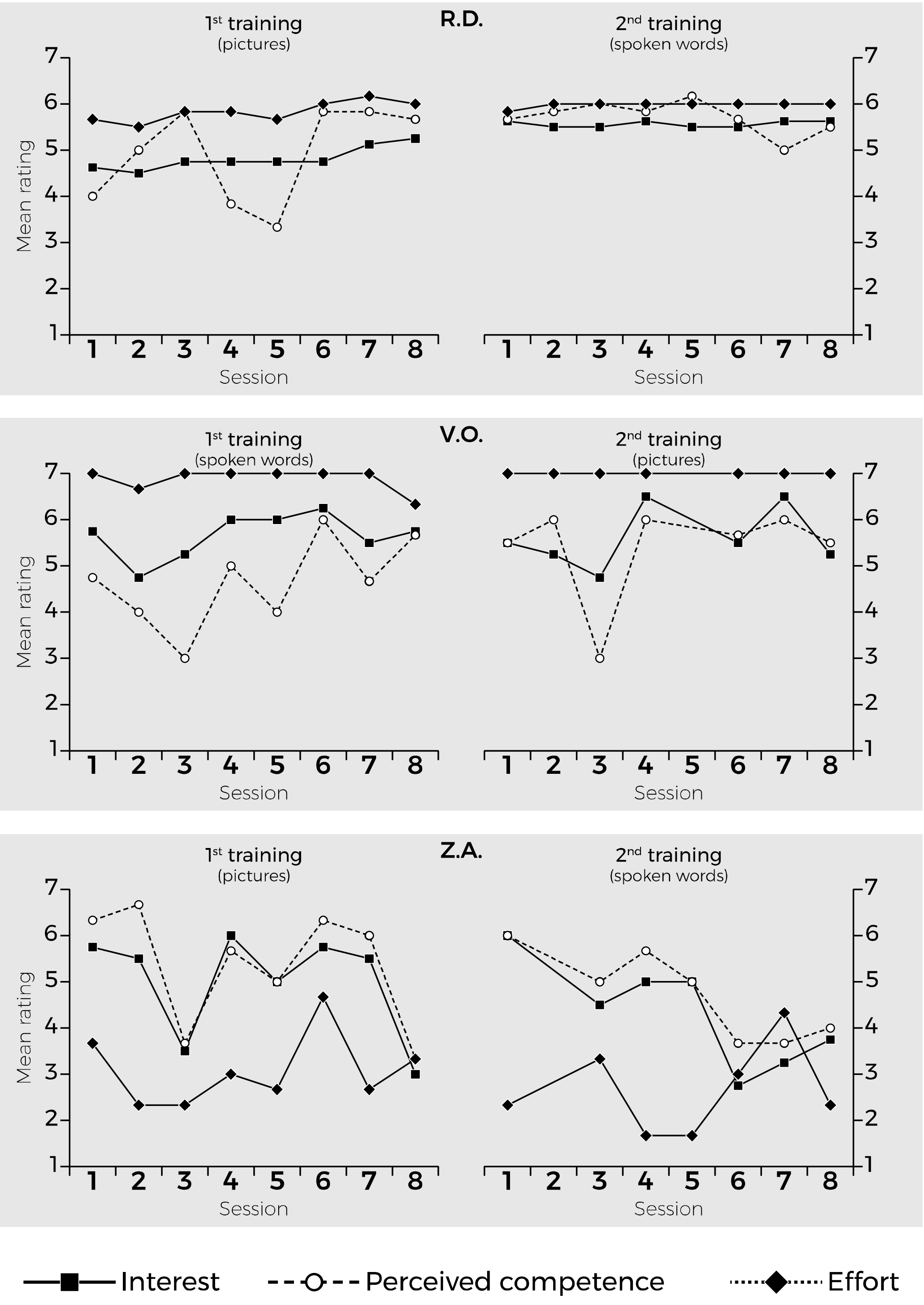

Figure 8 Mean scores of interest/enjoyment, perceived competence, and effort/importance over the sixteen sessions of training

6.4 Discussion

In this study, we investigated whether WM training effects transferred to unpractised WM and spoken sentence comprehension tasks, as well as to functional communication and everyday memory. The training targeted different components of WM, such as maintaining and updating WM representations and interference control. Consistent with previous results in related studies (e.g., Eom & Sung, 2016; Paek & Murray, 2015), participants showed improvements in the training tasks. However, the patterns of improvement were not consistent across the two training blocks: two participants improved only in the first block comprising the *n*-back with pictures. Because these two participants practised the training tasks in the same order (first *n*-back with pictures, and then *n*-back with spoken words), it is not possible to tease apart stimulus type and practice order effects. Nevertheless, performance patterns suggest different underlying mechanisms for the lack of improvement in the second training block (i.e., *n*-back with spoken words) for these two cases: R.D. seemed to reach asymptote by the fourth session in the second training block and changes in her performance may have gone undetected due to statistical properties of the Spearman correlation coefficient (i.e., it measures linear relationships) used to test for performance improvements. In case of Z.A., however, results of the motivation questionnaires suggest that the lack of improvement may be due to decreasing motivation and engagement with the training activity and/or to the fact that an *n*-back task including spoken stimuli was more difficult for her than the one including pictures. In sum, participants improved in the training tasks, and more importantly, the amount of improvement was comparable to that observed in healthy young adults in similar tasks (Novick et al., 2014).

Consistent with our previous study (Zakariás et al., 2018), we detected a mixed pattern of transfer. With respect to far transfer, all participants improved at least in three outcome measures out of the five. These included measures of spoken sentence comprehension (i.e., TROG-D, Sätze verstehen, Token test), functional communication (ANELT), and everyday memory (Everyday memory questionnaire). Crucially, follow-up results suggest that improvements in spoken sentence comprehension were also maintained at six weeks after completion of the study for two participants. Although psychometric properties are not known for all the far transfer tasks we used, results of a previous study indicates that the TROG has high test-retest reliability ($r = .99$ in a group of five people with different types and severity of aphasia, see Zakariás et al., 2018). Furthermore, the two parallel versions of the ANELT correlate with each other to a very high degree (Blomert et al., 1994). Both the TROG and ANELT could be used to evaluate treatment effects in spoken sentence comprehension and functional communication respectively. The current results are in line with previous findings of Eom and Sung (2016) and Zakariás et al. (2018), who also found improvement after WM training on spoken sentence comprehension. To our knowledge, this is the first study showing transfer effects after WM training on functional communication in aphasia.

With regards to the specificity of transfer effects on spoken sentence comprehension, we detected improvements on: (1) non-canonical number marked (object-verb-subject) sentences, (2) non-canonical sentences including varying syntactic structures, such as case marked and number marked object-verb-subject sentences and right-branching and centre-embedding object relative clauses, and (3) canonical sentences including case marked and number marked subject-verb-object sentences and right-branching and centre-embedding subject relative clauses in some individuals. What mechanisms can account for these improvements? A number of studies have suggested that WM supports parsing and interpretation (i.e., construction of the syntactic structure of a sentence and the use of this structure to determine sentence meaning, respectively) and is majorly involved in processing syntactically complex sentences, such as object-relative clauses (see Just and Carpenter, 1992 for review; Haarmann et al., 1997). Just and Carpenter (1992) argued that the same pool of WM resources tapped by WM tasks is also used in sentence processing. By contrast, Caplan and colleagues (2013) proposed that memory mechanisms captured by traditional WM tasks (e.g., WM span and *n*-back) do not support the on-line, automatic processing of syntactic information, but are engaged in a later stage of sentence comprehension, namely the revision of the previously encountered, inaccurately interpreted information, and the use of the product of the comprehension to perform a task (e.g., in a picture-matching task keeping sentence meaning in mind while analysing and interpreting the visual scenes and comparing them to the meaning of the sentence). This is called post-interpretive or expanded comprehension (Caplan et al., 2013). Our results showing improvements on both canonical and non-canonical structures after WM training in IWA suggest that the use of WM in sentence processing is less specific to syntactic structures but may play a role in more general processes involved in the later stage of sentence comprehension (post-interpretive comprehension). This aspect is particularly important in everyday tasks that involve sentence comprehension (e.g., extracting meaning from conversations, understanding information from the news).

With respect to near transfer, the pattern of improvements in the WM tasks suggests that very near transfer occurred. Gains detected in the *n*-back with letters but not in the running span suggest that the improvements were task specific rather than process specific.

What linguistic and cognitive profiles make participants likely benefit from WM training? According to Caplan et al. (2013) and Fedorenko (2014), WM provides extra computational resources or alternative routes for resolving the possible problems encountered during language comprehension. These theories also imply that WM training can be most beneficial for IWA demonstrating substantial WM deficits, because potential improvements on WM allow them to utilize extra resources during language comprehension. To investigate the potential relationship between initial WM, language comprehension abilities, and improvement on spoken sentence comprehension after training, we performed a Spearman rank correlation on the data of the current study and the data collected in our previous study (Zakariás et al., 2018). We could not find, however, any relationship between initial WM and improvement on spoken sentence comprehension after training. The results of this joint analysis revealed a relationship between initial spoken sentence comprehension ability and training outcome (*rho* = -.754, *p* = .084), suggesting that the more severe the spoken sentence comprehension deficit was at the beginning of training, the more it improved after training. The lack of a significant correlation between initial WM and improvement in spoken sentence comprehension could be due to the small number of data entered into the analyses.

Results of Zakariás et al. (2018) and the present study also suggest that the extent of improvement in an *n*-back training task is not necessarily proportional to the improvement in the transfer tasks in aphasia, as proposed by others investigating transfer in other populations, such as children and healthy young adults (e.g., Jaeggi et al., 2014; Waris et al., 2015, respectively). This may be related to the fact that in aphasia the extent of improvement on the trained processes and a complex interaction of intact and impaired functions affect training outcome and benefit to untrained functions.

The lack of significant improvement in everyday memory can be explained by the fact that participants had only mild impairments already at the beginning of the training, and therefore, there was not enough room for improvement. However, it is still difficult to interpret the negative change in performance for Z.A. One possible explanation could be that Z.A. and her daughter (who rated the everyday memory questionnaire) did not have everyday contact and communication during the time of study (i.e., they did not live together). Insufficient communication or biases might have led to false estimation (i.e., in this case overestimation) of the memory failures.

6.4.1 Limitations of the present study

There are a few limitations of the current study that could inform future research on WM training in aphasia. Because we assessed experimental outcome measures using a multiple baseline design, we did not include a control group. For a few conditions, however, baseline variability was too large to provide stable baseline estimates, which could have led to some outcome effects overestimated or going undetected. For these conditions therefore, both significant and non-significant effects should be interpreted with caution. In particular, experimental tasks used for multiple baseline assessments can benefit from more baselines.

A further concern relates to the Sätze verstehen test to assess specificity of transfer effects. The lack of significant effects in most conditions in this measure (despite significant effects on aggregated scores) could be a result of low statistical power due to only a small number of sentences per condition. Future research with a larger number of examples of each sentence type might allow for a better understanding of the underlying mechanisms of transfer effects on sentence comprehension.

Although the single-case experimental design employed ensures valid estimation of effects at the individual level, the large individual differences call for future research to clarify how far these results are generalizable to population level of IWA. A more feasible goal for prospective studies would be to identify a few mechanisms that may generalize to at least a subpopulation of IWA. To achieve this goal, more detailed analyses of individual differences on larger but yet homogenous samples may be required.

