Åbo Akademi University

Daniel Fellman

Verbal Working Memory Training in Health and Disease: Outcomes and Underlying Mechanisms

Verbal Working Memory Training in Health and Disease: Outcomes and Underlying Mechanisms

Daniel Fellman

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I will start by thanking my supervisors Professor Matti Laine and Dr. Anna Soveri. Matti, I am deeply grateful that you gave me the chance to conduct research in cognitive psychology. Your excellent guidance and ability to inspire me with new challenges has made my time as a PhD student very joyful. Furthermore, I would like to say that our collaboration in Study IV was a prime example of teamwork at its best. I could not have wished for a better supervisor. Anna, I am deeply thankful for your guidance and supervision over the years, I have been fortunate to work with such a skilled researcher as you. I am also truly indebted to you for your patience with me and for always finding time to listen to my research-related concerns.

I would like to thank my pre-examiners, Professor Elisabeth Stine-Morrow and Dr. Jason M. Chein. Your invaluable comments were of great help in improving this thesis.

I would also like to express my sincerest gratitude to all the members in the BrainTrain project, it has been an honor to be part of this tremendous team of top-level researchers. I am particularly thankful to Dr. Otto Waris for his collaboration in Studies II and IV, and Dr. Juha Salmitaival for his collaboration in Study III. I would also like to thank Dr. Jussi Jylkkä for helping me with statistics and for our many excellent discussions. I would also like to thank Liisa Ritakallio for the exceptional assistance during the demanding data collections in Studies III-IV. My deepest gratitude also goes to Daniel Wärnå; without your programming work, none of the studies conducted in this thesis would have been possible. I would also like to thank all those individuals who took part as participants in the studies.

I wish to thank all my fellow PhD students at the department of Psychology at Åbo Akademi university: Alar Kaskikallio, Alessandro Tadei, Carina Saarela, Francesco Pompedda, Jussi Jylkkä, Karolina Lukasik, Linda Karlsson, Marianne Källström, Minja Westerlund, and Thomas Nyman. I would especially like to thank Thomas Nyman. You have been extremely supportive since day 1, and our friendship and interesting discussions (spanning from research methodology to life in general) over the years has been invaluable to me.

I would like to thank my friends Christian Ahonen, David Nyman, Denis Snellman, Jonas Grönholm, Kristoffer Knuts, Linus Korkea-Aho, Martin Kjellman,

Oscar Damén, Rahul Raj, and Robin Julin. It is a privilege to share friendship with you.

I would like to thank my brother Walter Fellman and my sister Jennifer Hellman. I feel fortunate to be a big brother to such amazing siblings. I am also deeply thankful for all the support over the years from my mother Camilla Fellman, my father Porfirio Otero and my grandmother Margareta Fellman.

Lastly, I would like to thank the person to whom I am dedicating this thesis, my grandfather Håkan Fellman. Håkan, no part of this thesis would have been possible without your support and guidance over the years. Not only do you believe in me when I do not, support me when I need it most, but you are also the one who has taught me what is important in life.

LIST OF ORIGINAL PUBLICATIONS

I.   Fellman, D., Soveri A., Viktorsson, C., Haga, S., Nylund, J., Johansson, S., Edman, J., von Renteln., F., & Laine, M. (2018). Selective updating of sentences: Introducing a new measure of verbal working memory. *Applied Psycholinguistics, 39*, 275-301.
      doi:10.1017/S0142716417000182

II.   Fellman, D., Soveri, A., Waris, O., & Laine, M. (2017). Training of verbal working memory at sentence level fails to show transfer. *Frontiers in Communication, 2*, 14.
      doi: 10.3389/fcomm.2017.00014

III. Fellman, D., Salmi, J., Ritakallio, L., Ellfolk, U., Rinne. J.O., & Laine, M. (2018). Training working memory updating in Parkinson's disease: A randomized controlled trial. *Neuropsychological Rehabilitation*.
      doi:10.1080/09602011.2018.1489860 [Epub ahead of print]

IV.   Laine, M., Fellman, D., Waris, O., & Nyman, T. J. (2018). The early effects of external and internal strategies on working memory updating training. *Scientific Reports, 8*, 4045.
      doi:10.1038/s41598-018-22396-5

The original articles have been reproduced by permission of the copyright holders.

SVENSK SAMMANFATTNING

Arbetsminnesträning och dess möjliga generalisering (s.k. transfer) till andra kognitiva domäner har väckt stort forskningsintresse. Drygt ett decennium sedan utkom de första högintressanta resultaten som tydde på att en ny typ av kognitiv träning av arbetsminnet, s.k. *basträning* (eng. core training), har vidgående effekter: inte bara arbetsminnet blev bättre utan träningen förbättrade även andra komplexa kognitiva funktioner såsom prestationer på intelligenstest som har en stark koppling till arbetsminnet. Den repetitiv-adaptiva *basträningen* bygger på hypotesen att arbetsminnet är ett system med begränsad kapacitet som genom intensiv och utmanande träning kan utvidgas och därigenom frambringa förbättringar i andra kognitiva uppgifter där arbetsminnet spelar en viktig roll (även kallat *kapacitetshypotesen*). De underliggande mekanismerna i *basträning* är dock oklara och resultaten i arbetsminnesträningsstudierna som efterföljde de tidiga fynden påvisar en stor variation beträffande hur långt transfereffekterna sträcker sig. Dessutom tyder de senaste metaanalyserna på att de största effekterna av arbetsminnesträningen är begränsade till otränade varianter av de tränade uppgifterna. Dessa fynd talar för en alternativ hypotes som antar att träningsrelaterade förbättringar snarare leder till en utökad effektivitet av arbetsminnet som ett resultat av att man utvecklar effektiva strategier under träningsperioden, än en förändring i kapaciteten som sådan (även kallat *strategiförmedlingshypotesen*).

Det övergripande målet med denna avhandling var att undersöka effekterna av träning av verbalt arbetsminne på grund av verbala arbetsminnets avgörande roll i mänsklig kognition inklusive kommunikation och språkanvändning, samt att testa strategiförmedlingshypotesen som den underliggande mekanismen till träningseffekterna.

Studie I var ett metodologiskt preludium vars mål var att utveckla ett nytt verbalt test på meningsnivå som kräver uppdatering i arbetsminnet. Testet kallas Selective Updating of Sentences (SUS). SUS-testet förväntades erbjuda ett alternativ till de nuvarande verbala arbetsminnestesterna som innehåller mera artificiella stimulussekvenser. I SUS-testet måste deltagarna ständigt uppdatera det semantiska innehållet i en mening genom att ersätta orden i meningarna med nya innehållsord, vilket kan likna vissa verkliga kommunikativa situationer. Resultaten ur två experiment visade att SUS-testet innehade adekvata psykometriska egenskaper och att det korrelerade positivt med test som mäter verbalt episodiskt minne på menings- och paragrafnivå, vilket indikerar att SUS-testet också är relaterat till minne för mera vardagligt språkmaterial.

Studie II-III undersökte effekterna av hembaserad *basträning av arbetsminnet* med fokus på nära transfer (dvs. träningsrelaterad förbättring i uppgifter inom arbetsminnesdomänen) hos friska vuxna (**Studie II)** och hos patienter med Parkinsons sjukdom (**Studie III**). Båda studierna gav stöd för träningsrelaterade förbättringar i en del av de tränade uppgifterna. Uppgiftsspecifik nära transfer till otränade arbetsminnesuppgifter med samma uppgiftsparadigm som den tränade uppgiften observerades i **Studie III**, men inte i **Studie II.**

Ingen av de två studierna visade nära transfer till arbetsminnesuppgifter som inte delade uppgiftsparadigm med de tränade uppgifterna, eller avlägsen transfer till otränade uppgifter som mätte en annan kognitiv funktion än den som hade tränats.

Studie IV fokuserade specifikt på strategiförmedlingshypotesen genom att testa effekterna av en externt given strategi vs. internt genererade strategier under en 30-minuters session av arbetsminnesträning som krävde uppdatering. Resultaten visade att endast en träningssession med en externt given strategi kan ge upphov till det transfermönster som man typiskt ser efter 4-6 veckor av oinstruerad arbetsminnesträning. Resultaten visade också att strategityp och mängden detaljer i de självgenererade strategierna i arbetsminnesuppgiften som krävde uppdatering, var starkt associerat med bättre prestation i samma uppgiftsparadigm.

Sammanfattningsvis tyder resultaten på att träning av verbalt arbetsminne förbättrar prestationen främst i de tränade uppgifterna och i otränade varianter av den tränade uppgiften både hos friska individer och hos patienter med Parkinsons sjukdom. Dessa förbättringar verkar vara medierade av uppgiftsspecifika strategier som deltagarna förvärvar under träningsperioden. Fynden i denna avhandling är således snarare i linje med strategiförmedlingshypotesen än med kapacitetshypotesen.

ABSTRACT

Working memory (WM) training and its possible generalization (transfer to other cognitive domains) has stirred considerable interest over the past years. About a decade ago, the first rather spectacular results indicated that a new type of cognitive training of WM, so-called *core training*, yielded widespread improvements not only in the WM domain, but also in other complex cognitive functions such as intelligence that were related to WM. The repetitive, adaptive *core training* was based on the so-called *Working Memory Capacity Hypothesis*. According to this hypothesis, the inherently limited capacity of WM can be expanded through intensive and challenging training, eliciting improvements also in other cognitive domains where WM has a pivotal role. However, the mechanisms underlying core training remain unclear and the transfer results in subsequent WM training studies have been mixed. Moreover, recent meta-analytic evidence indicates that any substantial transfer following WM training is limited to untrained variants of the trained tasks. These findings speak for an alternative account, namely *the Strategy Mediation hypothesis*. This hypothesis claims that WM training mainly leads to an increased efficiency via employment of task-specific strategies generated during training, while WM capacity as such does not change.

The main aims with the present thesis was to investigate the effects of verbal WM training because of the crucial role of verbal WM in human cognition, including verbal communication and language use, and to examine the viability of the *Strategy Mediation hypothesis* as an account for training-related changes in WM.

Study I served as a methodological prelude by developing and testing a novel sentence-level WM updating task coined as Selective Updating of Sentences (SUS). The development work was motivated by the concern that most commonly used verbal WM tasks typically employ rather artificial materials. The SUS task prompts one to constantly update the contents of semantically and grammatically feasible sentences by replacing content words with new ones, thus bearing similarity to some real-life communicative situations. The results of two experiments showed that the SUS task had adequate psychometric properties and correlated positively with sentence- and paragraph-level verbal episodic measures, indicating that it also taps on memory for more real-life language materials.

Studies II-III examined the effects of home-based core WM training, focusing on near transfer effects (i.e., training-related generalization within the WM domain) in healthy adults **(study II)**, and in patients with Parkinson's disease (PD) **(study III)**.

Both studies showed the expected practice effects in some of the trained WM tasks. Task-specific near transfer to untrained WM tasks structurally similar to the trained ones was observed in **study III**, but not in **study II**. These studies did not show near transfer to untrained structurally dissimilar WM tasks, or far transfer to other cognitive domains following training.

Study IV focused specifically on the Strategy Mediation hypothesis by testing the effects of externally provided vs. internally generated strategies in a 30-minute session of WM updating training. The results showed that just a single session with an externally provided strategy elicited a similar transfer pattern as is typically seen following a 4-6 week long uninstructed WM training period. Moreover, type and level of detail of self-generated strategies used in the WM updating task were strongly associated with WM updating performance.

In conclusion, the present results show that in both normals and in patients with PD, WM training elicits mostly improvements in the trained WM tasks, and in untrained WM tasks structurally similar to the trained WM tasks. A significant part of these improvements could be mediated by task-specific strategies employed by participants during training. Thus, the present findings are more in line with the Strategy Mediation hypothesis than the Working Memory Capacity hypothesis.