In conclusion, the present study is the first systematic investigation of transfer effects of training higher-level WM functions on functional communication and everyday functions in aphasia. Our results suggest that WM can improve through intensive computerized training in chronic aphasia and this improvement can lead to improvements in spoken sentence comprehension and functional communication.

Acknowledgements

We are grateful to R.D., V.O., and Z.A. for their participation, and thank Leonie Lampe for her help with preparing the assessment and training materials, Helene Gerda Killmer, Susanne Pelke, Thomas Hinsdorf, and Vanessa Dawn Löffler for their help with the data collection, Jan Ries for creating the figures, and Attila Keresztes for his helpful insights during the study.

6.5 Appendix

Table A1 *Stimuli and procedure of the listening span task (based on Tompkins et al., 1994)*

Level 1	
Set 1	Set 1
Fische schwimmen im **Wasser** (T)	Fish swim in **water** (T)
Set 2	Set 2
Menschen putzen ihre Zähne mit einem **Löffel** (F)	People clean their teeth with a **spoon** (F)
Set 3	Set 3
Es gibt Gras im **Park** (T)	There is grass in a **park** (T)
Level 2	
Set 4	Set 4
Im Kino schaut man einen **Film** (T)	In the cinema you can watch a **movie** (T)
Milch ist **rot** (F)	Milk is **red** (F)
Set 5	Set 5
Kinder mögen **Eis** (T)	Children like **ice-cream** (T)
Schweine können **fliegen** (F)	Pigs can **fly** (F)
Set 6	Set 6
Die Erde hat einen **Mond** (T)	The earth has a **moon** (T)
Papier kann **kochen** (F)	Paper can **boil** (F)
Level 3	
Set 7	Set 7
Zucker ist **süß** (T)	Sugar is **sweet** (T)
Berlin liegt neben **Rom** (F)	Berlin is next to **Rome** (F)
Menschen essen Frühstück am **Abend** (F)	People eat breakfast in the **evening** (F)
Set 8	Set 8
Menschen schlafen in einem **Bett** (T)	People sleep in a **bed** (T)
Karotten können **tanzen** (F)	Carrots can **dance** (F)
Äpfel wachsen am **Baum** (T)	Apples grow on **tree** (T)
Set 9	Set 9
Deutschland hat einen **König** (F)	Germany has a **king** (F)
Kühe essen gerne **Gras** (T)	Cows like to eat **grass** (T)
Ein Kapitän steuert ein **Schiff** (T)	A captain steers a **ship** (T)
Level 4	
Set 10	Set 10
Giraffen haben einen langen **Hals** (T)	Giraffes have a long **neck** (T)
Stühle können **essen** (F)	Chairs can **eat** (F)
Ein Fahrrad ist langsamer als ein **Bus** (T)	A bicycle is slower than a **bus** (T)
Auf Konzerten gibt's **Musik** (T)	At concerts there is **music** (T)
Set 11	Set 11
Häuser sind aus **Käse** (F)	Houses are made of **cheese** (F)
Katzen mögen **schlafen** (T)	Cats like to **sleep** (T)
Fleischer machen **Brot** (F)	Butchers make **bread** (F)
Worte bilden einen **Satz** (T)	Words form a **sentence** (T)
Set 12	Set 12
Hasen können **lesen** (F)	Rabbits can **read** (F)
Hühner essen **Holz** (F)	Chickens eat **wood** (F)
Kinder gehen in die **Schule** (T)	Children go to **school** (T)
Ein Zug fährt auf einem **See** (F)	A train drives on a **lake** (F)

Table A1 ***(Continued)***

Level 5	
Set 13	Set 13
Hamster können **reden** (F)	Hamsters can **talk** (F)
Blei ist schwerer als **Papier** (T)	Lead is heavier than **paper** (T)
Eis ist **heiß** (F)	Ice is **hot** (F)
Häuser haben eine **Tür** (T)	Houses have a **door** (T)
Blumen brauchen **Licht** (T)	Flowers need **light** (T)
Set 14	Set 14
Menschen haben eine **Nase** (T)	People have a **nose** (T)
Saft enthält viel **Fett** (F)	Juice contains lots of **fat** (F)
Eine Rose ist ein **Tier** (F)	A rose is an **animal** (F)
Eine Maus ist kleiner als ein **Hund** (T)	A mouse is smaller than a **dog** (T)
Ein Auto kann **fahren** (T)	A car can **race** (T)
Set 15	Set 15
Ein Schuh hat einen **Kopf** (F)	A shoe has a **head** (F)
Pferde rennen im **Himmel** (F)	Horses run in the **sky** (F)
Eine Uhr zeigt die **Zeit** (T)	A clock shows the **time** (T)
Ein Buch kann **laufen** (F)	A book can **run** (F)
Ein Lachs ist ein **Fisch** (T)	A salmon is a **fish** (T)
Practice sets	
Level 1	
Set 1	Set 1
Menschen lesen Bücher im **Ofen** (F)	People read books in the **oven** (F)
Set 2	Set 2
Hunde haben einen **Schwanz** (T)	Dogs have a **tail** (T)
Level 2	
Set 3	Set 3
Die Zwiebel ist ein **Obst** (F)	People read books in the **oven** (F)
Ein Elefant hat einen **Rüssel** (T)	Dogs have a **tail** (T)
Set 4	Set 4
Menschen sehen Löwen im **Zoo** (T)	An elephant has a **trunk** (T)
Zwölf ist gleich **Duzend** (T)	People see lions in the **zoo** (T)

Note. Target items are highlighted with bold; T = True; F = False.

The task was modified for use in aphasia. Target items (in German) are controlled for frequency and length (high frequency, one- or two-syllable long words). The task includes only simple sentences. Each stimulus set of the task was checked by a native German speaker for phonological and semantic similarity to avoid interference across items to ensure the highest recall rate possible.

Procedure. Immediately after hearing each sentence, participants were asked to judge it as true or false by pointing a check mark or cross on a sheet of paper. Concurrently, they were asked to retain the final word of each sentence in each set for spoken recall, immediately after the entire set was presented. Practice sets were included to familiarize participants with the task procedure and to ensure that they understood the task. Practice sets were performed before presenting the experimental sets.

Table A2 *Stimuli of the training tasks with English translations in square brackets*

N-back with pictures

Semantic category	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8
Animal	Elefant [elefant]	Esel [donkey]	Zebra [zebra]	Giraffe [giraffe]	Katze [cat]	Maus [mouse]	Bär [bear]	Fisch [fish]
Furniture	Stuhl [chair]	Sofa [sofa]	Kühlschrank [fridge]	Herd [stove]	Fernseher [television]	Schreibtisch [desk]	Kommode [chest of drawers]	Lampe [lamp]
Clothes	Hose [trausers]	Stiefel [boots]	Mütze [cap]	Rock [skirt]	Bluse [blouse]	Socke [socks]	Hut [hat]	Kleid [dress]
Body parts/tools	Säge [saw]	Mund [mouth]	Daumen [thumb]	Nase [nose]	Finger [finger]	Leiter [ladder]	Fuß [foot]	Ohr [ear]
Vehicles/musical instruments	Auto [car]	Hubschrauber [helicopter]	Zug [train]	Bus [bus]	Geige [violin]	Horn [horn]	Trommel [drum]	Klavier [piano]
Food	Kartoffel [potato]	Karotte [carrot]	Banane [banana]	Kürbis [pumpkin]	Kirsche [cherry]	Tomate [tomato]	Salat [salat]	Pilz [mushroom]
Toys/household items	Drachen [kite]	Kreisel [spinning top]	Brille [brille]	Bleistift [pencil]	Fahne [flag]	Puppe [doll]	Ball [ball]	Tennisschläger [tennis racket]
Kitchen/home	Glas [glass]	Tasse [cup]	Vase [vase]	Flasche [bottle]	Kerze [candle]	Löffel [spoon]	Gabel [fork]	Schüssel [bowl]

N-back with spoken words

Semantic category	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8
Animals	Bär [bear]	Frosch [frog]	Maus [mouse]	Kuh [cow]	Hase [rabbit]	Katze [cat]	Biene [bee]	Löwe [lion]
Vegetables/ drinks	Gurke [cucumber]	Erbse [pea]	Bohne [bean]	Salat [salat]	Lauch [leek]	Pilz [mushroom]	Tee [tea]	Mais [corn]
Vehicles	Schlitten [sled]	Taxi [taxi]	Traktor [tractor]	Fähre [ferry]	Kran [crane]	Bus [bus]	Kahn [boat]	Boot [boat, ship]
Furniture/house	Couch [couch]	Dach [roof]	Schrank [cupboard]	Thron [throne]	Sessel [armchair]	Regal [shelf]	Sofa [sofa]	Truhe [chest]
Musical instruments/toys	Geige [violin]	Trommel [drum]	Orgel [organ]	Klavier [piano]	Maske [mask]	Puppe [doll]	Schaukel [swing]	Kreisel [spinning top]
Tools	Axt [axe]	Stift [pin]	Sieb [sieve]	Beil [axe]	Zange [pliars]	Hammer [hammer]	Säge [saw]	Bürste [brush]
Clothes	Anzug [suit]	Bluse [blouse]	Mütze [cap]	Jacke [jacket]	Schal [scarf]	Tuch [scarf]	Hemd [shirt]	Rock [skirt]
Professions	Boxer [boxer]	Schwimmer [swimmer]	Bauer [farmer]	Tänzer [dancer]	Forscher [researcher]	Bäcker [baker]	Friseur [hairdresser]	Gärtner [gardener]

Table A3 *Statements of the Everyday memory questionnaire in German and English*

Speech		
1	Er/Sie bringt Namen von Freunden oder Verwandten durcheinander oder nennt sie bei einem falschen Namen.	He/She confuses the names of friends or relatives or calls them by the wrong names.
2	Er/Sie bringt Namen von geläufigen Dingen durcheinander oder nennt sie bei einem falschen Namen.	He/She confuses the names of common things or uses the wrong names.
3	Ihm/Ihr liegen Wörter auf der Zunge. Er/Sie kennt das Wort aber kann es nicht finden.	He/She has words on the tip of his/her tongue. He/She knows what it is but can't quite find it.
4	Er/Sie vergisst Dinge, die einige Minuten zuvor gesagt wurden. Zum Beispiel etwas, was der Ehepartner oder ein Freund gerade gesagt hat.	He/She forgets something that he/she was told a few minutes earlier; for instance, something his/her spouse or a friend has just said.
5	Er/Sie vergisst, was ihm/ihr gestern oder vor einigen Tagen erzählt wurde.	He/She forgets something he/she was told yesterday or a few days earlier.
6	Er/Sie wiederholt Dinge, die er/sie kurz zuvor gesagt hat oder stellt die gleichen Fragen mehrmals.	He/She repeats something he has just said or asks the same question several times.
7	Er/Sie vergisst, was er/sie gerade gesagt hat. Dabei sagt er/sie möglicherweise etwas wie “Worüber habe ich gerade gesprochen?”	He/She forgets what he has just said. Thereby, he possibly says something like “What have I just been talking about?”
8	Er/Sie ist nicht in der Lage, dem zu folgen, was jemand erzählt. In einem Gespräch verliert er/sie den Faden.	He/She loses track of what someone tells him/her. During a conversation, he loses the thread.
9	Er/Sie beginnt etwas zu sagen, vergisst dann aber, worüber er/sie eigentlich sprechen wollte.	He/She starts to say something, but then forgets what he actually wanted to talk about.
10	Er/Sie schweift ab und spricht über unwichtige und irrelevante Dinge.	He/She gets off the point and speaks about unimportant or irrelevant things.
11	Er/Sie vergisst, anderen wichtige Dinge zu erzählen. Zum Beispiel vergisst er/sie, eine Nachricht weiterzuleiten oder jemanden an etwas zu erinnern.	He/She forgets to tell others something important. For instance, he forgets to pass on a message or to remind someone of something.
12	Er/Sie bringt Details von dem durcheinander, was ihm/ihr jemand erzählt hat.	He/She mixes up the details of what someone has told him.
13	Er/Sie wiederholt Geschichten oder Witze, die er/sie bereits erzählt hat.	He/She repeats a story or joke he has said before.
Faces and places		
14	Er/Sie vergisst, wo er/sie Dinge hingelegt hat. Er/Sie verlegt Dinge im Haus.	He/She forgets where he put something. He misplaces things around the house.
15	Er/Sie erkennt Angehörige und Freunde nicht.	He/She does not recognise relatives and friends.
16	Er/Sie erkennt Fernsehcharaktere oder andere Berühmtheiten nicht.	He/She does not recognise television characters or other famous people.
17	Er/Sie verläuft sich oder geht auf einem Weg oder Spaziergang in die falsche Richtung, den er/sie schon oft gegangen ist.	He/She gets lost or takes the wrong direction on a route or walk that he went on often.
18	Er/Sie erkennt Orte nicht, von denen ihm/ihr gesagt wurde, dass er/sie dort schon oft gewesen sei.	He/She does not recognise places he was told that he has often been to before.
19	Es fällt ihm/ihr schwer, im Fernsehen der Handlung zu folgen.	It is hard for him/her to follow the storyline when watching TV.

Table A3 ***(Continued)***

Actions		
20	Er/Sie vergisst regelmäßige Handlungen, die er/sie sonst ein- oder zweimal am Tag durchführen würde.	He/She forgets regular activities that he would normally do once or twice a day.
21	Er/Sie stellt fest, dass er/sie eine regelmäßige Handlung ausversehen zweimal durchgeführt hat.	He/She discovers that he did some regular activity twice by mistake.
22	Er/Sie muss überprüfen, ob er/sie alles getan hat, was er/sie tun sollte.	He/She has to check whether he has done everything he ought to.
23	Er/Sie vergisst, was er/sie gestern gemacht hat oder bringt die Details von dem durcheinander, was passiert ist.	He/She forgets what he did yesterday or getting the details of what happened mixed up and confused.
24	Er/Sie fängt an, Dinge zu tun und vergisst aber währenddessen, was er/sie eigentlich tun wollte. Dabei sagt er/sie möglicherweise etwas wie “Was tue ich hier?”	He/She starts doing something, but then forgets what he was intending to do. Thereby, he possibly says something like “What am I doing here?”
25	Er/Sie ist geistesabwesend. Er/Sie tut Dinge, die er/sie nicht wirklich vorhatte.	He/She is absentminded. He does things that he did not really intend to do.
Learning new things		
26	Er/Sie erinnert sich nicht an den Namen von jemandem, den er/sie vor kurzem zum ersten Mal getroffen hat.	He/She is not able to remember the name of someone he met for the first time recently.
27	Er/Sie erkennt Menschen nicht, die er/sie vor kurzem zum ersten Mal getroffen hat.	He/She does not recognise people he met for the first time recently.
28	Er/Sie verläuft sich auf einem Weg oder Spaziergang, den er/sie vorher nur ein- oder zweimal gegangen ist.	He/She gets lost on a route or walk that he has only gone on once or twice before.
29	Es gelingt ihm/ihr nicht, eine neue Fertigkeit, wie z.B. ein Spiel oder den Umgang mit einem Gerät, zu erlernen, wenn er/sie es ein- oder zweimal geübt hat.	He/She is not able to pick up a new skill, such as a game or handling a new gadget, if he practised it once or twice.
30	Er/Sie kann mit Veränderung im Tagesablauf nicht umgehen. Er/Sie verfolgt dann irrtümlicherweise weiterhin die alte Routine.	He/She is not able to cope with changes in his daily routine. He then mistakenly keeps following the former routine.
31	Er/Sie vergisst, sich an Verabredungen zu halten.	He/She forgets to stick to agreements.

Rating scales for questionnaire presentations	
Speech	
(4) In etwa 60% oder mehr Fällen pro Tag	(4) About 60 % or more of the cases in a day
(3) In weniger als 60% der Fälle pro Tag	(3) Less than 60 % of the cases in a day
(2) Etwa einmal am Tag	(2) About once each day
(1) Ein- oder zweimal in der Woche	(1) Once or twice in a week
(0) Seltener als einmal in der Woche oder nie	(0) Less than once a week
Faces and places and Actions	
(4) Mehrere Male am Tag	(4) Several times in a day
(3) Etwa einmal am Tag	(3) About once each day
(2) Ein- oder zweimal in der Woche	(2) Once or twice in a week
(1) Seltener als einmal in der Woche	(1) Less than once a week
(0) Nie	(0) Never

Table A3 ***(Continued)***

Learning new things	
(4) Jedes Mal	(4) On every occasion
(3) Häufiger	(3) On every other occasion
(2) Nur manchmal	(2) Only sometimes
(1) Sehr selten	(1) Very rarely
(0) Nie	(0) Never

Note. The questionnaire was developed based on Sunderland et al. (1983).

Table A4 ***Questions of the Motivation questionnaire in German and English***

Interest/enjoyment		
1	Wie gut hat Ihnen die Aufgabe heute gefallen? (überhaupt nicht gut; sehr gut)	How much did you enjoy the activity today? (not at all; a lot)
2	Wie viel Spaß hat Ihnen die Aufgabe gemacht? (überhaupt keinen Spaß; sehr viel Spaß)	How much fun was the activity to do? (not at all; a lot of fun)
3	Wie aufregend/spannend war die Aufgabe heute? (überhaupt nicht spannend; sehr spannend)	How exciting was the activity today? (not exciting at all; very exciting)
4	Wie gerne würden Sie die Aufgabe weiter üben, wenn wir Zeit dafür hätten? (überhaupt nicht gerne; sehr gerne)	How happily would you further practice the task if we had time?” (not gladly at all; very gladly)
Perceived competence		
5	Wie gut waren Sie heute in dieser Aufgabe? (überhaupt nicht gut; sehr gut)	How good were you at this activity today? (not good at all; very good)
6	Wie gut haben Sie heute in dieser Aufgabe abgeschnitten, im Vergleich zu anderen Tagen? (überhaupt nicht gut; sehr gut)	How well did you do at this activity today, compared to other days? (not well at all; very well)
7	Wie zufrieden sind Sie mit Ihrer Leistung heute? (überhaupt nicht zufrieden; sehr zufrieden)	How satisfied are you with your performance today? (not satisfied at all; very satisfied)
Effort/importance		
8	Wie sehr haben Sie sich heute angestrengt? (überhaupt nicht; sehr)	How much effort did you put into this today? (no effort at all; a lot of effort)
9	Wie viel Mühe haben Sie sich heute mit dieser Aufgabe gegeben? (überhaupt keine Mühe; sehr viel Mühe)	How hard did you try on this activity today? (not hard at all; very hard)
10	Wie wichtig war es Ihnen, gut in dieser Aufgabe zu sein? (überhaupt nicht wichtig; sehr wichtig)	How important was for you to do well on this task? (not important at all; very important)

Note. The questionnaire was developed based on Jaeggi et al. (2011) and McAuley et al. (1989).

7 Study 3: The methodological quality of short-term/working memory treatments in post-stroke aphasia: A systematic review[13]

Abstract

Purpose: The aim of this systematic review is to provide an overview of short-term memory (STM) and working memory (WM) treatments in stroke aphasia and to systematically evaluate the internal and external validity of these 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 8 targeted attention and WM usually involving more complex treatment procedures. 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 everyday life functioning were tested only in 5 studies.

Conclusions: Methodological limitations 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

---

[13] This chapter is under revision as:
Zakariás, L., Kelly, H., Salis, C., & Code, C. The methodological quality of short-term/working memory treatments in post-stroke aphasia: A systematic review. *Journal of Speech, Language, and Hearing Research.* https://pubs.asha.org/journal/jslhr

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, communication, and psychosocial functioning more generally to demonstrate the ecological validity of STM and WM treatments.

7.1 Introduction

Memory as well as language deficits (i.e., aphasia) are a frequent occurrence after stroke and both deficits interfere with inter-personal communication, rehabilitation as well as wellbeing (Barker-Collo et al., 2010; Cruice, Worrall, Hickson, & Murison, 2003; Worrall & Holland, 2003). Importantly, memory and language deficits seem to interact: For instance, the two related memory systems (short-term and working memory) have been shown to interfere with language processing after stroke. However, the introduction of experimental treatments to ameliorate such memory deficits in aphasiological research is a relatively new development. As a construct, short-term memory (STM) refers to the ability to temporarily maintain and retrieve information (Baddeley, 2012; Engle, Tuholski, Laughlin, & Conway, 1999). Relatedly, working memory (WM) refers to a complex cognitive construct that, beyond the temporary maintenance of information, also supports 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 describe the relationship between STM and WM (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 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).

STM and WM are associated with various brain regions, such as the frontal lobes, and in particular, the dorsolateral prefrontal cortex (D'Esposito et al., 1995; Miyake, Friedman, et al., 2000; E. E. Smith, Jonides, Marshuetz, & Koeppe, 1998), the left inferior frontal gyrus (Botvinick, Cohen, & Carter, 2004; E. E. 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 (N. Martin & Ayala, 2004; N. Martin & Reilly, 2012; Murray et al., 2018; Warrington & Shallice, 1969). Such deficits may persist even in cases where aphasia has resolved (e.g., Vallat et al., 2005). Moreover, impairments of STM/WM can negatively influence individuals' language comprehension (N. Martin et al., 2012; Novick et al., 2009; Robinson et al., 1998; Sung et al., 2009; Wright et al., 2007), reading (Caspari et al., 1998) and other aspects of verbal communication (Frankel et al., 2007; Fridriksson et al., 2006; Keil & Kaszniak, 2002; Luna, 2011; Penn et al., 2010; Ramsberger, 2005).

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 highlighted an overlap of language and STM/WM at cognitive and neural levels (e.g., Fedorenko et al., 2012; Geranmayeh et al., 2014). Second, studies of people with aphasia provided 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). Together, these findings galvanized a promising hypothesis, namely, that improvements in STM/WM would lead to improvements in language abilities that rely on STM/WM functions in aphasia. This hypothesis is particularly pertinent in rehabilitation research because it relates directly to the concept of generalization (Webster et al., 2015), or transfer of skills (Klingberg, Forssberg, & Westerberg, 2002), according to which treatment enhances not only the targeted skill but also performance on similar skills or even language abilities (e.g., sentence comprehension, naming).

The first generation of studies mainly used repetition tasks as treatment, relying heavily on both language and STM/WM skills. In the first reported STM treatment study, Peach (1987) delivered a treatment including 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 introduced, using various protocols and involving individuals with a range of types and severities of aphasia, as well as different cognitive-linguistic profiles (e.g., Eom & Sung, 2016; Harris et al., 2014; Kalinyak-Fliszar et al., 2011; Lee & Sohlberg, 2013; Vallat-Azouvi et al., 2014). Treatment effects on further domains of language, for instance, spoken sentence comprehension and verbal communication, have also been examined.