1 INTRODUCTION

Working memory (WM) refers to the ability for temporarily maintaining and manipulating incoming information in a readily accessible form (Baddeley, 2000; Cowan, 2014), thus serving as the mental workspace for ongoing cognitive activities. WM is considered to be one of the most influential cognitive constructs in psychology, and it has attracted intense research interest. WM is also a cognitive bottleneck as it can hold only a limited amount of information (Dehn, 2015). While there is high inter-individual variation in WM capacity and the efficiency with which the WM system is utilized, it has been suggested that maximum 4-7 items can be maintained in WM during cognitive processing (Cowan, 2001). Interestingly, inter-individual variability in WM performance has been shown to correlate with nearly all cognitive and metacognitive functions (Dehn, 2015). For example, WM has shown to be highly associated with reasoning (Conway, Kane, & Engle, 2003; Engle, Laughlin, Tuholski, & Conway, 1999; Oberauer, Süß, Wilhelm, & Wittmann, 2008; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002), multitasking (Hambrick, Oswald, Darowski, Rench, & Brou, 2010; König, Bühner, & Mürling, 2005), and executive control (Kane et al., 2007; Poole & Kane, 2009). Importantly, WM has shown to be a significant predictor of several skills crucial for success in life, such as mathematical abilities (Gathercole & Pickering, 2001; Raghubar, Barnes, & Hecht, 2010; Swanson & Beebe-Frankenberger, 2004), reading comprehension (Daneman & Carpenter, 1980; Turner & Engle, 1989), and performance in college entrance exams (Cowan et al., 2005; Turner & Engle, 1989).

Due to its capacity limitation and high relevance to various cognitive abilities and skills, more recent attempts to enhance WM through computerized training have stirred particular interest. While the early studies suggested extensive transfer effects (i.e., generalization to untrained tasks) across several cognitive domains (e.g., Carretti, Borella, Zavagnin, & De Beni, 2013; Chein & Morrison, 2010; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008), many more recent WM training studies and subsequent meta-analyses have indicated only weak or non-existent transfer beyond the WM domain (Guye & von Bastian, 2017; Melby-Lervåg, Redick, & Hulme, 2016; Schwaighofer, Fischer, & Bühner, 2015; Soveri, Antfolk, Karlsson, Salo, & Laine, 2017; von Bastian & Eschen, 2016). The present thesis tackled this controversy by examining the extent of transfer following WM training in health and disease, and by probing into the hitherto unknown cognitive mechanisms underlying WM training. More specifically, the extent of transfer was examined both with a newly developed verbal WM training task and with other commonly employed verbal WM tasks. These studies were conducted in healthy younger adults and in patients with Parkinson's disease (PD). Concerning the underlying mechanisms, in the last study self-reports

were employed to examine whether use of task-specific strategies during training would account for training-related performance improvements.

1.2 Hypotheses on the underlying mechanisms of WM training

As noted above, the extent to which WM is malleable to training is controversial, and the cognitive mechanisms underlying training-induced changes are unclear. Nevertheless, two alternative hypotheses have been put forth to explain training-related changes in WM. The first hypothesis, coined as the *WM Capacity hypothesis,* assumes that the limited capacity system of WM can be enhanced through intensive and challenging training close to individual maximum performance level. Such a capacity increase is assumed to stem from plastic changes in WM-related brain networks (Dahlin, Bäckman, Neely, & Nyberg, 2009), thereby eliciting widespread improvements not only in the WM domain, but also in other complex cognitive functions that are related to WM (Morrison & Chein, 2011; von Bastian & Oberauer, 2014). However, the WM Capacity hypothesis has been criticized for adopting a rather naïve view concerning the malleability of WM, as it compares the brain to a "muscle" that through repeated, adaptive practice would elicit plastic changes in WM-related neural networks relevant for several cognitive tasks (Melby-Lervåg & Hulme, 2013; Morra & Borella, 2015). Consequently, the WM Capacity hypothesis has been challenged by the claim that improved WM functioning could instead reflect a more efficient use of available WM resources (Ericsson & Kintsch, 1995). Thus, an alternative hypothesis, coined as the *Strategy Mediation hypothesis*, assumes that WM is a relatively finite and fixed system, and that training-related increases in WM performance are actually due to compensatory strategies that participants develop during training (Dunning & Holmes, 2014; McNamara & Scott, 2001). As strategies are expected to be more or less specific to the task(s) one is training (e.g., Chase & Ericsson, 1982; Lustig et al., 2009; Maguire, Valentine, Wilding, & Kapur, 2003), the Strategy Mediation hypothesis does not presuppose widespread transfer to other cognitive domains following training. Rather, it assumes that enhanced WM functioning due to employment of task-specific strategies enhances performance mostly in such tasks where the same strategies can be applied, and not in other untrained tasks where the corresponding strategies would be ineffective or unsuitable (Lustig et al., 2009; Morrison & Chein, 2011).

Studies addressing the WM Capacity hypothesis in the context of WM training have typically employed a so-called *core training* approach (Morrison & Chein, 2011; Shipstead, Redick, & Engle, 2012). A crucial component in such a protocol is its adaptivity: task difficulty is adjusted depending on the individual's performance, thus forcing the participants to constantly perform the training task(s) close to their

maximum WM capacity. A typical setup would include four to six weeks of training, with three to four 30-minute sessions per week. As regards strategies, it has been emphasized that core training needs to be entirely unsupervised, so that participants would be less likely to adopt compensatory strategies (e.g., chunking or rehearsal), as strategies are expected to constrain the generalization effects (Lustig et al., 2009). Besides these features, it has been suggested that the administered core training task(s) should prompt active manipulation of information in WM (Jaeggi et al., 2010), avoid automatization (Morrison & Chein, 2011), and emphasize rapid encoding and retrieval demands of information in WM (Morrison & Chein, 2011). The Strategy Mediation hypothesis, on the other hand, has been examined with adaptive WM training combined with external strategy instructions (Bailey, Dunlosky, & Hertzog, 2014; Carretti, Borella, & De Beni, 2007; McNamara & Scott, 2001; Peng & Fuchs, 2017; Turley-Ames & Whitfield, 2003), and/or by evaluating the use of participant-generated strategy use through self-reports (Bailey et al., 2014; Dunning & Holmes, 2014; McNamara & Scott, 2001). However, most of the WM training studies have either explicitly or implicitly relied on the WM Capacity hypothesis that characterized the influential adaptive WM training experiments published about a decade ago, and focused on the extent of transfer rather than on its underlying mechanisms. Thus, especially the less studied Strategy Mediation hypothesis deserves further scrutiny.

1.3 Earlier research on WM training

This chapter provides an overview of previous studies examining the transfer effects of core WM training that is related to the WM Capacity hypothesis, and WM training strategy studies addressing the Strategy Mediation hypothesis. Central in the context of earlier WM training research is the concept of *transfer* where most studies make a distinction between *near* and *far* transfer. Near transfer occurs when improvement is seen in tasks that measure the same cognitive domain that has been practiced. Far transfer, on the other hand, occurs when improvement is seen in other cognitive domain(s). Thus, near transfer effects would correspond to improvements on untrained WM tasks, while far transfer would correspond to improvements in tasks measuring, for instance, fluid intelligence, or set shifting. Besides this division, some recent studies have separated near transfer further into two different categories based on their closeness to the trained task(s). One of these categories (herefort *task-general near transfer*) refers to generalization to untrained WM tasks that do not employ the same task paradigm as the trained ones, while the other category (herefort *task-specific near transfer*) refers to generalization to untrained WM tasks that are based on the same task paradigm as the trained ones (Soveri et al., 2017). To take an example, improvement on a digit span task following letter span training

would represent task-specific near transfer, while letter span training leading to improvement on an n-back task would indicate task-general near transfer.

1.3.1 Studies addressing the WM Capacity hypothesis

Since the early 2000s, the initial core WM training studies attracted widespread interest due to their promising results. In a pioneering training study by Klingberg et al. (2002) with children having ADHD, the results showed far transfer effects to fluid intelligence, and even decreased hyperactivity in everyday situations. Another remarkable study by Jaeggi and colleagues (2008) showed that core training enhanced fluid intelligence in healthy young adults, a finding that the authors succeeded to replicate a few years later (Jaeggi et al., 2010; Jaeggi, Buschkuehl, Shah, & Jonides, 2014). Also other early studies pointed to the same direction, namely that core WM training yielded transfer to skills relevant for everyday life, such as reading comprehension (Chein & Morrison, 2010), mathematics (Holmes, Gathercole, & Dunning, 2009), and sustained attention (Brehmer, Westerberg, & Bäckman, 2012).

However, following these rather intriguing far transfer effects, a second wave of research identified several critical methodological shortcomings in the early studies, such as selective reporting of significant and nonsignificant results (McCabe, Redick, & Engle, 2016), not including an active control group (Morrison & Chein, 2011), failures of controlling for between-group expectancies (Boot, Simons, Stothart, & Stutts, 2013; Simons et al., 2016), employing small samples (Melby-Lervåg et al., 2016), only using a single task for measuring the domain of interest (Shipstead et al., 2012), or applying unjustified one-tailed *t*-tests (McCabe et al., 2016). Subsequent methodologically more sound core WM training studies have shown highly mixed results regarding the extent of transfer (e.g., Bäckman et al., 2017; Baniqued et al., 2015; Bergman-Nutley & Klingberg, 2014; Chooi & Thompson, 2012; Clark, Lawlor-Savage, & Goghari, 2017; Colom et al., 2013; Lawlor-Savage & Goghari, 2016; Lilienthal, Tamez, Shelton, Myerson, & Hale, 2013; Owen et al., 2010; Salminen, Strobach, & Schubert, 2012; Swanson & McMurran, 2018; Thompson et al., 2013; Waris, Soveri, & Laine, 2015). What makes the situation even more complicated is that most studies have focused mainly on far transfer effects, and have not systematically investigated the transfer pattern within the WM domain by taking into account how much the untrained WM tasks (where more substantial transfer effects are observed) resemble the training task(s) (Soveri et al., 2017).

Only a few core WM training studies with larger samples and with adequate methodology have investigated the nature of the near transfer effects in more detail, by taking into account the important separation of task-specific- and task-general near transfer measures (De Simoni & von Bastian, 2018; Guye & von Bastian, 2017;

Sprenger et al., 2013; von Bastian & Eschen, 2016).[1] Sprenger et al. (2013) investigated the extent of transfer in healthy young adults following 3-6 consecutive weeks of core WM training in two separate experiments ($N = 253$). Both experiments employed a WM training group that practiced with eight training tasks over 20 training sessions (30 min/session). Experiment 1 included passive controls as the reference group, and experiment 2 active controls that practiced with a recognition task for the same amount of sessions as the WM training group. The results from both experiments indicated improvements on the trained tasks and on tasks that either shared the task paradigm or task stimuli with the training tasks, while transfer to other untrained WM measures or far transfer measures was absent. Similarly to the Sprenger et al. (2013) study, von Bastian and Eschen (2016) examined the transfer effects following 20 sessions of WM training over a 4-week period in healthy younger adults. The participants were randomly allocated either to an adaptive WM training group ($n = 34$), a WM training group that practiced with a randomly adjusted task difficulty ($n = 30$), a WM training group that practiced with a self-selected task difficulty ($n = 34$), or an active control group that received trivia- and general knowledge training ($n = 32$). The pre- and posttest tasks included three training tasks that probed for the training-related practice effects, three task-general near transfer measures, and five far transfer measures. Compared to the active controls, the WM training groups showed substantial improvements in the trained WM tasks, but no transfer was seen either to untrained structurally dissimilar WM tasks or to tasks tapping other cognitive domains.

Guye and von Bastian (2017) investigated the effects of WM training (altogether 25 training sessions over a 5-week period) with a sizable sample of 142 healthy older adults (WM training group $n = 68$, active control group $n = 74$), and used a pre- and posttest battery of tasks measuring task-general near transfer and three far transfer domains (fluid intelligence, set shifting, and inhibition). Using Bayesian linear mixed modeling, their results provided consistent evidence for the absence of near transfer to untrained structurally dissimilar WM tasks, and a substantial to strong evidence for the absence of far transfer to other cognitive domains. Employing an identical analytical approach and a partially similar training procedure as in Guye and von Bastian (2017), De Simoni and von Bastian (2018) recently investigated the transfer effects by randomly allocating healthy younger adults to two WM training groups and one active control group. The training groups received either WM training with updating ($n = 81$) or WM training with binding ($n = 77$), and the control group participated in visual search training ($n = 75$). Each group completed 20 training

---

[1] The main objective in these studies was not only to investigate the extent of transfer within the WM domain, but to examine whether the transfer effects extended to other cognitive domains as well.

sessions over a five-week period. The transfer pattern was examined with a pre-post battery comprising of the training tasks, untrained structurally dissimilar WM updating and binding tasks, and four far transfer measures tapping on reasoning, set shifting, speed, and inhibition. Albeit each group improved significantly in their respective training task following training, the results provided substantial and consistent evidence for the absence of task-general near transfer and far transfer, suggesting that training did not enhance WM capacity.

Based on previous studies accumulated over a decade or so, several meta-analyses have also investigated the transfer pattern following core WM training. The meta-analytic results are unequivocal with respect to the far transfer effects: in healthy adults, only small, yet statistically significant effect sizes have been observed to other cognitive domains such as fluid intelligence (Au et al., 2015; Karbach & Verhaeghen, 2014; Soveri et al., 2017), cognitive control (Soveri, et al., 2017; Weicker, Villringer, & Thöne-Otto, 2016), verbal ability (Schwaighofer et al., 2015), and arithmetics (Melby-Lervåg et al., 2016). With regard to the near transfer effects, the results are slightly less congruent, mainly stemming from differences in how the near transfer tasks have been categorized in the meta-analyses. Those meta-analyses that have merged task-specific transfer tasks and task-general transfer tasks into one single WM domain indicate improvements with small-to-moderate effect sizes (e.g., Melby-Lervåg et al., 2016; Weicker et al., 2016), while the most recent meta-analysis that separated these two near transfer domains indicated a moderate task-specific near transfer effect, and only a small task-general near transfer effect that was equal to the weak far transfer effects observed (Soveri et al., 2017). All in all, the most recent meta-analyses indicate that core WM training fails to yield any substantial improvements in cognitive functioning beyond the WM domain. This is against the predictions of the WM Capacity hypothesis.

1.3.2 Studies addressing the Strategy Mediation hypothesis

As previously noted, only a limited amount of studies have explored the outcomes of strategy-based WM training (Bailey et al., 2014; Carretti et al., 2007; Dunning & Holmes, 2014; McNamara & Scott, 2001; Peng & Fuchs, 2017; Turley-Ames & Whitfield, 2003). McNamara and Scott (2001) examined whether strategy use was related to improved WM performance following training in two experiments. Experiment 1 was a within-subjects design where participants ($n = 21$) attended a two-week study encompassing a pretest, four sessions of short-term memory training (comprised of 22 to-be-recalled word lists across the four training sessions) with an explictly given storytelling strategy, and a posttest. The results showed that the four strategy training sessions did not only improve performance on the short-term memory training task, but training also led to an increased performance in a complex

span task. Experiment 2 extended the first experiment by allocating participants to either a WM strategy training group (n = 30) receiving a chaining strategy during a four-session training period with word list recall (identical to the task employed in Experiment 1), or a control group (n = 30) that practiced for the same amount of time with the same task but without any strategy instructions. Following training, the strategy training group showed reliable gains in a complex span task compared with the active controls. Moreover, a retrospective evaluation of individual strategy reports using a 4-point categorical scale (0 = no strategy use; 1 = rehearsal; 2 = combining rehearsal with more semantic processing; 3 = semantic processing by connecting the words to sentences, stories or mental images) showed that the more strategic participants also had better WM performance.

Turley-Ames and Whitfield (2003) investigated the effects of explicitly given strategies on WM performance in three experiments. Experiment 1 encompassed a pretest-posttest design administered on two consecutive sessions where half of the enrolled participants (n = 66) received a rehearsal strategy prior to posttest, while the other half (n = 58) did not receive any strategy instructions. The pre- and posttest battery included a WM task (an Operation span task) and a reading ability task (the Nelson Reading test). The results showed that those receiving the rehearsal strategy showed significantly better Operation span performance at posttest compared to the no-strategy group, but no group difference was seen in reading ability. The design and the pre-and posttest battery in Experiment 2 was similar to Experiment 1, but included three experimental groups, each of which were given a different strategy prior to posttest (rehearsal n = 90, imagery n = 90, and semantics, n = 90). Compared with a no-strategy group (n = 90), the rehearsal group and the semantics group improved more from pre- to posttest in the Operation span task. Moreover, the rehearsal strategy proved to be particularly beneficial for those with poor WM performance at pretest. However, in line with Experiment 1, no group differences were seen in the reading ability test at posttest. Experiment 3 employed a similar setup as in Experiment 2, the only difference being the time for solving the processing and storage task that was constrained to 3 and 4 seconds respectively in the Operation span task. Here, the results showed no group differences in neither the Operation span task nor the Nelson reading test at posttest. However, an additional analysis showed that the correlation between the Operation span task and the Nelson Reading test at posttest was almost twice as high in the rehearsal strategy group (r = .79; n = 45) than in the control group (r = .47; n = 45), suggesting that strategies may suppress the true relationship between WM performance and higher-order cognition such as reading comprehension.

Using a design similar to McNamara and Scott (2001), Carretti et al. (2007) allocated healthy younger adults (n = 14) and healthy older adults (n = 13) to groups

that practiced with an imagery-based strategy during three sessions of memory training, or to age-matched control groups (younger controls $n = 14$, older controls $n = 14$) that trained for the same period of time but without strategy instructions. The memory training comprised of a immediate list recall task that included word lists ranging from 10-15 words. Training gains were measured with a pre- and posttest version of the trained list recall task, and training- related transfer was measured with a complex span task tapping on WM. The results showed that the strategy instruction was beneficial, yielding training effects in the immediate list recall task at posttest in both the younger and older adults as compared to the controls. Importantly, improvements were seen also in the complex span task in both age groups, thus concurring with the findings observed in McNamara & Scott (2001).

Bailey et al. (2014) examined to what extent strategy-based WM training generalized to other untrained tasks in both younger and older adults. In both age groups, the participants were randomized to a WM training group (younger $n = 41$, older $n = 39$) instructed to use mnemonic strategies when practicing with a span task, or to an active control group (younger adults $= 38$, older adults $n = 39$) training with a filler task that loaded only weakly on memory processes. Both groups took three training sessions (30-60 min/session) before completing the posttest. Compared to the controls, for both age groups the WM training group improved on an untrained WM task (i.e., a complex span task) similar to the trained one, but no transfer was seen to the two untrained tasks that tapped on other domains, namely a Paired-Associate Cued recall task and a Self-Ordered Pointing task (note that these untrained transfer tasks were administered only at posttest).

Of particular relevance to the Strategy Mediation hypothesis in the context of typical core WM training is the recent study by Dunning and Holmes (2014). Here, the participants were randomized either to a WM training group ($n = 14$), an active control group receiving non-adaptive WM training ($n = 14$), or a passive control group ($n = 15$). The study was otherwise similar to the typical core training studies, the only exception being that the training period encompassed 10 training sessions (i.e., slightly less than in the typical training protocols), and that participants' strategy reports were gathered during pre- and posttest. The results showed a significant increase in the self-reported use of a Grouping strategy (e.g., shaping separate chunks of multiple individual items) for the adaptive training group as compared to the active and passive controls in most of the pre-post tasks. Moreover, performance improvements and the Grouping strategy co-occurred in two untrained WM tasks that shared both the task paradigm and stimulus materials with the training tasks, thus indicating an increased use of internally generated strategies on the same tasks that had been practiced during training.

1.3.3 Summary of earlier research

The initial studies on core WM training clearly supported the WM Capacity hypothesis, as they indicated far transfer effects to other cognitive domains (e.g., Borella et al., 2010; Jaeggi et al., 2008; Richmond, Morrison, Chein, & Olson, 2011). However, the more recent and methodologically more rigorous studies and meta-analyses indicate much more limited transfer that does not support the idea of an increased WM capacity following training (De Simoni & von Bastian, 2018; Guye & von Bastian, 2017; Melby-Lervåg et al., 2016; Soveri et al., 2017). Rather, the limited generalization to untrained WM tasks following training provides indirect evidence for the Strategy Mediation hypothesis. According to this hypothesis, WM capacity as such does not change as a result of training, but training leads to an improved WM efficiency via employment of strategies that are adopted during training. Taking all the current available evidence on WM training effects into account, two outstanding questions of both theoretical and practical relevance emerge:

A) If the transfer effects are as limited as the recent meta-analytic evidence indicates, how should one move forward with core WM training studies?

B) How could the hithertho unknown mechanisms underlying WM training be clarified?

To address the first question, one option would be to start employing core WM training paradigms that are closer to "real-life" demands and thus potentially more directly relevant for the participants (Soveri, Karlsson, Waris, Grönholm-Nyman, & Laine, 2017). Here an attempt in this direction was to develop a training paradigm that employs meaningful sentences rather than random sequences of isolated items as training stimuli. As most people in modern countries read texts on a daily basis, practicing with sentences might have influence in real-life settings if it would enhance verbal WM relevant for sentence processing. Verbal WM is considered to be crucial for sentence processing, because handling of sentences requires rapid computation of linguistic relations between temporally distal parts of the sentence (Lewis, Vasishth, & Van Dyke, 2006).

To shed light on the unresolved mechanisms underlying WM training effects, the present thesis includes a study that tested directly the Strategy Mediation hypothesis. The recent meta-analysis by Soveri et al. (2017) suggests a transfer pattern that seems to fit with this hypothesis, namely that more substantial transfer from WM training is limited to untrained WM tasks structurally similar to the trained ones. In other words, task-specific strategies that participants adopt during training could easily be applied to structurally similar untrained WM tasks, but would not necessarily benefit

performance on other, dissimilar pre-post tasks. Only a few studies have specifically probed the Strategy Mediation hypothesis, but they show substantial performance improvements when participants have been taught efficient strategies such as storytelling (McNamara & Scott, 2001), imagery (Carretti et al., 2007), and rehearsal (Turley-Ames & Whitfield, 2003). Moreover, they indicate that both externally given strategies and self-generated strategies developed during training improve WM performance mostly on the trained tasks (Bailey et al., 2014; Dunning & Holmes, 2014; McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003; but see Carretti et al., 2007). None of these studies have specifically examined the relationship between task-specific strategies and the recently identified near transfer pattern (task-specific but not task-general transfer) typically seen following WM training, thus prompting further research on the topic.