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 et al., 2017; Murray, 2012; Salis, Kelly, & Code, 2015), to our knowledge, systematic reviews of STM/WM treatment studies in stroke aphasia have not been reported previously. Crucially, such a systematic review could enhance and broaden the evidence-base of treatments in stroke aphasia and could provide guidelines to implement a key principle in evidence-based clinical practice, that is, the adoption of high quality studies (Greenhalgh, 2014). Importantly, such a review would also help resolve several major controversies in the STM/WM treatment literature that currently preclude identifying the best available evidence that could be adopted in clinical practice. 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 memory problems after stroke was an unmet rehabilitation need. Furthermore, there is growing acknowledgement that STM/WM deficits can influence rehabilitation decisions and outcomes (Balasooriya-Smeekens et al., 2016; Suleman & Kim, 2015), for example, returning to and staying in work (Balasooriya-Smeekens et al., 2016).

Additional questions arise, as to whether STM/WM treatments are indeed efficacious. In particular: Who (in relation to cognitive-linguistic profiles) might benefit from these treatments? Which treatment protocol and what dosage are most likely to produce improvements in STM/WM? More importantly, do STM/WM treatments also improve cognitive, language and other everyday functions that are not targeted during treatment, that is, do STM/WM treatments lead to transfer (near and far)[14] in aphasia? Addressing these questions would help better understand the underlying cognitive mechanisms of transfer effects of STM/WM treatments and, consequently, the nature of the relationship between these functions and the contribution of STM/WM to language and everyday functions.

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 STM/WM, language (e.g., spoken sentence comprehension, functional communication), and other everyday functions can benefit from STM/WM treatments in stroke aphasia.

7.2 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).

7.2.1 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

---

[14] Near transfer effects refer to the improvements on tasks that are similar to the treatment task, but were not practiced during treatment (e.g., new STM/WM tasks) (Jaeggi et al., 2008). Far transfer effects refer to the improvement on skills or abilities that were not targeted during treatment but depend on STM/WM (e.g., spoken sentence comprehension, verbal communication) (Jaeggi et al., 2008).

Register of Controlled Trials, PsycBITE™ and SpeechBITE™– in 2014 February by H.K., and then updated in 2016 December by two authors (H.K. and L.Z.).

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) short-term memory or working memory (and related terms), (iii) rehabilitation (and related terms). In addition, reference lists of included studies, conference abstracts (Clinical Aphasiology Conference) and three relevant reviews (Brady et al., 2016; Cotoi et al., 2016; Murray, 2012) were screened for potentially eligible studies. After removing duplicates, study titles and abstracts from the search were screened against eligibility 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.). Any disagreements were resolved through discussion with all authors, and if we opted for exclusion, we recorded the rationale for doing so.

Inclusion criteria 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; 3) intervention protocol included treatment of STM/WM; 4) outcomes included STM/WM data; 5) in case of mixed etiology groups, it was 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.

7.2.2 Data extraction

For each study meeting the eligibility criteria, the following data were extracted: 1) Study aims; 2) study method information (design, randomization, blinding patients/practitioners/assessors, inter-rater reliability of practitioners 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 practitioner, etc.); 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.) performed data extraction for each study.

7.2.3 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 studies with a single-case experimental design. It 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 in the outcome measures) are attributable to introduction of the independent variable (STM/WM treatment) and not some other factors (spontaneous recovery, charm of the practitioner, etc.). 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; Tate et al., 2015). Importantly, assessment of both types of validities largely depend 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/or L.Z.) independently rated each included study. In case any of the included studies were authored by any of the independent raters, they were not involved in rating their own study. Any disagreements were resolved through discussion with all authors.

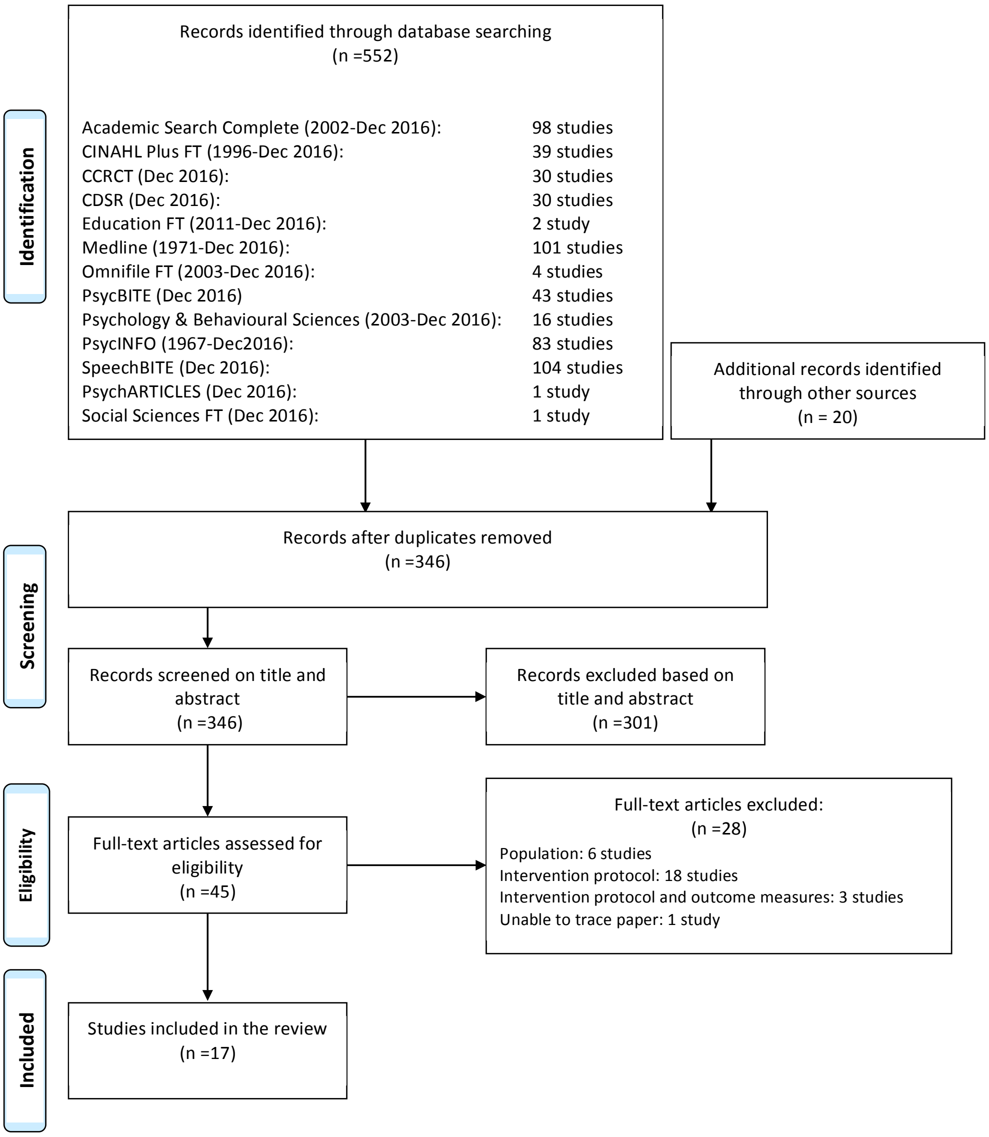

Figure 9 Flow diagram of the identification-inclusion process

7.3 Results

The electronic searches generated 552 studies. 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 were obviously irrelevant (n = 301) were excluded. The remaining 45 full text articles were assessed for eligibility by two authors (C.S., H.K., and/or L.Z.). After a 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 9. Three authors were contacted to obtain further details from their studies. One study was not available in the English language and was excluded. Studies included in the review are shown in Table 7.

7.3.1 Participants and treatments

Table 7 provides details of the participants' characteristics. A total of 37 people with post-stroke aphasia took part in the selected studies (24 male, 13 female). Aphasia type was specified in 14 of the 17 studies (82%). Eighteen of the 37 participants (49%) had some form of fluent aphasia (eight conduction, seven anomic, one Wernicke, and two broadly described as having fluent aphasia). Nine participants (24%) had non-fluent aphasia (three Broca's, three transcortical motor, and three noted as having non-fluent aphasia), seven (19%) were not reported, and three (8%) were reported to have recovered from aphasia by the time of study. Participants predominantly presented with mild or moderate severity of aphasia (14 and 12, respectively, and one presented with mild-moderate aphasia). Assessment data on severity was provided in 11 studies. Two studies reported severity without providing data on it. Only three participants had severe aphasia (8%), and seven (19%) were not reported or had largely recovered by the time of study. In terms of the participants' native language, six participants spoke English, 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 at a hospital in the US or UK, another two studies with two participants used test batteries in French). In terms of time post stroke, the majority of participants (27/37) was beyond eight months, representing chronic aphasia. Time post stroke for nine participants was not reported and one was 11 days post-onset at the time of study (Peach, 1987). Information about coexisting impairments of STM/WM and attention was provided in 13 studies involving 24 participants.

Table 8 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). The rest of the treatments (8/17) targeted attention and WM. Most of these treatments had a complex structure in that they involved more than one task with several different types practiced during a session (see Table 8).

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 8 indicates that the frequency of treatment administration varied greatly across studies (0.7-5 times per week, average frequency: 2.5 times per week), with the duration of treatment ranging from 1-19 months in total (mean of 23 weeks per study).

Table 7 ***Participant characteristics***

Study	N	Age (mean)	Education (mean)	Gender	Handedness	TPO (mean)	Aphasia type	Aphasia severity	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	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	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	1 pSTM impairment 1 sSMT impairment
Kalinyak-Fliszar et al. (2011)	1	55	High school diploma	F	R	29 m	Conduction	Moderate	Verbal STM impairment
Koenig-Bruhin and Studer-Eichenberger (2007)	1	47	University degree	M	R	34 m	Conduction	NR	Verbal STM impairment
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	1 Mild AOS 4 Impaired attention
Majerus et al. (2005)	1	50	NR	F	NR	NR	NR	NR	pSTM impairment
Mayer and Murray (2002)	1	62	University degree and some work towards doctorate	M	R	NR	Fluent	Mild	R-sided weakness; impaired STM, WM, and attention

Table 7 ***(Continued)***

Study	N	Age (mean)	Education (mean)	Gender	Handedness	TPO (mean)	Aphasia type	Aphasia severity	Reported coexisting impairments
Murray et al. (2006)	1	57	University degree	M	R	4 y	Conduction	Mild	Verbal WM deficit, mildly impaired attention
Peach (1987)	1	53	NR	F	NR	11 d	Conduction	Moderate	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	2 Attention impairment
Salis (2012)	1	73	University degree	F	L	5 y	TMA	Severe	R Hemiplegia, verbal and visuo-spatial STM impairment
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	3 Mild AOS 2 Moderate AOS 5 Verbal STM impairment 4 Visuo-spatial STM impairment
Vallat et al. (2005)	1	53	High school diploma	Male	R	14 m	Recovered from conduction	Recovered	WM deficit
Vallat-Azouvi et al. (2014)	1	38	NR	F	NR	> 4 y	Recovered from conduction	Recovered	WM deficit
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	NR

Note. TPO = time post onset; M = male; F = female; R = right; m = months; y = years; d = days; AOS = apraxia of speech; NR = not reported; WM = working memory; TMA = transcortical motor aphasia.