2 AIMS AND RESEARCH QUESTIONS

The main aim of the present thesis was twofold: to investigate the extent of transfer following core training of verbal WM, and to examine the viability of the Strategy Mediation hypothesis as an account for training-related changes in WM. The specific aims of the four studies were as follows:

Study I. As a methodological prelude for the present thesis, a novel sentence-level WM task coined as the Selective Updating of Sentences (SUS) task was developed and tested. This was motivated by the concern that most commonly used verbal WM tasks typically employ rather artificial materials. Based on the updating component of WM (Miyake et al., 2000), the SUS task prompts participants to constantly update the semantic contents of feasible sentences by replacing their constituent words with more recently presented words. In two experiments, we sought to examine the psychometric properties of the SUS task, and probed for its "ecological relevance" by investigating its associations with verbal episodic memory tasks at sentence and paragraph level. Apart from the validation, we also tested the viability of the SUS paradigm as a core WM training task in the subsequent training studies (i.e., **Studies II-III** as listed below).

Study II. Using a randomized controlled design, this study investigated the transfer effects following home-based sentence-level core verbal WM training in healthy young adults. The employment of a sentence-level based training paradigm was motivated by promising results in a recent study (Payne & Stine-Morrow, 2017) where a partly similar core training regime was employed. In that study, transfer was observed not only to untrained structurally dissimilar verbal WM tasks, but also to measures tapping on sentence recall, comprehension of syntactically ambiguous sentences, and verbal fluency.

Study III. Employing a randomized controlled design on patients with Parkinson's disease (PD), this study examined the efficacy of home-based core training targeting mainly verbal WM updating. With a carefully compiled pre- and posttest battery, including a number of cognitive tasks and self-assessment measures tapping on executive and WM functioning as well as symptoms of depression, we charted the near and far transfer effects of the training regime. This study grew out from the notion that surprisingly few cognitive training studies have been administered in home-based settings in clinical groups, and to our knowledge, no previous study has specifically investigated the detailed pattern of transfer effects following core WM training in patients with PD.

Study IV. This study tested the viability of the Strategy Mediation hypothesis as an account for training-related changes in WM. In a randomized controlled design encompassing two intervention groups (strategy-based training vs. uninstructed training) and a passive control group, transfer effects were examined after just a single 30-minute session of WM updating training. Besides the effects of external strategy that enabled us to make causal claims, we also examined whether type and level of detail of internal, self-generated strategies was associated with an improved performance in WM updating tasks.

3 PARTICIPANTS

In all studies, the participants (see Table 1) had to give their informed consent. Moreover, all enrolled participants had to complete a prescreening questionnaire targeting key background variables (e.g., age, education, possible neurological and psychiatric diseases, and reading difficulties) for being eligible to take part in the study.

3.1 Study I

This study consisted of two experiments. The participants in both of them comprised of neurologically and psychiatrically healthy university and polytechnics students. Moreover, only Finnish native speakers with little or no exposure to any other language during childhood were included. Experiment 1 comprised of 170 participants with a mean age of 25.2 years ($SD = 5.2$) out of which 76.9% were females. Upon study completion, the participants were rewarded with a movie ticket. The data in Experiment 2 stems from the pretest session of the WM training study reported in **Study II**, consisting 80 participants with a mean age of 24.4 years ($SD = 3.6$). Also in this Experiment, most of the enrolled participants were females (74.7%).

3.2 Study II

Altogether 68 healthy Finnish university and polytechnics students were included in the study ($M_{age} = 24.6$, $SD = 4.0$) after excluding 22 participants for not meeting the inclusion criteria and 11 participants for withdrawing during the intervention period (see Figure 2 for more details regarding the exclusion criteria and the attrition rate during the study). The participants were randomly allocated to either a WM training group ($n = 31$) or an active control group ($n = 37$). The groups were comparable with regard to age, gender, education, and (non-clinical) depressive symptoms as measured with the Beck Depression Inventory-II (BDI-II; Beck, Steer, & Brown, 2004). The participants were compensated with 70 euros upon study completion.

3.3 Study III

In total, 64 patients with PD were recruited to take part in the study. Participants with severe disabilities (i.e., unable to live an independent life), severe motor fluctuations or dyskinesias, or severe health problems, were excluded from the study. Moreover, those patients that had been exposed to two or more languages before the age of six, or had participated in previous WM training studies, were deemed as illegible participants. The final sample in the analyses consisted of 52 participants ($M_{age} = 65.1$, $SD = 5.5$) after excluding six participants that did not meet our inclusion

criteria, and six participants for withdrawing following pretest or during intervention (for more details concerning the attrition rate, see Figure 3). The participants were block-randomized pairwise to either a WM training group (n = 26) or an active control group (n = 26). There were no significant group differences with regard to age, gender, education, age at diagnosis, or disease duration between the two intervention groups. The participants did not receive any monetary compensation upon study completion.

3.4 Study IV

For this study, we recruited 127 healthy Finnish-speaking university and polytechnics students free from psychiatric and neurological illnesses (including dyslexia). Five of these participants withdrew from the study during the intervention and two participants were excluded because of previous participation in cognitive training studies. Moreover, one participant was excluded for not taking the posttest on time, one for being a multivariate outlier, and one due to missing training data (for a more detailed attrition rate, see Figure 4). This yielded a final sample size of 116 participants (M_{age}, = 24.2, SD = 4.4). The participants were randomized to three groups: a WM training group (n = 40), an active control group (n = 37), or a passive control group (n = 39). These three groups did not differ significantly with regard to age, gender, education, or well-being as measured with the Patient Health Questionnaire (PHQ-9; Spitzer, Kroenke, & Williams, 1999). Upon study completion, the participants were compensated with 40 euros.

Table 1

Summary table of the research aims and participant characteristics for the four studies.

Study	Research aim	Participants	N	Age in years	F/M
I	Validating a novel WM updating task	Young adults	Experiment 1: 170	19-47	131/38*
			Experiment 2: 80	18-39	59/21
II	Sentence-level WM training	Young adults	31 training group	18-39	24/7
			37 control group	19-39	28/9
III	Verbal WM updating training	Patients with Parkinson's disease	26 training group	55-72	17/9
			26 control group	45-72	17/9
IV	Instructed vs. uninstructed WM training	Young adults	40 Strategy training group	19-40	29/11
			37 Active control group	20-37	28/9
			39 Passive control group	19-38	28/11

*One participant chose "other" for the gender variable.

4 METHODS

This chapter provides a brief summary of the methods used in the four studies including the procedure, the outcome measures employed (see also Table 2 for summary of the used cognitive tasks), and the training tasks administered during the intervention period in **Study II-IV**. More detailed descriptions of the procedures are given in the original articles.

4.1 Study I

Procedure

In Experiment 1, we employed a fully online procedure in which the participants completed a set of cognitive tasks on their own in a quiet place where they would not be disturbed or interrupted. In Experiment 2, the participants completed the cognitive tasks in computer classes.

The cognitive tasks employed in Experiment 1

The cognitive task battery in Experiment 1 comprised of four cognitive tasks. One of these tasks was the to-be-validated Selective Updating of Sentences (SUS) task, and the three others were different complex span tasks (CSTs), namely a Reading span task, an Operation task, and a Symmetry task.

Selective updating of sentences (SUS). In this novel verbal WM task, participants had to memorize semantically feasible sentences and concurrently update them when some constituent words were replaced with newly presented words (see Figure 1 for an illustration of the SUS task). Each trial started with a sentence that was presented on the computer screen for 4000 ms, so that each word appeared in a box. The task was to keep in mind the words in the initial sentence before it disappeared. After this, a blank screen was displayed for 500 ms, whereafter an updating stage occurred for 4000 ms with a new row of boxes. The number of updating stages varied between trials, but common to all of them was that two of the boxes comprised of new words while the rest of the boxes were empty. The participants had to replace the old words with the new ones, while at the same time maintaining those words in WM that did not change. When the sentences were gradually updated and replaced by new words, they remained semantically and syntactically feasible. At the end of each trial, the participants had unlimited time to recall the latest version of the sentence, including the most recently updated words. The percentage of correctly recalled words on all trials was used as the dependent variable.
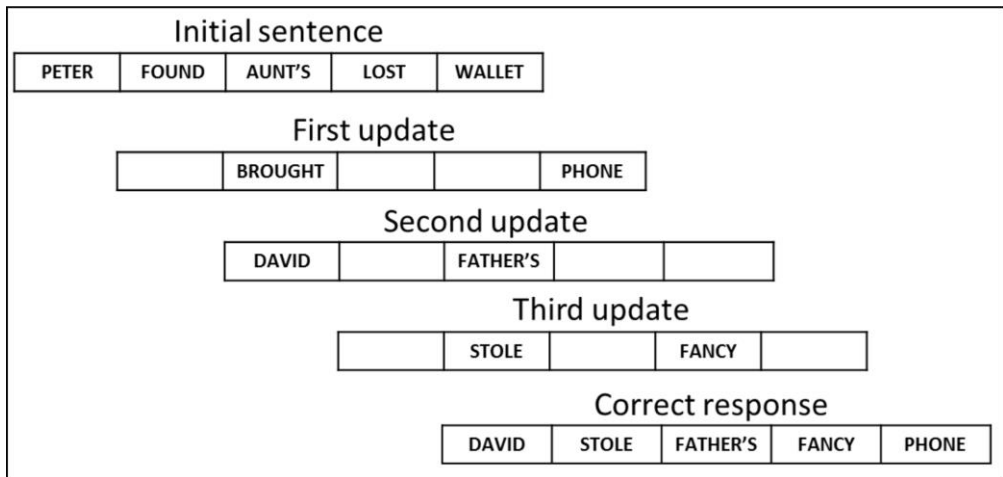
Figure 1. An English translation of a sample trial in the SUS task including a sentence with five words, and three updating stages. The SUS task encompassed altogether 12 trials that occurred in a pseudorandomized order (one trial comprised of an initial sentence followed by its updating stages). The 12 trials were split into three blocks, with 4 trials in each block. The blocks differed in terms of the number of updating stages (2 updates, 3 updates, and 5 updates) while being similar in terms of sentence length so that one trial of each sentence length (range 4-7 words) was presented in all blocks. The Finnish sentences in the SUS task comprised of ordinary declarative sentences with a canonical SVX word order of both transitive and intransitive type. The sentences, included e.g., predicative clauses, transitive clauses, intransitive clauses, ownership clauses, and existential clauses. Note that the Finnish language has no articles.

Reading span task. We employed the Reading span task following Daneman and Carpenter (1980). In each trial, the participants had to memorize the final words of a set of sentences and to simultaneously judge whether the sentence was semantically correct or not. The sentences were displayed for 8000 ms, and the next sentence occurred directly after a response was provided. At the end of each trial, the participants had unlimited time to recall the final words by typing them up into empty boxes that were shown on the computer screen. The participants completed seven trials, with one trial per list length (list lengths ranged from two to eight words). The order of trials was randomized. The dependent variable was based on a partial credit scoring (Conway et al., 2005), consisting of the number of correctly recalled words per trial, regardless of trial length.

Operation span task. Following Turner and Engle (1989), we administered the Operation span task. This task prompted the participants to recall a series of unrelated digits (exposure time = 1000 ms). Between the to-be-remembered digits, the participants were to solve simple math problems that served as distractors (exposure

time = 6000 ms). At the end of each trial, the participants were to recall the digits in the order they were presented. The Operation span task included six trials (list lengths ranged from four to nine digits) that were presented in a randomized order. The dependent variable was the partial credit scoring employed in the Reading span task.

Symmetry span task. In this computerized Symmetry span task (Kane et al., 2004), the participants were to memorize a series of positions of black squares (exposure time = 1000 ms) in a 3 x 3 grid, and at the end of each trial, to recall them in a correct serial order. The to-be-remembered stimuli were interleaved by distractor tasks (exposure time = 6000 ms) where the participants were to mentally merge two 3×3 matrix patterns in order to decide whether their combination corresponded to a third pattern shown on the screen. The participants completed five trials (list lengths ranged from three to seven items) with one trial per list length. The presentation order of the trials was randomized. As in the other complex span tasks, the dependent variable was based on the partial credit scoring.