Table 8 ***Treatment characteristics***

Study	Design	Target cognitive construct	Treatment procedure	Schedule and treatment duration
Berthier et al. (2014)	Pre-post intervention	Treatment 1: Language skills such as naming, repetition, spoken word comprehension and conversation Treatment 2: Auditory-verbal STM	1. Distributed speech-language therapy (DSLT) combined with the cholinesterase inhibitor donepezil (DP) 2. Massed sentence repetition therapy (MSRT) combined with DP	DSLT: ~2.5hrs per week, for 16 weeks (40 hours) MSRT: 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 (40 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	Treatment 1: Phonological STM Treatment 2: Semantic STM	1. Repetition and recognition tasks with non-words 2. Repetition and recognition tasks with words	1×90-min therapy session per week + 20 home trials per week, for 10 weeks (for both treatments)
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
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

Table 8 *(Continued)*

Study	Design	Target cognitive construct	Treatment procedure	Schedule and treatment duration
Lee and Sohlberg (2013)	ABA 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	Treatment 1: Text level reading Treatment 2: Attention and WM	1. Repeated oral reading of text (Modified MOR, based on Beeson, 1998) 2. Sequence Exercises for Working Memory (modified reading span task with grammaticality judgments and spoken recall of semantic categories)	2hs/session, 11 sessions
Murray et al. (2006)	ABA 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
Peach (1987)	ABA 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 (1987)	ABA 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

Table 8 *(Continued)*

Study	Design	Target cognitive construct	Treatment procedure	Schedule and treatment duration
Salis (2012)	AB+follow-up with multiple baseline	Auditory-verbal STM	Matching listening span	2×30 min per week, for 13 weeks
Salis et al. (2017)	AB+follow-up with multiple baseline and probe task	Auditory-verbal STM	Matching listening span with feedback	~27-30 sessions
Vallat et al. (2005)	ABA 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
Vallat-Azouvi et al. (2014)	ABA multiple baseline across behaviors	Treatment 1: Phonological loop Treatment 2: Visuo-spatial sketchpad Treatment 3: Central executive	1. 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 2. 2-D mental imagery on a chessboard, 2-D mental imagery on a calculator keyboard, 3-D mental imagery, visual n-back 3. N-back with words, n-back questions, reading span tasks	1×2hs per week, for 19 months
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

7.3.2 Methodological quality appraisal of included studies: Internal and external validity

Table 9 provides the appraisal scores for the scientific quality of each study. The total quality score ranged from 4 to 17 (mean = 9.5) on a 0–30 scale, 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) on a 0–14 scale. 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 N^{-1} trial 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. 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 practitioners 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, however, they were not "blind" to study phase (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) on a 0-16 scale. 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, in most studies, a justification for the suitability of the statistical procedure or the results of the analyses were not reported. Six studies replicated their findings, with four studies involving four or more participants (Eom & Sung, 2016; Lee & Sohlberg, 2013; Peach, Nathan, & Beck, 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 9 *RoBiNT scores of included studies*

Study	Design with control	Randomization	Sampling of behavior	Blinding of people involved in the intervention	Blinding of assessor(s)	Inter-rater 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	Generalization	**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-Eichenberger (2007)	0	0	1	0	0	0	0	**1**	2	0	1	1	0	1	0	0	**5**	**6**
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**

Table 9 *(Continued)*

Study	Design with control	Randomization	Sampling of behavior	Blinding of people involved in the intervention	Blinding of assessor(s)	Inter-rater 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	Generalization	**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**

7.3.3 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 10 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. Among the auditory-verbal STM tasks, digit span forward was the most popular task (used in 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 (in the latter case requiring identity, semantic or rhyming judgments), comprising single words/non-words, word/non-word pairs or triplets, letter or digit strings. Treatment effects on WM outcome measures 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 with subtests that assess WM, the Test of Everyday Attention (TEA; Robertson, Ward, Ridgeway, & Nimmo-Smith, 2001) (3/17). In terms of response demands, the majority of studies (14/17) employed tests involving 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 requiring 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).

Effects of STM/WM treatment on language (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. 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 (Lee & Sohlberg, 2013; Murray et al., 2006). Language production, in particular verbal communication, was investigated only in six studies, using picture-description tasks (Berthier et al., 2014; Koenig-Bruhin & Studer-Eichenberg, 2007), or self- and/or spouse-reported questionnaires (Murray et al., 2006; Peach et al., 2017; Salis et al., 2017; Vallat et al., 2005). Change due to treatment in aphasia severity (Aphasia Quotient of the Western Aphasia Battery, WAB) was assessed in two studies (Berthier et al., 2014; Eom & Sung, 2016). Change in general language profile, assessed by the WAB (Kertész, 1982; 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).

With respect to the ecological validity of the studies, only five studies included measures of everyday life functioning. Note that treatment effects on everyday life functioning are also a form of far transfer effects according to the definition we adopted here. Of these, four studies used a verbal communication questionnaire (mentioned also in the previous paragraph on outcome measures of verbal communication). Two of these used a self-reported measure (Peach et al., 2017; Vallat et al., 2005), one study used a spouse-reported measure (Salis et al., 2017), and one study used both (Murray et al., 2006). Two studies used a self-reported WM questionnaire (Vallat et al., 2005; Vallat-Azouvi et al., 2014) and two others used an attention questionnaire (both self-reported and spouse-reported in Murray et al. (2006), and self-reported in Peach et al. (2017)). One study used the self-reported Communication Outcomes After Stroke (COAST; Long, Hesketh, Paszek, Booth, & Bowen, 2008) (Salis et al., 2017).

Table 10 ***Outcome measures and their results***

Study	N	Near transfer	Far transfer
Berthier et al. (2014)	3	**STM:** PALPA word^{*1} and non-word^{*3} repetition, digit span, word pair repetition^{*2,3}, word triplet repetition ^{*1,2,3}, sentence repetition^{*1,2,3}	**Aphasia severity:** WAB AQ^{*1,2,3} **Spoken discourse:** WAB picture description$^{\#}$
Eom and Sung (2016)	6	**STM/WM:** Digit span forward* and backward* (WAIS), word span forward* and backward (Sung, 2011), sentence repetition*	**Aphasia severity:** WAB AQ* **Language profile:** WAB fluency, comprehension, repetition*, and naming, BNT **Spoken sentence comprehension:** Sentence picture matching* (Sung, 2015)
Francis et al. (2003)	1	**STM/WM:** Digit span forward and backward* (WAIS-3), letter span, sentence repetition*, Recognition Memory Test Words* (Warrington, 1984)	**Spoken sentence comprehension:** RTT*, TROG*, 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^{*1,2}, word span^{*1,2}, semantic^{*1} and rhyming span^{*1,2}, sentence repetition (Hanten & Martin, 2000)^{*2} **General cognitive profile:** Birmingham Cognitive Screen (Humphreys et al., 2012)	**Spoken sentence comprehension:** Anomaly detection^{*1} (Hanten & Martin, 2000), PALPA spoken sentence picture matching^{*1}
Kalinyak-Fliszar et al. (2011)	1	**STM:** TALSA phoneme discrimination$^{\#}$, rhyme pair judgment$^{\#}$, synonym triplet judgment$^{\#}$, rhyming triplet judgment$^{\#}$, lexical comprehension, category judgment–pictures$^{\#}$, picture naming$^{\#}$, word and non-word repetition$^{\#}$, word pair repetition$^{\#}$, sentence repetition$^{\#}$, digit span–pointing, digit span–repetition, word and non-word span$^{\#}$, repetition span for words$^{\#}$, probe memory span	**Spoken sentence comprehension:** TALSA sentence comprehension$^{\#}$

Table 10 *(Continued)*

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

Table 10 *(Continued)*

Study	N	Near transfer	Far transfer
Peach (1987)	1	**STM:** Sentence repetition (probe task)$^{\#}$	**Language profile:** BDAE $^{\#\text{Naming, Repetition}}$
Peach et al. (2017)	4	**Attention and WM:** TEA map search$^{\#1}$, elevator counting, elevator counting with distraction$^{\#1, 3, 4}$, visual elevator accuracy$^{\#1}$ and timing$^{\#1}$, elevator counting with reversal$^{\#1}$, telephone search, telephone search with counting$^{\#1}$, lottery, Paced Auditory Serial Addition Test (PASAT)$^{\#1, 4}$ **Interference control:** Stroop Test$^{\#1, 4}$	**Aphasia severity:** WAB AQ$^{\#1, 4}$ **Language:** Discourse Comprehension Test, Object and Action Naming Battery$^{\#1}$ (Druks & Masterson, 2000) **Functional communication:** American Speech-Language Hearing Association Quality of Communication Life Scale **Cognitive functioning in everyday life:** Rating Scale of Attention Behaviors$^{\#1, 2}$
Salis (2012)	1	**STM/WM:** PALPA digit listening span$^{\#}$, digit span forward$^{\#}$ and backward–repetition (WMS-R)	**Spoken sentence comprehension:** TT, TROG* **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 **Functional communication:** CETI^{*3}, Communication Outcome After Stroke (COAST)
Vallat et al. (2005)	1	**STM/WM:** Digit span forward* and backward, visuo-spatial span forward and backward, letter span, word span, Brown-Peterson paradigm (consonant recall with or without interference)*, arithmetic problem solving task*	**Spoken sentence comprehension:** Text oral comprehension task (based on Ducarne de Ribeaucourt, 1989)* **Functional communication:** Verbal communication questionnaire* (Darringrand, 2000) **Cognitive functioning in everyday life:** Working memory questionnaire*

Table 10 *(Continued)*

Study	N	Near transfer	Far transfer
Vallat-Azouvi et al. (2014)	1	**Attention, STM/WM:** Digit span forward* and backward, visuo-spatial span forward and backward, Brown-Peterson paradigm (consonant recall with or without interference)*, *n*-back, TAP divided attention and mental flexibility, arithmetic problem solving	**Spoken sentence comprehension:** Text oral comprehension task (Ducarne de Ribeaucourt, 1989), TT **Cognitive functioning in everyday life:** 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$^{*1, 2}$, *n*-back with computer-generated tones	**Spoken sentence comprehension:** TROG$^{*1, 3}$ **Control task:** BNT

Note. * and # in superscripts indicate an improvement in the task based on statistics and descriptive analysis, respectively; numbers in superscripts refer to the participants; N = number of participants included in the study; PALPA = Psycholinguistic Assessments of Language Processing in Aphasia (Kay, Coltheart, Lesser, 1992); WAB = Western Aphasia Battery (Kertész, 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).

7.3.4 The effects of treatments

Table 10 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. The main characteristics of Phase 1 studies are that they 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)[15]. 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). Eighty-five and 82% of the studies that investigated effects on STM and WM, respectively, reported improvements in these domains. Similarly, 77% of the studies that included outcome measures on spoken sentence comprehension (7/9) reported improvements in these measures. Understanding of spoken paragraphs showed a small, subtle improvement in one study as measured by decreased 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

---

[15] 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.

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 verbal communication questionnaire, only one was successful in demonstrating improvements (Vallat et al., 2005). Results in measures of everyday life functioning (i.e., WM and attention questionnaires) were variable across studies. Two studies reported improvements in WM and attention (Vallat et al., 2005; Peach et al., 2017, respectively), whereas two studies did not report any changes in these domains (Murray et al., 2006; Vallat-Azouvi et al., 2014).

7.4 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 cognitive, language, and everyday functions (i.e., transfer effects) in 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.