The cognitive tasks employed in Experiment 2

The cognitive task battery employed in Experiment 2 included five WM tasks, and five tasks tapping verbal episodic memory. The Selective updating of sentences task, Reading span task, and the Operation span task in Experiment 2 were identical to the ones employed in Experiment 1. The other administered tasks are shortly described below.

Minus 2 span task. In the Minus 2 span task (Waters & Caplan, 2003), the participants were prompted to memorize a set of digits in a serial order (with list lengths ranging from two to nine digits), while simultaneously subtracting 2 from each digit. For example, if a digit sequence 4-9-3-6-8-3 was presented, the correct response would have been 2-7-1-4-6-1. After each presented sequence that was shown for 1000 ms, a recall grid was shown where the participants typed in their response. The participants completed 12 trials in a randomized order, with two sequences per list length. The dependent variable was the total number of digits recalled in the correct serial position.

Alphabet working memory task. The Alphabet working memory task followed the description in Was et al. (2011). Here, the participants were presented with either one letter or two alphabetically nonadjacent letters for 2500 ms, followed by a transformation phase where direction and number cues (−3, −2, −1, +1, +2, +3) were given. In this transformation phase, the participants were to move either up or down in the alphabet according to the aforementioned cues (e.g., the letter pair GB + 3 = ID). The transformation cue was displayed on the screen until the participants decided to proceed with the task. Following this, they were to give their response. The participants completed 18 trials in a randomized order, with half of the trials including

one letter, and the other half two letters. The dependent variable was the proportion of correctly recalled items per minute.

Sentence recall. We administered both a Finnish and an English version of this task. These two versions were similar regarding the number of words in the sentences and the scoring procedure. Here, words of a sentence were presented successively on a computer screen at a rate of one word per 1000 ms. Immediately after the sequence ended, the participants were prompted to recall the sentence by typing it up in an empty column. Both the Finnish and the English versions had five sentences that were presented in a randomized order. As the dependent variable, we employed a verbatim scoring procedure, that is, the proportion of correctly recalled words, regardless of the order in which they were recalled.

Paragraph memory. In this task, the participants were presented with two separate paragraphs one at a time in random order for an unlimited time. The participants' task was to recall the key points using the original words from the paragraph. When the participants decided to proceed, an empty box appeared on the screen, and they were to write down as much of the paragraph as possible. We employed two separate dependent variables in this task; a verbatim scoring procedure comprised of the proportion of correctly recalled words, and a semantically based scoring procedure yielded the proportion of correctly recalled semantic contents that had been determined beforehand when designing the paragraphs.

Word fluency. We also employed a computerized version of the word fluency task (Benton & Hamsher, 1978). Here, a letter was displayed on the computer screen, and the participants were to type as many words as possible beginning with that letter for 60 seconds. The participants completed three trials with different letters, the order of which was randomized. The sum of the unique correctly reported words across the three trials served as the dependent variable in this task.

Statistical analyses

In Experiment 1, we examined the internal consistency of the SUS task using the Cronbach's alpha coefficient, and by investigating the item-total correlations (ITCs) between the SUS items. The internal consistency was also examined in the other WM tasks that the SUS task was validated against. As we expected that the SUS trials including more words and updating stages would be more difficult, we analyzed the task manipulations (i.e., number of updating stages; sentence length) with two-way repeated measures analysis of variance (ANOVA). The concurrent validity of the SUS task was investigated by examining its intercorrelations with the three CSTs, using the Pearson correlation coefficients.

Also in Experiment 2, the internal consistency of the SUS task was examined with the Cronbach's alpha coefficient, and the ITCs. Likewise, the effects of task

manipulations (i.e., number of updating stages; sentence length) in the SUS task were analyzed with the two-way ANOVA. The test-retest reliability and concurrent validity of the SUS task were examined with the Pearson correlation coefficients. Moreover, hierarchical two-step multiple regression analyses were used to explore whether the SUS task predicted performance in the sentence and paragraph recall tasks over and beyond the CSTs. The CSTs were thus set as predictors at stage 1 of the model, and the SUS task was added as a predictor at stage 2.

4.2 Study II

Procedure

All participants took part in the pretest before being allocated to either a WM training group, or an active control group (see Fig. 2). Following training, all participants completed the posttest. The training period took four weeks during which both groups practiced four times per week, each training session lasting for 30 minutes. The pre- and posttests were administered as computer class sessions where several participants completed the computerized tasks simultaneously. The training was performed in home-based settings via the participants' home computers.
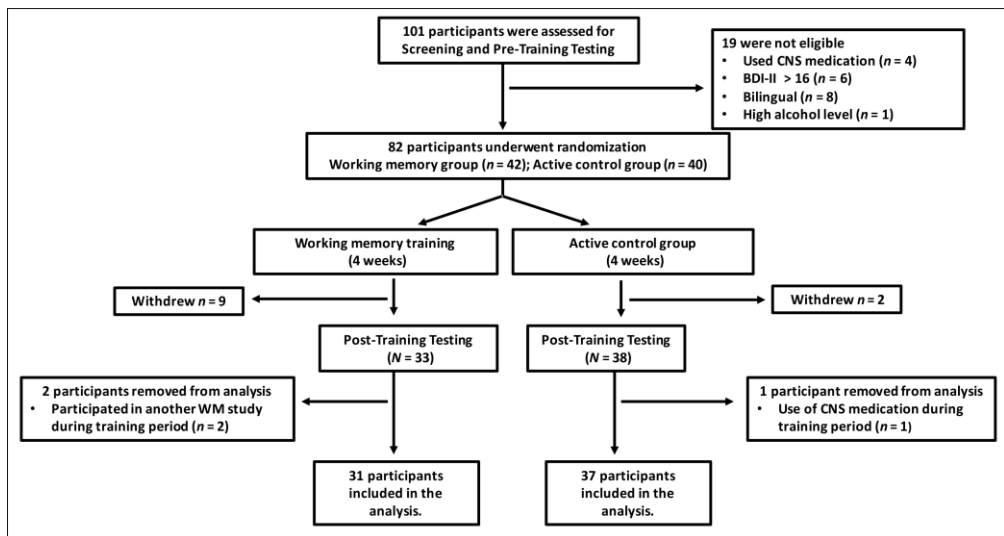
Figure 2. Study II: Flowchart of the attrition rate at the different stages of the study.

Training paradigm

The WM training group practiced with two adaptive sentence-level WM tasks. One of the training tasks was based on the Selective Updating of Sentences (SUS) paradigm that was validated in **Study I**. This training variant, coined as the Selective Updating of Sentence Training (SUST) task, was identical to the SUS task except for

its adaptive nature and a sizable pool of unique sentences used as training stimuli (altogether 640 unique trials; a single trial consisted of an initial sentence and its updating stages). As regards the SUST adaptivity, altogether 40 levels were created. The first eight levels consisted of sentence trials (range 4-7 words) that were presented for 4000 ms, the difficulty level of which was increased by adding more updating stages (level 1 = one updating stage, level 8 = eight updating stages). If the participants successfully completed level 8, they proceeded to level 9 that comprised of sentence trials with one updating stage, but the presentation time was now decreased to 3000 ms. The participants then advanced with this presentation time through updating stages 1-8 (i.e., level 9-16). Consequently, the same adaptive procedure was employed at levels 7-24 (2000 ms presentation time) and levels 33-40 (500 ms presentation time). The sentence trials at each difficulty level appeared in blocks of four. At least three of the four trials had to be correctly recalled before proceeding to the next difficulty level. With two correctly recalled trials, one stayed at the same level, whereas less than two correctly recalled trials resulted to a decrease in level.

The second training task was based on the well-established Reading span paradigm. The task was otherwise identical to the Reading span task in **Study I**, but here the difficulty level was adjusted depending on the participant's performance (i.e., higher difficulty entailed more to-be remembered final words in a sequence).

The active control group practiced for the same amount of time on a quiz task tapping on general knowledge on the website www.alypaa.com.

Tasks administered at pre- and posttest

The pre-and posttest tasks were the same as described in Experiment 2 in **Study I**, namely the Selective updating of sentences task, Reading span task, Operation span task, Alphabet working memory task, Minus 2 span task, Sentence recall in Finnish, Sentence recall in English, Paragraph memory, Paragraph semantic memory, and Word fluency. The SUS task and the Reading span task were criterion tasks, that is, these were pre-posttest versions of the training tasks used during the intervention. The Operation span task served as a task-specific transfer measure (i.e., it shared the same task paradigm with the trained Reading span task), whereas the Alphabet working memory task and the Minus 2 span task served as task-general transfer measures (i.e., these were untrained and structurally dissimilar WM tasks). The Sentence and Paragraph recall tasks and the Word fluency task were far transfer tasks.

Statistical analysis

We conducted independent samples t-tests for examining baseline differences for continuous variables, and χ^2 tests for categorical variables between the two training groups (i.e., WM training group vs. active controls). To examine the progression

during the training period in the WM training group, we employed one-way repeated measures ANOVAs separately for both training tasks. Pre-post changes on the cognitive measures between the two groups were examined using analysis of covariance (ANCOVAs) with posttest performance set as the dependent variable, group as the fixed factor, and pretest performance as the covariate.

4.3 Study III

This study was a fully home-based randomized controlled design spanning over an eight-week period (see Figure 4). All participants took part in pre- and posttest assessment. After completing the pretest during the first week of the study, the participants were randomized in pairs (according to age and gender) either into a WM training group or an active control group. Both groups practiced thrice a week (30 minutes per session) for the following five weeks. After a successfully completed training period, the participants completed the posttest during the next week.
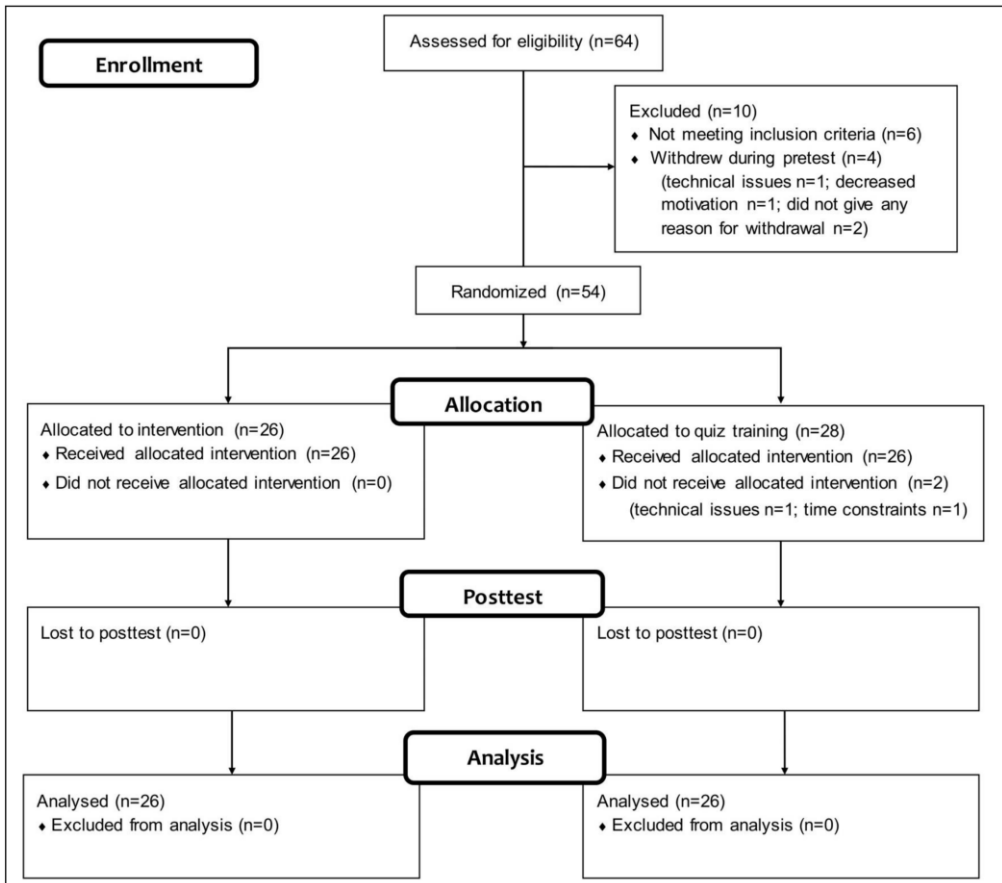

Figure 3. Study III: CONSORT flow diagram of the attrition rate at the different stages of the study.
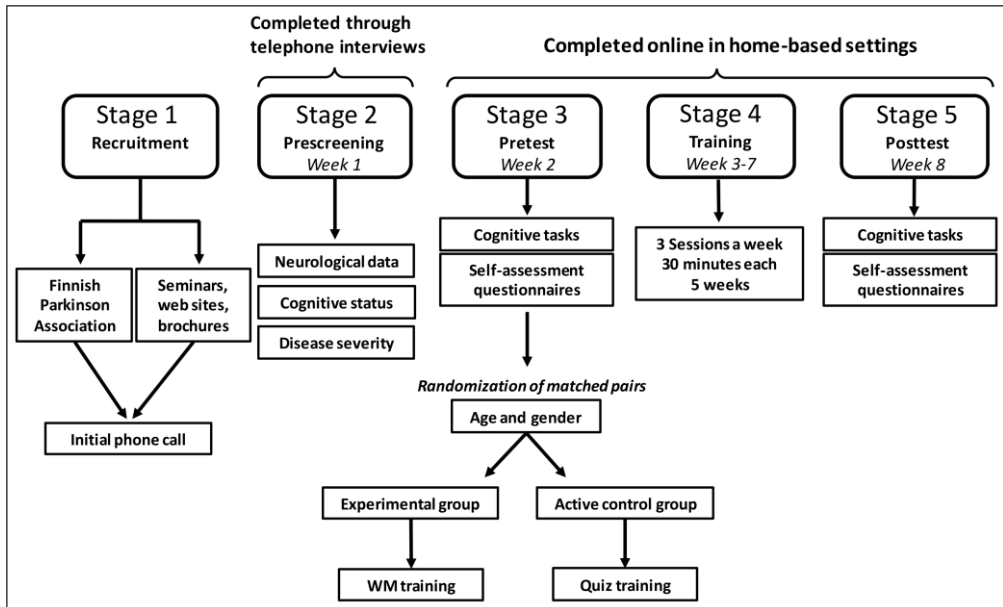
Figure 4. Study III: *An illustration of the five main stages in the study.*

Training paradigm

The WM training protocol comprised of three training tasks, two of which tapped WM updating, and one WM maintenance. One of the WM updating training tasks was the SUST paradigm used in **Study II**. The SUST task properties were almost identical to the one employed in **Study II**. The only difference was that we prevented adjacent words to be updated in **Study III** (i.e., there had to be at least one word between those words that were updated) while the updating stages in **Study II** were allowed to be updated irrespective of position.

The other updating training task was an n-back task (Kirchner, 1958). In the n-back task, the participants decide whether the currently displayed item is the same or not the same as the item presented *n* items earlier. In line with SUST, the n-back training was adaptive so that the level of *n* (i.e., how many items back one had to compare the current item with) was adjusted depending on how the participant performed. The third training task was a Forward span task (FSST) tapping WM maintenance. In this task, a sequence of items was presented on the computer screen in a serial order. The task was to recall the items in the same order they were presented. In line with the updating training tasks, the FSST was adaptive: the difficulty level was adjusted by increasing span length. In both the n-back training task and the FSST task, the training period encompassed three different stimulus types that were rotated from

session to session so that the same stimuli were practiced every third training session (1st session: letters, 2nd session: words, 3rd session: digits).

Tasks administered at pre- and posttest

In this study, we incorporated an extensive array of pre- and posttest tasks that were compiled into different composite scores based on their resemblance of the tasks that were practiced during the intervention period. Firstly, we employed pre- and posttest task versions of the trained tasks (coined as "Trained WM tasks") that were structurally similar and included the same stimuli that the WM training group practiced on. These three tasks were analyzed separately. Secondly, we administered pre- and posttest tasks that contained untrained stimuli but were otherwise identical to those that the WM training group practiced with. These three tasks were combined into a task-specific near transfer measure. Thirdly, we implemented a task-general near transfer composite score consisting of three WM tasks with different task paradigms than those that the WM training group practiced with. Lastly, we administered four far transfer measures tapping on other cognitive domains than WM. Two of these tasks (a sentence recall task, and a word list recall task) formed a verbal episodic memory composite. We also employed a cognitive interference measure (a computerized Stroop task), and an inhibition measure (Continuous performance task), and these two tasks were combined into an executive functioning and attention composite score.

Besides the computerized tasks, we also administered self-assessed pre-post questionnaires to examine whether the participants experienced subjective changes in WM functioning, executive functioning, or depressive symptoms following the training period.

Trained WM tasks

Selective Updating of Sentences task (SUS). This task was identical to the SUS task employed in **Study I and II.**

N-back task with digits. In the pre-post version of this training task, the participants were to perform two fixed n-back blocks, one with a 1-back sequence, and another with a 2-back sequence. The order of items in the sequences was randomized. Both n-back sequences included 48 trials, out of which 16 were targets and 32 non-targets. Sixteen of the non-targets were so-called lures (8 n+1, 8 n-1 lures). Each trial in a block was displayed for 1000 ms followed by a blank screen displayed for 450 ms.[2]

---

[2] In this thesis, all training and pre-posttest versions of the n-back paradigm employed the same stimulus display time (1000 ms) and the interstimulus interval (450 ms).

As the dependent variable, we used the d-prime accuracy measure calculated following Stanislaw and Todorov (1999).

Forward digit span task. In this task, the participants were presented with sequences of digits, one digit at a time (exposure time = 1000 ms), and the task was to recall the digits in the same order they were presented. The participants completed seven trials (list lengths ranged from three to nine items) in a randomized order, with one trial of each list length. The total number of correctly recalled digits from all the sequences served as the dependent variable.

Task-specific near transfer tasks

Selective Updating of Digits (SUD). Following Murty et al. (2011), we administered a WM updating task coined as Selective updating of digits (SUD). Here, five digits ranging from 1-9 were displayed on the computer screen in a row of five boxes (presentation time 4000 ms). The participants were to memorize the digit sequence, after which it disappeared, and a new row of five boxes (i.e., an updating stage) occurred on the screen (presentation time 4000 ms). In the updating stage, two of the new boxes contained digits, while the rest of the boxes were empty. The task was to replace the old digits with the recently shown digits in the memorized sequence, while simultaneously maintaining the unchanged digits from the initial digit sequence. At the end of each trial, one was to recall the whole sequence that included the most recent updated digits. The SUD task included altogether four baseline trials (i.e., no updating stages), four trials with two updating stages, four trials with three updating stages, and four trials with five updating stages. The trials were presented in a randomized order. In the analyses, we used the percentage of the total number of correctly recalled digits (in correct order) in the updating trials as the dependent variable.

N-back task with colors. The properties of this task were identical to the n-back task with digits, the only exception being that the stimuli in this task were colors instead of digits.

Forward color span. The Forward color span task was identical to the Forward digit span, except that the stimuli were color patches instead of digits.

Task-general near transfer tasks

Running memory with digits. Following Pollack et al. (1959), we administered a Running memory task. Here, participants were to memorize the last 4 digits of stimulus lists of unpredictable length (list lengths ranged from 4-11 digits). The participants completed one trial per list length (i.e., altogether eight trials), the order of which was randomized. The digits were displayed one at a time on the screen for 1000 ms. At the end of each list, the participants were to report the last four items in

the exact order in which they had been presented. The dependent variable was the total number of correctly recalled digits in correct position.

Alphabet working memory task. This task was identical to the one used in **Study I and II.**

Minus 2 span task. This task was identical to the one used in **Study I and II**.

Far transfer tasks

Sentence recall. This task was identical to the one used in **Study I and II**.

Word list recall. In this task, 10 words were displayed one at a time for 1000 ms. The task was to memorize each word in correct serial order, and finally recall them by typing in the words in empty columns. The participants completed altogether three trials, with each trial consisting of the same 10 words but in a different randomized order. The dependent variable was a true recall score, that is, the total number of correctly recalled words in the three trials minus repetitions, perseverations and additions.

Continuous performance task. The present version followed the task properties of Conners et al. (2003). Here, the participants were to press the spacebar when any letter except "X" was displayed on the screen. Altogether 360 letters were presented one at a time, each letter for 250 ms. These letters were presented in 18 consecutive blocks of 20 trials. However, the blocks had different interstimulus intervals (ISIs), namely 1, 2, or 4 seconds. The ISIs were block-randomized so that all three ISI conditions occurred in every three blocks but in a different order. The percentage of trials with letters other than "X" was 90%, and this percentage was constant across all blocks. As the outcome variable, we used the rate of commission errors from the 18 consecutive blocks, that is, the number of times when the participant pressed the spacebar when the letter "X" was presented.

Stroop. As the fourth far transfer measure, we adminstered a computerized Stroop task (Bartsch & Kothe, 2016; Salo, Henik, & Robertson, 2001; Zysset, Müller, Lohmann, & Von Cramon, 2001). Here, the participants were shown with two stimulus rows on a computer screen. In the upper row, a word (or, in the neutral condition, a series of x) was writtten in either red, green, blue, or yellow ink. In the lower row, there was a color name written in black ink. The task was to judge whether the ink color of the upper stimulus matched the color name below. Altogether six blocks were administered: two blocks included neutral conditions, two blocks congruent conditions, and two blocks incongruent conditions. As the dependent variable, we used the interference score, calculated as Neutral mean RT minus Incongruent mean RT.

Self-assessment measures

A set of self-report questionnaires tapping apathy, working memory, executive functioning, and mood were implemented at pre- and posttest. These were used to assess possible subjective changes following the intervention period.

Working Memory Questionnaire. At pre- and posttest, the participants were asked to complete the Working Memory Questionnaire (WMQ; Vallat-Azouvi, Pradat-Diehl, & Azouvi, 2012). This questionnaire includes three subscales, namely short-term storage, attention, and executive functioning. It contains 30 questions, with each question being rated on a 0-4 point Likert scale (0 = "no problem at all", 4 = "very severe problems in everyday life"). In the statistical analyses, the dependent variable was the sum score of the three subscales (score range: 0-120), with higher scores indicating more difficulties.

The Behavior Rating Inventory of Executive Functioning. To measure potential pre-post changes in self-estimated executive functioning, we employed the Behavior Rating Inventory of Executive Functioning (BRIEF-A; Roth, Isquith, & Gioia, 2005). It includes 75 questions divided into nine non-overlapping subscales, namely inhibition, shifting, emotional control, self-monitoring, initiation, working memory, plan/organize, task monitor, and organization of materials. As the dependent variable in the analyses, we calculated the sum score (t-score) of the aforementioned nine subscales.

Geriatric Depression Scale-30. Potential pre-post changes in depressive symptoms were evaluated with GDS-30 (Yesavage et al., 1983). Higher scores represent a higher degree of depression (range from 0 to 30).