7.4.1 Methodological quality

Among the studies reviewed, only 53% (9/17) used a single-case methodology with factors (e.g., multiple baseline measurements, inclusion of untreated control measures; Tate et al., 2015) that introduced experimental control and helped mitigate or isolate sources of experimental bias. A positive feature of such designs is that a systematic manipulation of the intervention and target behavior assessment allows for participants to serve as their own controls (Tate et al., 2015). Forty-one percent of the studies (7/17) described individuals, but rather than implementing single-case methodology, employed instead a pre-post intervention design. Pre-post intervention designs are 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 et al., 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 et al., 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 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 them (11/17) scored zero on the first three items (i.e., design, randomization, and sampling of behavior). Similarly, external validity scores, although higher than for internal validity, were still low. Such important methodological limitations make 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 treatments to date because of the major issues our review identified in the included studies. 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 in clinical practice.

7.4.2 Participants and treatments

With respect to the participants included in the reviewed studies, 73% 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 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. Participants' cognitive profile was taken into account in 13 studies when selecting 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, most frequently comprising words or sentences. The remaining treatments focused on attention and WM, and were usually complex in terms of their structure (e.g., APT-2, Sohlberg et al, 2010; L-SAT, Peach et al., 2017). 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 et al., 2014), one can assume 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 current data due to large differences in treatment outcome both across and within studies. This assumption should be specifically tested in future studies.

7.4.3 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 functional communication; improvements in this domain were reported in some but not all studies. Studies rarely included patient-reported outcome measures of everyday life and psychosocial functioning (e.g., questionnaires assessing everyday life problems related to deficits of STM/WM). When they did, results were inconclusive. As we noted before, due to low methodological validity of the reviewed studies, any conclusions regarding the positive effects of treatments have to be viewed with caution.

7.4.4 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) , 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 (Crawford & Garthwaite, 2005), and the Wilcoxon signed-rank test (Field, 2009) can be used in case-control designs and randomized control trials, respectively. 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 of everyday life and psychosocial functioning (e.g., verbal communication questionnaires) should be included in future studies to demonstrate clinically significant improvements. Finally, future attempts should include recommendations about a Core Outcome Set for use in STM/WM trials with aphasia, that is, the minimum set of outcomes that should be measured and reported in STM/WM trails. This would potentially enable comparison of equivalent research and use of this data in meta-analysis.

7.5 Appendix

Search terms

1. aphasi*
2. dysphasi*
3. 1 OR 2
4. Working memory
5. Short-term memory
6. Immediate memory
7. Acoustic memory
8. Echoic memory
9. Verbal short-term memory
10. Primary memory
11. Auditory-verbal short-term memory
12. STM
13. Sensory memory
14. Active memory
15. Transient memory
16. OR/4-15
17. Rehabilitat*
18. Therap*
19. Treat*
20. Train*
21. OR/17-20
22. 3 AND 16 AND 21

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

Alexander, M. P. (2006). Impairments of procedures for implementing complex language are due to disruption of frontal attention processes. *Journal of the International Neuropsychological Society*, *12*(2), 236–247. https://doi.org/10.1017/S1355617706060309

Allen, C. M., Martin, R. C., & Martin, N. (2012). Relations between short-term memory deficits, semantic processing, and executive function. *Aphasiology*, *26*(3–4), 428–461. https://doi.org/10.1080/02687038.2011.617436

Alloway, T. P., Gathercole, S. E., Adams, A.-M., Willis, C., Eaglen, R., & Lamont, E. (2005). Working memory and phonological awareness as predictors of progress towards early learning goals at school entry. *British Journal of Developmental Psychology*, *23*(3), 417–426. https://doi.org/10.1348/026151005X26804

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

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review*, *22*(2), 366–377. https://doi.org/10.3758/s13423-014-0699-x

Avons, S. E., Wragg, C. A., Cupples, W. L., & Lovegrove, W. J. (1998). Measures of phonological short-term memory and their relationship to vocabulary development. *Applied Psycholinguistics*, *19*(4), 583–601. https://doi.org/10.1017/S0142716400010377

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. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

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

Baddeley, A., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89. https://doi.org/10.1016/S0079-7421(08)60452-1

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

Barker-Collo, S., Feigin, V. L., Parag, V., Lawes, C. M. M., & Senior, H. (2010). Auckland Stroke Outcomes Study. Part 2: Cognition and functional outcomes 5 years poststroke. *Neurology*, *75*(18), 1608–1616. https://doi.org/10.1212/WNL.0b013e3181fb44c8

Barrouillet, P., & Camos, V. (2001). Developmental increase in working memory span: Resource sharing or temporal decay? *Journal of Memory and Language*, *45*(1), 1–20. https://doi.org/10.1006/jmla.2001.2767

Basso, A. (2003). *Aphasia and its therapy*. New York: Oxford University Press.

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

Ben-Shachar, M., Hendler, T., Kahn, I., Ben-Bashat, D., & Grodzinsky, Y. (2003). The neural reality of syntactic transformations: Evidence from functional magnetic resonance imaging. *Psychological Science*, *14*(5), 433–440. https://doi.org/10.1111/1467-9280.01459

Ben-Shachar, M., Palti, D., & Grodzinsky, Y. (2004). Neural correlates of syntactic movement: Converging evidence from two fMRI experiments. *NeuroImage*, *21*(4), 1320–1336. https://doi.org/10.1016/j.neuroimage.2003.11.027

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

Biegler, K. A., Crowther, J. E., & Martin, R. C. (2008). Consequences of an inhibition deficit for word production and comprehension: Evidence from the semantic blocking paradigm. *Cognitive Neuropsychology*, *25*(4), 493–527. https://doi.org/10.1080/02643290701862316

Bishop, D. V. M. (1989). *Bishop, Dorothy VM. Test for reception of grammar* (2nd ed.). University of Manchester: The Author, Age and Cognitive Performance Research Centre.

Blomert, L. (1992). The Amsterdam — Nijmegen Everyday Language Test (ANELT). In *Neuropsychological Rehabilitation* (pp. 121–127). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-77067-8_14

Blomert, L., Kean, M. L., Koster, C., & Schokker, J. (1994). Amsterdam—Nijmegen everyday language test: Construction, reliability and validity. *Aphasiology*, *8*(4), 381–407. https://doi.org/10.1080/02687039408248666

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

Boyle, W., Lindell, A. K., & Kidd, E. (2013). Investigating the role of verbal working memory in young children's sentence comprehension. *Language Learning*, *63*(2), 211–242. https://doi.org/10.1111/lang.12003

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

Brownsett, S. L. E., Warren, J. E., Geranmayeh, F., Woodhead, Z., Leech, R., & Wise, R. J. S. (2014). Cognitive control and its impact on recovery from aphasic stroke. *Brain*, *137*(1), 242–254. https://doi.org/10.1093/brain/awt289

Brunner, C., & Steiner, J. (1994). *ANELT Amsterdam-Nijmegen Everyday Language Test – Deutsche Fassung*. Developed by Blomert & Buslach.

Burchert, F., Lorenz, A., Schröder, A., De Bleser, R., & Stadie, N. (2011). *Sätze verstehen. Neurolinguistische Materialien für die Untersuchung von syntaktischen Störungen beim Satzverständnis*. Hofheim: NAT-Verlag.

Byng, S., & Black, M. (1999). *Reversible Sentence Comprehension Test*. London: Winslow. Retrieved from http://discovery.ucl.ac.uk/186014/

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

Caplan, D., & Waters, G. (2013). Memory mechanisms supporting syntactic comprehension. *Psychonomic Bulletin & Review*, *20*(2), 243–268. https://doi.org/10.3758/s13423-012-0369-9

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

Chein, J. M., & Morrison, A. B. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task.

Psychonomic Bulletin & Review, *17*(2), 193–199. https://doi.org/10.3758/PBR.17.2.193

Christensen, S. C., & Wright, H. H. (2010). Verbal and non-verbal working memory in aphasia: What three n-back tasks reveal. *Aphasiology*, *24*(6–8), 752–762. https://doi.org/10.1080/02687030903437690

Christensen, S. C., Wright, H. H., & Ratiu, I. (2018). Working memory in aphasia: Peeling the onion. *Journal of Neurolinguistics*. https://doi.org/10.1016/j.jneuroling.2018.02.001

Cicerone, K. D. (2002). Remediation of "working attention" in mild traumatic brain injury. *Brain Injury*, *16*(3), 185–195. https://doi.org/10.1080/02699050110103959

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

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, *386*(6625), 604–608. https://doi.org/10.1038/386604a0

Collette, F., Van der Linden, M., Laureys, S., Arigoni, F., Delfiore, G., Degueldre, C., … Salmon, E. (2007). Mapping the updating process: Common and specific brain activations across different versions of the running span task. *Cortex*, *43*(1), 146–158. https://doi.org/10.1016/S0010-9452(08)70452-0

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

Coolican, H. (2014). *Research Methods and Statistics in Psychology* (6th ed.). London: Psychology Press. Retrieved from https://www.taylorfrancis.com/books/9781444170122

Cotoi, A., Salter, K., Jinah, A., Brar, J., Macaluso, S., Orange, J., & Teasell, R. (2016). Aphasia and apraxia (Evidence tables). Retrieved from http://www.ebrsr.com/evidence-review/14-aphasia-and-apraxia

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, *169*, 323–338. https://doi.org/10.1016/S0079-6123(07)00020-9

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

Cruice, M., Worrall, L., Hickson, L., & Murison, R. (2003). Finding a focus for quality of life with aphasia: Social and emotional health, and psychological well-being. *Aphasiology*, *17*(4), 333–353. https://doi.org/10.1080/02687030244000707

Dahlin, E., Neely, A. S., Larsson, A., Bäckman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science*, *320*(5882), 1510–1512. https://doi.org/10.1126/science.1155466

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

de Jong, P. F., & Olson, R. K. (2004). Early predictors of letter knowledge. *Journal of Experimental Child Psychology*, *88*(3), 254–273. https://doi.org/10.1016/j.jecp.2004.03.007

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

del Río, D., Maestú, F., López-Higes, R., Moratti, S., Gutiérrez, R., Maestú, C., & del-Pozo, F. (2011). Conflict and cognitive control during sentence comprehension: Recruitment of a frontal network during the processing of Spanish object-first sentences. *Neuropsychologia*, *49*(3), 382–391. https://doi.org/10.1016/j.neuropsychologia.2010.12.005

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037//0033-295X.93.3.283

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

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: Changes in grey matter induced by training. *Nature*, *427*(6972), 311–312. https://doi.org/10.1038/427311a

Duffin, J. T., Collins, D. R., Coughlan, T., O'Neill, D., Roche, R. A. P., & Commins, S. (2012). Subtle memory and attentional deficits revealed in an Irish stroke patient sample using domain-specific cognitive tasks. *Journal of Clinical and Experimental Neuropsychology*, *34*(8), 864–875. https://doi.org/10.1080/13803395.2012.690368

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.

Fox-Boyer, A. V. (2013). *TROG-D: Test zur Überprüfung des Grammatikverständnisses*. Schulz-Kirchner Verlag GmbH.

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

Franklin, R. D., Allison, D. B., & Gorman, B. S. (2014). *Design and analysis of single-case research.* New York: Psychology Press.

Freedman, M., Alexander, M. P., & Naeser, M. A. (1984). Anatomic basis of transcortical motor aphasia. *Neurology*, *34*(4), 409–417.

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

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

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

Friedmann, N., & Gvion, A. (2003). Sentence comprehension and working memory limitation in aphasia: A dissociation between semantic-syntactic and phonological reactivation. *Brain & Language*, *86*(1), 23–39. https://doi.org/10.1016/S0093-934X(02)00530-8

Geranmayeh, F., Brownsett, S. L. E., & Wise, R. J. S. (2014). Task-induced brain activity in aphasic stroke patients: What is driving recovery? *Brain*, *137*(10), 2632–2648. https://doi.org/10.1093/brain/awu163

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

Goodglass, H. (1993). *Understanding aphasia*. San Diego: Academic Press.

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

Green, D. W., Grogan, A., Crinion, J., Ali, N., Sutton, C., & Price, C. J. (2010). Language control and parallel recovery of language in individuals with aphasia. *Aphasiology*, *24*(2), 188–209. https://doi.org/10.1080/02687030902958316

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

Guye, S., Röcke, C., Mérillat, S., von Bastian, C. C., & Martin, M. (2016). Adult lifespan. In T. Strobach & J. Karbach (Eds.), *Cognitive training: An overview of features and applications* (pp. 45–55). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-42662-4_5

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

Hallowell, B., & Chapey, R. (2008). Introduction to language intervention strategies in adult aphasia. In R. Chapey (Ed.), *Language intervention strategies in aphasia and related neurogenic communication disorders* (5th ed., pp. 3–19). Philadelphia, USA: Lippincott Williams & Wilkins.

Hamilton, A. C., & Martin, R. C. (2007). Proactive interference in a semantic short-term memory deficit: Role of semantic and phonological relatedness. *Cortex*, *43*(1), 112–123. https://doi.org/10.1016/S0010-9452(08)70449-0

Harnish, S. M., Schwen Blackett, D., Zezinka, A., Lundine, J. P., & Pan, X. (2018). Influence of working memory on stimulus generalization in anomia treatment: A pilot study. *Journal of Neurolinguistics*. https://doi.org/10.1016/j.jneuroling.2018.02.003

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.

Helm-Estabrooks, N., & Albert, M. L. (1991). *Manual of aphasia therapy*. Austin, TX: Pro-Ed.

Hockey, A., & Geffen, G. (2004). The concurrent validity and test–retest reliability of a visuospatial working memory task. *Intelligence*, *32*(6), 591–605. https://doi.org/10.1016/j.intell.2004.07.009

Holmes, J., Gathercole, S. E., & Dunning, D. E. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*(4), F9–F15. https://doi.org/10.1111/j.1467-7687.2009.00848.x

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

Hsu, N. S., Jaeggi, S. M., & Novick, J. M. (2017). A common neural hub resolves syntactic and non-syntactic conflict through cooperation with task-specific networks. *Brain & Language*, *166*, 63–77. https://doi.org/10.1016/j.bandl.2016.12.006

Huber, W. (1983). *Aachener Aphasie-Test (AAT)*. Hogrefe, Verlag für Psychologie.

Hussey, E. K., Harbison, J. I., Teubner-Rhodes, S. E., Mishler, A., Velnoskey, K., & Novick, J. M. (2017). Memory and language improvements following cognitive control training. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(1), 23–58. https://doi.org/10.1037/xlm0000283

Hussey, E. K., & Novick, J. M. (2012). The benefits of executive control training and the implications for language processing. *Frontiers in Psychology*, *3*, 1–14. https://doi.org/10.3389/fpsyg.2012.00158

Ivanova, M. V., Dragoy, O. V., Kuptsova, S. V., Ulicheva, A. S., & Laurinavichyute, A. K. (2015). The contribution of working memory to language comprehension: Differential effect of aphasia type. *Aphasiology*, *29*(6), 645–664.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829–6833. https://doi.org/10.1073/pnas.0801268105

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences*, *108*(25), 10081–10086.

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

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149. https://doi.org/10.1037/0033-295X.99.1.122

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

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the n-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 615–622. https://doi.org/10.1037/0278-7393.33.3.615

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217. https://doi.org/10.1037/0096-3445.133.2.189

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *Boston Naming Test (Revised)*. Philadelphia: Lea & Febiger.

Karbach, J., & Verhaeghen, P. (2014). Making working memory work: A meta-analysis of executive-control and working memory training in older adults. *Psychological Science*, *25*(11), 2027–2037. https://doi.org/10.1177/0956797614548725

Katz, B., Jaeggi, S., Buschkuehl, M., Stegman, A., & Shah, P. (2014). Differential effect of motivational features on training improvements in school-based cognitive training. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00242

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

Kelly, M. E., Loughrey, D., Lawlor, B. A., Robertson, I. H., Walsh, C., & Brennan, S. (2014). The impact of cognitive training and mental stimulation on cognitive and everyday functioning of healthy older adults: A systematic review and meta-analysis. *Ageing Research Reviews*, *15*, 28–43. https://doi.org/10.1016/j.arr.2014.02.004

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.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352–358.

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

Kühn, S., & Lindenberger, U. (2016). Research on human plasticity in adulthood: A lifespan agenda. In K. W. Schaie & S. Willis (Eds.), *Handbook of the psychology of aging* (8th ed., pp. 105–123). San Diego: Academic Press. https://doi.org/10.1016/B978-0-12-411469-2.00006-6

Lambon Ralph, M. A., Graham, K. S., Ellis, A. W., & Hodges, J. R. (1998). Naming in semantic dementia—what matters? *Neuropsychologia*, *36*(8), 775–784. https://doi.org/10.1016/S0028-3932(97)00169-3

Lambon Ralph, M. A., Snell, C., Fillingham, J. K., Conroy, P., & Sage, K. (2010). Predicting the outcome of anomia therapy for people with aphasia post CVA: Both language and cognitive status are key predictors. *Neuropsychological Rehabilitation*, *20*(2), 289–305. https://doi.org/10.1080/09602010903237875

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)

Li, S.-C., Schmiedek, F., Huxhold, O., Röcke, C., Smith, J., & Lindenberger, U. (2008). Working memory plasticity in old age: Practice gain, transfer, and maintenance. *Psychology and Aging*, *23*(4), 731–742. https://doi.org/10.1037/a0014343

Lilienthal, L., Tamez, E., Shelton, J. T., Myerson, J., & Hale, S. (2013). Dual n-back training increases the capacity of the focus of attention. *Psychonomic Bulletin & Review*, *20*(1), 135–141. https://doi.org/10.3758/s13423-012-0335-6

Lind, M., Kristoffersen, K. E., Moen, I., & Simonsen, H. G. (2009). Semi-spontaneous oral text production: Measurements in clinical practice. *Clinical Linguistics & Phonetics*, *23*(12), 872–886. https://doi.org/10.3109/02699200903040051

Lindeløv, J. K., Dall, J. O., Kristensen, C. D., Aagesen, M. H., Olsen, S. A., Snuggerud, T. R., & Sikorska, A. (2016). Training and transfer effects of N-back training for brain-injured and healthy subjects. *Neuropsychological Rehabilitation*, *26*(5–6), 895–909. https://doi.org/10.1080/09602011.2016.1141692

Lindenberger, U., Wenger, E., & Lövdén, M. (2017). Towards a stronger science of human plasticity. *Nature Reviews Neuroscience*, 261–262. https://doi.org/10.1038/nrn.2017.44

Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory & Cognition*, *22*(4), 395–410. https://doi.org/10.3758/BF03200866

Lomas, J., Pickard, L., Bester, S., Elbard, H., Finlayson, A., & Zoghaib, C. (1989). The Communicative Effectiveness Index: Development and psychometric evaluation of a functional communication measure for adult aphasia. *Journal of Speech and Hearing Disorders*, *54*(1), 113–124. https://doi.org/10.1044/jshd.5401.113

Long, A., Hesketh, A., Paszek, G., Booth, M., & Bowen, A. (2008). Development of a reliable self-report outcome measure for pragmatic trials of communication therapy following stroke: The Communication Outcome after Stroke (COAST) scale. *Clinical Rehabilitation*, *22*(12), 1083–1094. https://doi.org/10.1177/0269215508090091

Loosli, S. V., Buschkuehl, M., Perrig, W. J., & Jaeggi, S. M. (2012). Working memory training improves reading processes in typically developing children. *Child Neuropsychology*, *18*(1), 62–78. https://doi.org/10.1080/09297049.2011.575772

Lorenz, A., Heide, J., & Burchert, F. (2014). Compound naming in aphasia: Effects of complexity, part of speech, and semantic transparency. *Language, Cognition and Neuroscience*, *29*(1), 88–106. https://doi.org/10.1080/01690965.2013.766357

Lukács, Á., Győri, M., & Rózsa, S. (2012). A TROG pszichometriai jellemzőinek magyar vizsgálata, a normák kialakítása. In D. V. M. Bishop (Ed.), *TROG - Test for Reception of Grammar Handbook* (pp. 47–86). Budapest, Hungary: OS Hungary.

Lukic, S., Barbieri, E., Wang, X., Caplan, D., Kiran, S., Rapp, B., … Thompson, C. K. (2017). Right hemisphere grey matter volume and language functions in stroke aphasia. *Neural Plasticity*, 1–15. https://doi.org/10.1155/2017/5601509

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/

Maas, E., Robin, D. A., Hula, S. N. A., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology*, *17*(3), 277–298. https://doi.org/10.1044/1058-0360(2008/025)

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

Majerus, S., Attout, L., Artielle, M.-A., & Van der Kaa, M.-A. (2015). The heterogeneity of verbal short-term memory impairment in aphasia. *Neuropsychologia*, *77*, 165–176. https://doi.org/10.1016/j.neuropsychologia.2015.08.010

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

Majerus, S., Poncelet, M., Greffe, C., & Van der Linden, M. (2006). Relations between vocabulary development and verbal short-term memory: The relative importance of short-term memory for serial order and item information. *Journal of Experimental Child Psychology*, *93*(2), 95–119. https://doi.org/10.1016/j.jecp.2005.07.005

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLOS ONE*, *7*(8), e43230. https://doi.org/10.1371/journal.pone.0043230

Mårtensson, J., Eriksson, J., Bodammer, N. C., Lindgren, M., Johansson, M., Nyberg, L., & Lövdén, M. (2012). Growth of language-related brain areas after foreign language learning. *NeuroImage*, *63*(1), 240–244. https://doi.org/10.1016/j.neuroimage.2012.06.043

Martin, N., & Ayala, J. (2004). Measurements of auditory-verbal STM span in aphasia: Effects of item, task, and lexical impairment. *Brain & Language*, *89*(3), 464–483. https://doi.org/10.1016/j.bandl.2003.12.004

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. (1993). Short-term memory and sentence processing: Evidence from neuropsychology. *Memory & Cognition*, *21*(2), 176–183. https://doi.org/10.3758/BF03202730

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, R. C., Shelton, J. R., & Yaffee, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language*, *33*(1), 83–111. https://doi.org/10.1006/jmla.1994.1005

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., Mitchinson, S. I., & Murray, L. L. (2017). Addressing concomitant executive dysfunction and aphasia: Previous approaches and the new brain budget protocol. *Aphasiology*, *31*(7), 837–860. https://doi.org/10.1080/02687038.2016.1249333

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

Mayer, J. F., & Murray, L. L. (2012). Measuring working memory deficits in aphasia. *Journal of Communication Disorders*, *45*(5), 325–339. https://doi.org/10.1016/j.jcomdis.2012.06.002

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport*, *60*(1), 48–58. https://doi.org/10.1080/02701367.1989.10607413

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 817–835. https://doi.org/10.1037//0278-7393.27.3.817

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

McNeil, M. R., & Prescott, T. E. (1978). *The Revised Token Test*. Baltimore, MD: University Park Press.

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., Emerson, M. J., & Friedman, N. P. (2000). Assessment of executive functions in clinical settings: Problems and recommendations. *Seminars in Speech and Language*, *21*(2), 169–183. https://doi.org/10.1055/s-2000-7563

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.