Statistical analyses

Normality of data was tested using the Kolmogorov-Smirnov test. For baseline group differences, independent samples t-tests were conducted on continuous variables that were normally distributed, and Mann-Whitney U-tests on variables that were non-normally distributed. For the categorical variables, we used the χ^2 test or the Fisher's exact test. Motivation and alertness during training were analyzed using 3 x 2 mixed-model ANOVAs with time (start, halfway, end) and group (WM training group, active control group) as factors. Using growth curve analyses (Mirman, 2014), we examined the learning curves during the training period separately for each of the three WM training tasks. In the models, time served as the fixed effect (the fixed effect was coded as a linear contrast), comprising of the average performance over five training weeks (i.e., altogether five time points). To account for the variability between the participants, subject on the intercept served as the random effect in the models. In line with **Study II**, the pre-posttest changes between the groups were analyzed with ANCOVAs where posttest performance served as the dependent variable, group (WM

training vs. active controls) as the between-subjects factor, and pretest performance as the covariate.

4.4 Study IV

For each participant enrolled, the whole study took five weekdays. All participants had a pretest session on Monday, after which they were block-randomized either to an n-back training group that was given an explicit visuospatial strategy, an active control group training on the same n-back task but without any strategy instructions, or a passive control group that did not perform any training. The training sessions for the two training groups were administered in home-based settings, and had to be completed on either Tuesday, Wednesday, or Thursday. At Friday, all participants took part in a posttest session. The pre-and posttest assessments were administered as computer class sessions. For an illustration of the study design and the attrition rate, see Figure 5.
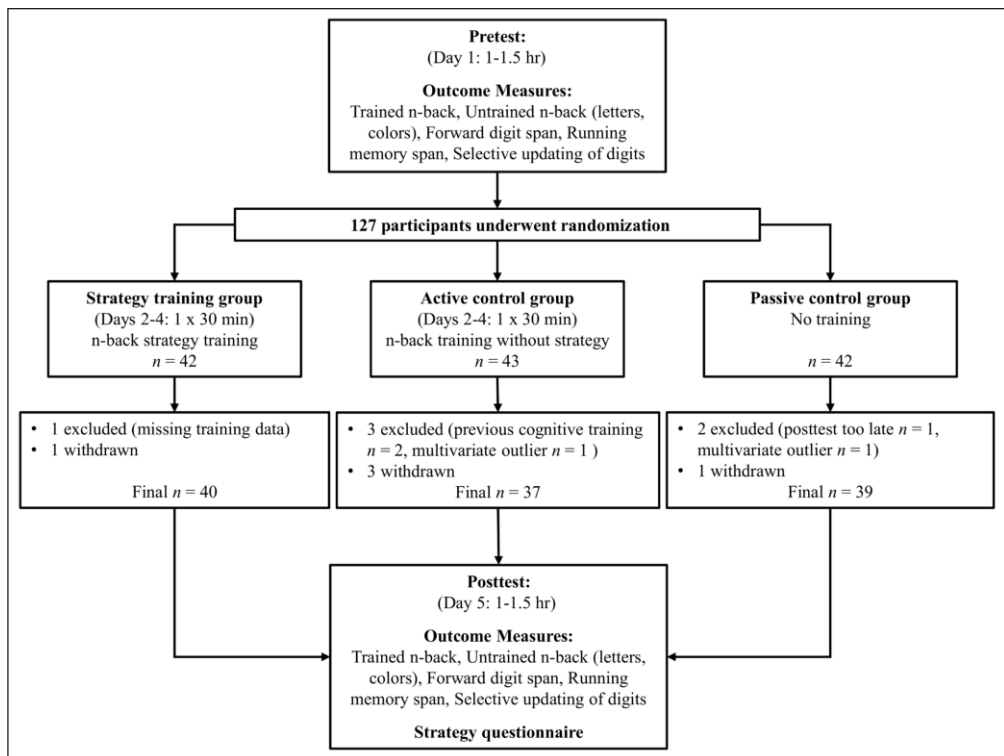

Figure 5. Study IV: *Flowchart of the attrition rate at the different stages of the study.*

Training paradigm

The training task comprised of one session of the commonly used adaptive n-back training paradigm using digits as stimuli. The primary reason for implementing n-back training in **Study IV**, in contrast to the SUS and Reading span training in **Studies II-III**, was the invention of a new visuospatial strategy specifically for this task paradigm. Both groups completed altogether 20 n-back blocks. In each block, six trials were targets while 14 trials were non-targets. The only difference between the groups was that one group received the visuospatial strategy instruction before each n-back block.

Tasks administered at pre- and posttest

The pre- and posttest battery included an adaptive n-back task with digits (i.e., the trained WM task), and two untrained n-back tasks using letters and colors as stimuli. Moreover, we included three untrained WM tasks, namely the Selective updating of digits (SUD) task, the Forward digit span task, and the Running memory task.

Trained n-back task. This task served as the criterion task in the pre- and posttest battery, as it was identical to the adaptive n-back task used in training. The only exception was that the pre- and posttest version included 10 n-back blocks instead of the 20 blocks used in training. As the dependent variable, we used both the maximum level on n achieved, and the average level of n achieved.

Untrained n-back tasks. Task specific-near transfer effects were measured with two untrained n-back tasks, one with letters and the other one with colors as stimuli. Both were non-adaptive n-back tasks, including one block of a fixed 2-back sequence, and one block of a fixed 3-back sequence. The task properties of the n-back sequences were identical to those employed in **Study III**. We used two dependent variables in the analyses. For measuring accuracy, we employed the d-prime (Stanislaw & Todorov, 1999). For reaction time, we used the mean reaction time (RT) on correct target responses.

Selective updating of digits (SUD). This task was similar to the SUD task employed in **Study III,** apart from two adjustments to ensure adequate task difficulty for the healthy young adults. The first adjustment concerned the presentation time of the updating stages that was fixed to 2000 ms in **Study IV**. The second adjustment pertained the trial sequences. In **Study IV**, the participants completed 10 baseline trials with no updating stages, and 10 trials with three updating stages. As the dependent variable, we used the percentage of correctly recalled items in the updating trials.

Forward digit span task. This task was similar to the digit span task administered in **Study III**, except that the list lengths (ranging from 4-10 items) were adjusted to ensure suitable task difficulty for the present healthy adults.

Running memory task. This task was identical to the Running memory task administered in **Study III.**

Statistical analyses

To examine whether the three groups were comparable at baseline, we employed one-way ANOVAs for the continuous variables age and education length, and the χ^2 test for the categorical variable gender. To examine improvements in the pre- and posttest tasks, we employed ANCOVA analyses with posttest performance serving as the dependent variable, pretest performance serving as the covariate, and group as the fixed factor. Those ANCOVA analyses that revealed a main effect of group were further followed up with planned contrasts. Here, we conducted two contrasts; the first one comparing the strategy training group to the active control group (-1, 1, 0), and the other one comparing the strategy training group to the passive controls (-1, 0, 1).

To investigate the role of spontaneous strategy use on n-back performance, we analyzed posttest strategy reports in the two control groups. Using a simple linear regression analysis, we first probed for whether the level of detail in self-generated strategies predicted n-back task performance. Through ANOVA analyses, we further investigated whether the type of strategy was related to n-back performance. Here, the strategy types were classified as No strategy, Rehearsal, Updating, Grouping, Grouping and comparison, and Other. In the ANOVAs, the No strategy group served as the baseline against which the other strategy groups were compared.

Table 2
Summary table of the cognitive tasks employed at pre- and posttest.

Tasks	Cognitive function	I	II	III	IV
		Study			
Selective updating of sentences	WM updating	x	x	x	
Selective updating of digits	WM updating			x	x
n-back with digits	WM updating			x	x
n-back with letters	WM updating				x
n-back with colors	WM updating				x
Running memory task	WM updating			x	x
Alphabet WM task	WM updating	x	x	x	
Minus 2 span task	WM updating	x	x	x	
Reading span task	Complex WM	x	x		
Operation span task	Complex WM	x	x		
Symmetry span task	Complex WM	x			
Forward digit span task	WM maintenance			x	x
Forward color span task	WM maintenance			x	
Sentence recall in Finnish	Episodic memory	x	x	x	
Sentence recall in English	Episodic memory	x	x		
Paragraph memory	Episodic memory	x	x		
Word fluency	Episodic memory	x	x		
Word list recall	Episodic memory			x	
Continuous performance task	Attention			x	
Stroop task	Interference			x	

5 RESULTS

5.1 Study I

In Experiment 1, the to-be-analyzed sample consisted of 169 participants as one participant was excluded for being a multivariate outlier. The results of Experiment 1 showed that our novel sentence-level WM updating task, coined as SUS, exhibited adequate internal consistency ($α = .740$). Next, we analyzed the effects of the of task manipulations in the SUS task. The task difficulty was manipulated as a function of increased number of updating stages (range: 2, 3, or 5 updates) and sentence length (range: 4-7 words). Two-way repeated measures ANOVA revealed a main effect of updating stage, indicating that the performance worsened as a function of number of updates. We also found a main effect of sentence length, indicating decreased performance as a function of an increased sentence length. In addition, ANOVA revealed a statistically significant updating stage × sentence length interaction, mainly stemming from a disproportionate performance decrease for the most demanding condition with 5 updates and 7-word sentences (see Figure 6A).

The concurrent validity of the SUS task was examined by investigating its intercorrelations against three complex span tasks (CSTs). The intercorrelations, illustrated in Figure 6B, showed that the SUS task correlated statistically significantly with all CSTs, indicating that the SUS paradigm was tapping on WM.
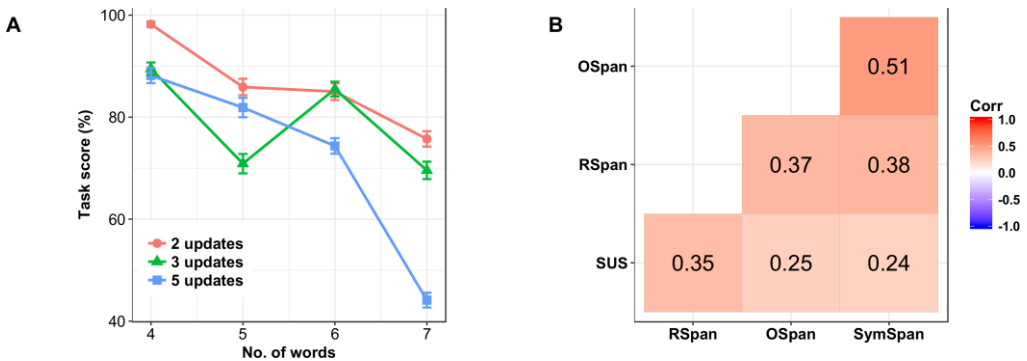

Figure 6. Study I, Experiment 1: (A) *Mean task performance as a function of updating stages and sentence length in the the Selective Updating of Sentences task in Experiment 1. Error bars represent standard errors of mean.* ***(B)*** *Intercorrelations for all variables in Experiment 1. All correlations were statistically significant (p < .01). SUS = Selective Updating of Sentences; RSpan = Reading span; OSpan = Operation span; SymSpan = Symmetry span.*

In line with Experiment 1, the internal consistency of the SUS task was adequate in Experiment 2 ($α = .775$). Also the task manipulation effects corresponded to those

in Experiment 1. Here, ANOVA revealed a main effect of updating stage, a main effect of sentence length, and an updating stage × sentence length interaction (see Figure 7A). Furthermore, the SUS task showed a five-week test-retest reliability of $r = .707$, indicating adequate external reliability (see Figure 7B).