Morgan, D. L., & Morgan, R. K. (2008). *Single-case research methods for the behavioral and health sciences*. Los Angeles: SAGE Publications.

Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, *18*(1), 46–60. https://doi.org/10.3758/s13423-010-0034-0

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

Murray, L. L., Salis, C., Martin, N., & Dralle, J. (2018). The use of standardised short-term and working memory tests in aphasia research: A systematic review. *Neuropsychological Rehabilitation*, *28*(3), 309–351. https://doi.org/10.1080/09602011.2016.1174718

Nicholas, M., Hunsaker, E., & Guarino, A. J. (2017). The relation between language, non-verbal cognition and quality of life in people with aphasia. *Aphasiology*, *31*(6), 688–702. https://doi.org/10.1080/02687038.2015.1076927

Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology*, *16*(10–11), 935–979. https://doi.org/10.1080/02687030244000563

Nickels, L., Howard, D., & Best, W. (1997). Fractionating the articulatory loop: Dissociations and associations in phonological recoding in aphasia. *Brain and Language*, *56*(2), 161–182. https://doi.org/10.1006/brln.1997.1732

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2014). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience*, *29*(2), 186–217. https://doi.org/10.1080/01690965.2012.758297

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

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(3), 263–281. https://doi.org/10.3758/CABN.5.3.263

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass*, *4*(10), 906–924. https://doi.org/10.1111/j.1749-818X.2010.00244.x

Oberauer, K., & Lewandowsky, S. (2011). Modeling working memory: A computational implementation of the Time-Based Resource-Sharing theory. *Psychonomic Bulletin & Review*, *18*(1), 10–45. https://doi.org/10.3758/s13423-010-0020-6

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, *19*(5), 779–819. https://doi.org/10.3758/s13423-012-0272-4

Osmánné Sági, J. (1991). Az afázia klasszifikációja és diagnosztikája. *Ideggyógyászati Szemle*, *44*, 339–362.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 46–59. https://doi.org/10.1002/hbm.20131

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.

Papagno, C., Cecchetto, C., Reati, F., & Bello, L. (2007). Processing of syntactically complex sentences relies on verbal short-term memory: Evidence from a short-term memory patient. *Cognitive Neuropsychology*, *24*(3), 292–311. https://doi.org/10.1080/02643290701211928

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

Pollack, I., Johnson, L. B., & Knaff, P. R. (1959). Running memory span. *Journal of Experimental Psychology*, *57*(3), 137–146. https://doi.org/10.1037/h0046137

Purdy, M. (2002). Executive function ability in persons with aphasia. *Aphasiology*, *16*(4–6), 549–557. https://doi.org/10.1080/02687030244000176

Racsmány, M., Lukács, Á., Németh, D., & Pléh, C. (2005). A verbális munkamemória magyar nyelvű vizsgálóeljárásai. *Magyar Pszichológiai Szemle*, *60*(4), 479–506. https://doi.org/10.1556/MPSzle.60.2005.4.3

Ramsberger, G. (2000). Executive functions: What are they and why do they matter? [Preface]. *Seminars in Speech and Language*, *21*(2), 93.

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

Raven, J. C., Raven, J. C., & Court, J. H. (1962). *Coloured Progressive Matrices*. London: HK: Lewis.

Redick, T. S., & Lindsey, D. R. B. (2013). Complex span and n-back measures of working memory: A meta-analysis. *Psychonomic Bulletin & Review*, *20*(6), 1102–1113. https://doi.org/10.3758/s13423-013-0453-9

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (2001). The Test of Everyday Attention (TEA), *4*(1), 5.

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

Rönnberg, J., Larsson, C., Fogelsjöö, A., Nilsson, L. G., Lindberg, M., & Angquist, K. A. (1996). Memory dysfunction in mild aphasics. *Scandinavian Journal of Psychology*, *37*(1), 46–61.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*(2), 217–236. https://doi.org/10.1068/p5117

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.

Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, *54*(2), 199–227. https://doi.org/10.1016/j.jml.2005.10.002

Schweizer, S., Hampshire, A., & Dalgleish, T. (2011). Extending brain-training to the affective domain: Increasing cognitive and affective executive control through emotional working memory training. *PLOS ONE*, *6*(9), e24372. https://doi.org/10.1371/journal.pone.0024372

Sharp, D. J., Turkheimer, F. E., Bose, S. K., Scott, S. K., & Wise, R. J. S. (2010). Increased frontoparietal integration after stroke and cognitive recovery. *Annals of Neurology*, *68*(5), 753–756. https://doi.org/10.1002/ana.21866

Shipstead, Z., Redick, T., & Engle, R. W. (2010). Does working memory training generalize? *Psychologica Belgica*, *50*(3–4). https://doi.org/10.5334/pb-50-3-4-245

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*(4), 628–654. https://doi.org/10.1037/a0027473

Sinotte, M. P., & Coelho, C. A. (2007). Attention training for reading impairment in mild aphasia: A follow-up study. *NeuroRehabilitation*, *22*(4), 303–310.

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.

Smith, G. E., Housen, P., Yaffe, K., Ruff, R., Kennison, R. F., Mahncke, H. W., & Zelinski, E. M. (2009). A cognitive training program based on principles of brain plasticity: Results from the Improvement in Memory with Plasticity-based Adaptive Cognitive Training (IMPACT) study. *Journal of the American Geriatrics Society*, *57*(4), 594–603. https://doi.org/10.1111/j.1532-5415.2008.02167.x

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

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.

Stadie, N., Cholewa, J., & De Bleser, R. (2013). *LEMO 2.0: Lexikon modellorientiert: Diagnostik für Aphasie, Dyslexie und Dysgraphie*. NAT-Verlag.

Stewart, F. M., Sunderland, A., & Sluman, S. M. (1996). The nature and prevalence of memory disorder late after stroke. *British Journal of Clinical Psychology*, *35*(3), 369–379. https://doi.org/10.1111/j.2044-8260.1996.tb01191.x

Stothard, S. E., & Hulme, C. (1992). Reading comprehension difficulties in children. *Reading and Writing*, *4*(3), 245–256. https://doi.org/10.1007/BF01027150

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

Sunderland, A., Harris, J. E., & Baddeley, A. D. (1983). Do laboratory tests predict everyday memory? A neuropsychological study. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 341–357. https://doi.org/10.1016/S0022-5371(83)90229-3

Sung, J. E. (2015). Effects of syntactic structure on sentence comprehension ability as a function of the canonicity of word-order and its relation to working memory capacity in Korean-speaking elderly adults. *Communication Sciences & Disorders*, *20*(1), 24–33. https://doi.org/10.12963/csd.15229

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

Szöllősi I., Lukács Á., & Zakariás L. (2015). A végrehajtó funkciók zavara afáziában. *Magyar Pszichológiai Szemle*, *70*(2), 349–369. https://doi.org/10.1556/0016.2015.70.2.4

Tan, Y., & Martin, R. C. (2018). Verbal short-term memory capacities and executive function in semantic and syntactic interference resolution during sentence comprehension: Evidence from aphasia. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2018.03.001

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

Thompson-Schill, S. L., Bedny, M., & Goldberg, R. F. (2005). The frontal lobes and the regulation of mental activity. *Current Opinion in Neurobiology*, *15*(2), 219–224. https://doi.org/10.1016/j.conb.2005.03.006

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

Tompkins, C. A., Bloise, C. G. R., Timko, M. L., & Baumgaertner, A. (1994). Working memory and inference revision in brain-damaged and normally aging adults. *Journal of Speech, Language, and Hearing Research*, *37*(4), 896–912. https://doi.org/10.1044/jshr.3704.896

Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language*, *49*(4), 446–468. https://doi.org/10.1016/S0749-596X(03)00095-0

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. (2012). The Working Memory Questionnaire: A scale to assess everyday life problems related to deficits of working memory in brain injured patients. *Neuropsychological Rehabilitation*, *22*(4), 634–649. https://doi.org/10.1080/09602011.2012.681110

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

Van der Molen, M. J., Van Luit, J. E. H., Van der Molen, M. W., Klugkist, I., & Jongmans, M. J. (2010). Effectiveness of a computerised working memory training in adolescents with mild to borderline intellectual disabilities. *Journal of Intellectual Disability Research*, *54*(5), 433–447. https://doi.org/10.1111/j.1365-2788.2010.01285.x

Villard, S., Noelle. (2016). *Between-session and within-session intra-individual variability in attention in aphasia*. Boston University.

Vuong, L. C., & Martin, R. C. (2011). LIFG-based attentional control and the resolution of lexical ambiguities in sentence context. *Brain & Language*, *116*(1), 22–32. https://doi.org/10.1016/j.bandl.2010.09.012

Vuong, L. C., & Martin, R. C. (2015). The role of LIFG-based executive control in sentence comprehension. *Cognitive Neuropsychology*, *32*(5), 243–265. https://doi.org/10.1080/02643294.2015.1057558

Waris, O., Soveri, A., & Laine, M. (2015). Transfer after working memory updating training. *PLOS ONE*, *10*(9), e0138734. https://doi.org/10.1371/journal.pone.0138734

Warrington, E. K., & Shallice, T. (1969). The selective impairment of auditory verbal short-term memory. *Brain*, *92*(4), 885–896. https://doi.org/10.1093/brain/92.4.885

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

Wenger, E., Kühn, S., Verrel, J., Mårtensson, J., Bodammer, N. C., Lindenberger, U., & Lövdén, M. (2017). Repeated structural imaging reveals nonlinear progression of experience-dependent volume changes in human motor cortex. *Cerebral Cortex*, *27*(5), 2911–2925. https://doi.org/10.1093/cercor/bhw141

Wilson, B., Cockburn, J., Baddeley, A., & Hiorns, R. (1989). The development and validation of a test battery for detecting and monitoring everyday memory problems. *Journal of Clinical and Experimental Neuropsychology*, *11*(6), 855–870. https://doi.org/10.1080/01688638908400940

Worrall, L., & Holland, A. (2003). Editorial: Quality of life in aphasia. *Aphasiology*, *17*(4), 329–332. https://doi.org/10.1080/02687030244000699

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

Wright, H. H., & Fergadiotis, G. (2012). Conceptualising and measuring working memory and its relationship to aphasia. *Aphasiology*, *26*(3–4), 258–278. https://doi.org/10.1080/02687038.2011.604304

Xing, S., Lacey, E. H., Skipper-Kallal, L. M., Jiang, X., Harris-Love, M. L., Zeng, J., & Turkeltaub, P. E. (2016). Right hemisphere grey matter structure and language outcomes in chronic left hemisphere stroke. *Brain: A Journal Of Neurology*, *139*(Pt 1), 227–241. https://doi.org/10.1093/brain/awv323

Ye, Z., & Zhou, X. (2009). Executive control in language processing. *Neuroscience & Biobehavioral Reviews*, *33*(8), 1168–1177. https://doi.org/10.1016/j.neubiorev.2009.03.003

Zakariás, L., Keresztes, A., Demeter, G., & Lukács, Á. (2013). A specific pattern of executive dysfunctions in transcortical motor aphasia. *Aphasiology*, *27*(12), 1426–1439. https://doi.org/10.1080/02687038.2013.835783

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

DECLARATION

I, Lilla Zakariás hereby declare that the doctoral thesis was written independently, without the help of third parties, and that the work presented in the thesis fully complies with all the rules of good scientific standards.

Potsdam, ______________ ________________________

Lilla Zakariás