The intercorrelations between the tasks in Experiment 2 are shown in Figure 7C. As expected, the SUS task correlated statistically significantly with all WM tasks, with the strongest correlations being observed between the SUS task and the Reading span task. Moreover, the intercorrelations showed that the SUS task, the CSTs, and the Minus 2 span task were most strongly related to the verbal episodic memory tasks. However, the SUS task was the only WM task that showed a positive correlation of a moderate strength ($r = .54$) with the Sentence recall in Finnish. Finally, we conducted hierarchical regression analyses to investigate whether the SUS task predicted performance on the verbal episodic memory tasks over and beyond the CSTs. The results showed that the SUS task shared unique variance with the Sentence recall in Finnish (see Figure 7D) and in English, but not with Paragraph recall or Paragraph semantic recall.
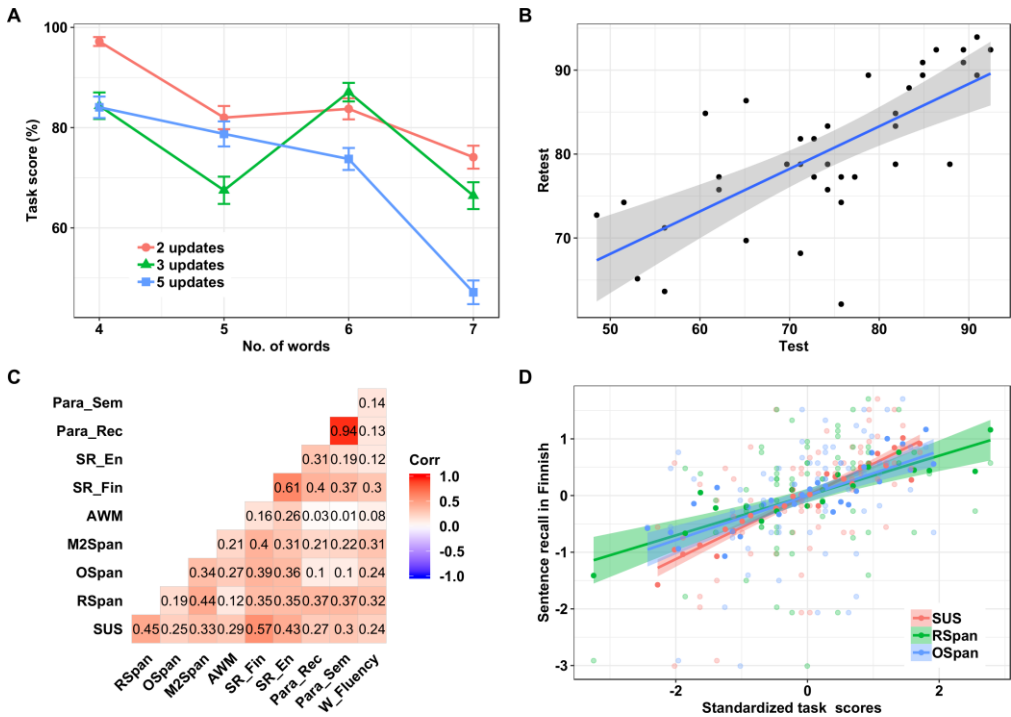
Figure 7. Study I, Experiment 2: (A) *Mean task performance as a function of updating stages and sentence length in the Selective Updating of Sentences (SUS) task in Experiment 2.* ***(B)*** *Test-retest reliability of the SUS task. Grey shaded regions represent 95% confidence intervals on the slope.* ***(C)*** *The intercorrelations for all variables in Experiment 2. All correlations were statistically significant (p < .01). SUS = Selective Updating of Sentences; RSpan = Reading span; OSpan = Operation span; M2Span = Minus 2 span task; AWM = Alphabet working memory task; SR_Fin = Sentence recall in Finnish; SR_En = Sentence recall in English; Para_Rec = Paragraph recall; Para_Sem = Paragraph semantic recall; W_Fluency = Word fluency.* ***(D)*** *Fitted slopes of the SUS task, the Reading span task, and the Operation span task on the Sentence recall in Finnish. X-axis represents standardized scores on the WM tasks, and Y-axis represents standardized scores of the Sentence recall in Finnish. The colored dots depict participants' performance separately for each WM task. Colored regions represent 95% confidence intervals.*

5.2 Study II

Figure 8 illustrates the improvements in the SUST task and the SRST task for the WM training group during the training period. The progression curve in the SUST task (Fig. 8A) shows a steady rate of improvement during the first 12 training sessions, after which performance stabilizes for the last four training sessions. In the SRST task (Fig. 8B), on the other hand, there is less improvement and more variability in

performance across the 15 training sessions. Here, a ca 20% higher span performance was reached during the training period, as compared with the span performance during the first training session. One-way repeated measures ANOVA showed a significant main effect of session in both the SUST task and the SRST task, indicating an improved performance over time in both training tasks.
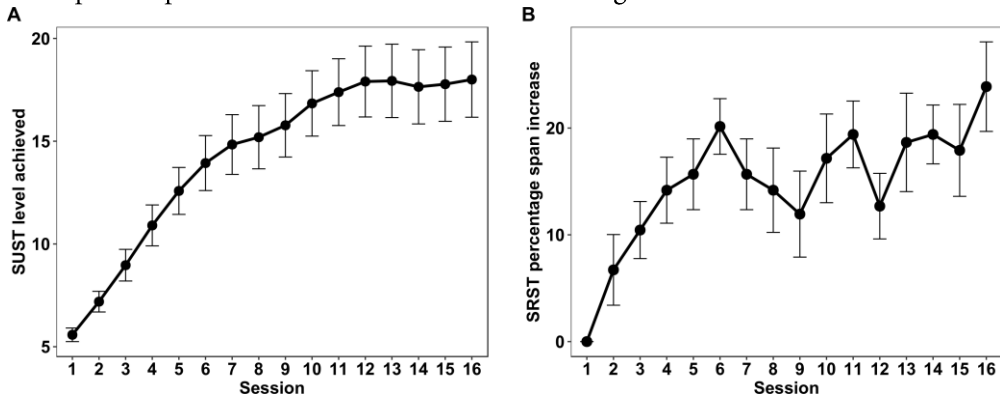

Figure 8. Study II: Training improvement in the SUST task (A), and the SRST task (B) in the WM training group. Error bars represent standard error of means.

The slopes of improvements in each pre- and posttest task are illustrated in Figure 9. The results from ANCOVAs on the criterion tasks (i.e., the trained WM tasks) showed no significant group differences at posttest (see Fig. 9A and 9B). However, when excluding one close-to-be multivariate outlier in the WM training group, the results in the SUS task became statistically significant, with the WM training group showing a higher accuracy at posttest compared with the active control group. The results of the ANCOVAs on the near transfer tasks showed no significant main effect of group in the Operation span task (see Fig. 9C) or in the Minus 2 span task (see Fig. 9D). However, albeit not reaching statistical significance, there was a trend towards a main effect of group in the Alphabet working memory task (see Fig. 9E), with the WM training performing slightly better at posttest compared with the active control group. As regards the far transfer tasks, no main effect of group was seen in the Sentence recall in Finnish (Fig. 9F), Sentence recall in English (Fig. 9G), Paragraph recall (Fig. 9H), Paragraph semantic recall (Fig. 9I), or the Word fluency task (Fig. 9J).
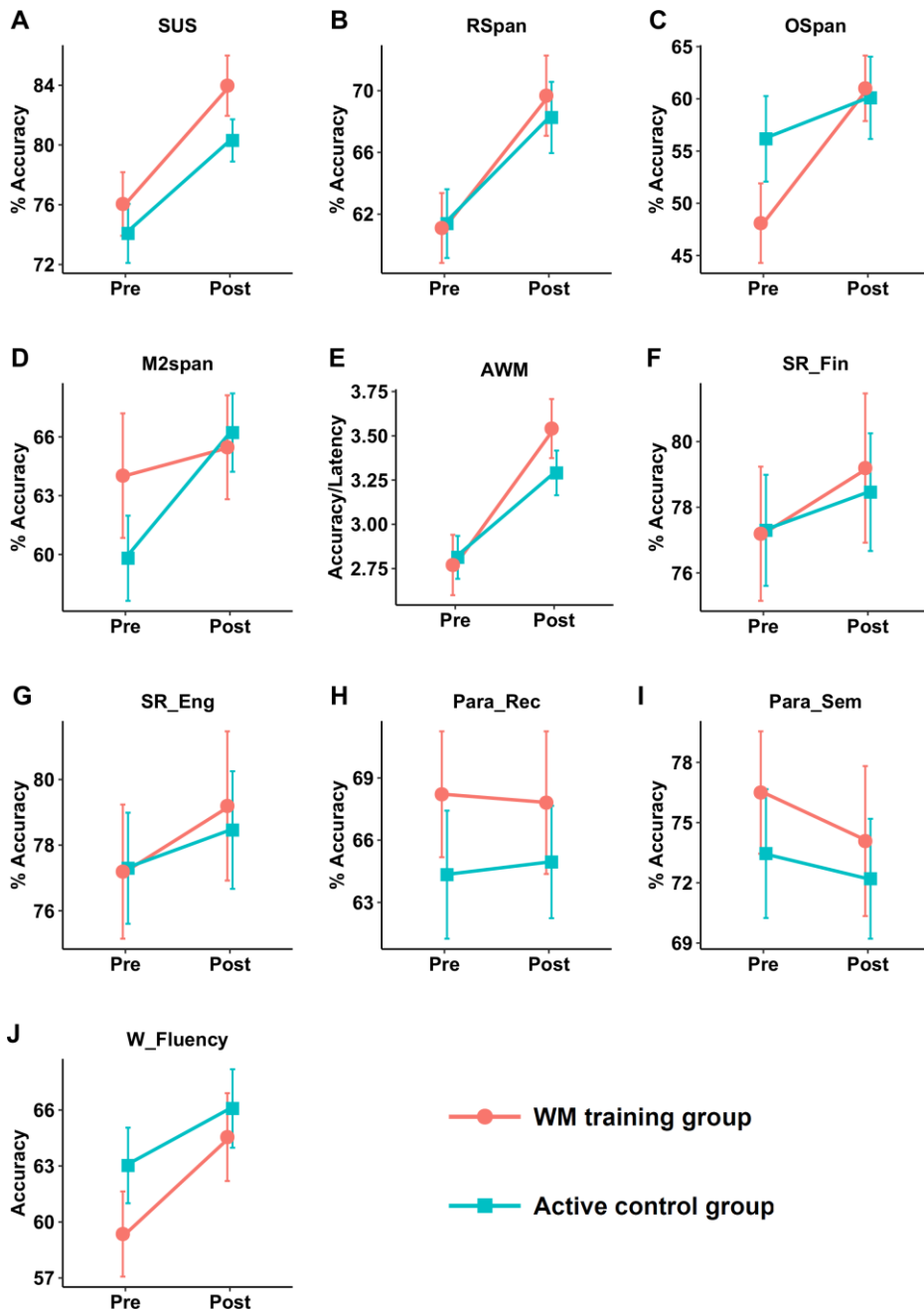
Figure 9. Study II: Pre-post means per group for the cognitive tasks. (A) Selective updating of sentences task; (B) Reading span (RSpan) task; (C) Operation span (OSpan) task; (D) Minus 2 span (M2span) task; (E) Alphabet WM (AWM) task; (F) Sentence recall (SR_Fin) in Finnish; (G) Sentence recall (SR_Eng) in English; (H) Paragraph recall (Para_Rec); (I) Paragraph semantic recall (Para_Sem); (J) Word Fluency (W_Fluency). Error bars represent standard error of mean.

5.3 Study III

The training improvements in the three WM tasks over the 15 training sessions are illustrated in Figure 10. In the n-back task (see Fig. 10B), a steady rate of improvement was seen during the training period, with the learning curve becoming more stable towards the end. In the SUST task (see Fig. 10A), the learning curve showed a steady rate of improvement over the 5-week training period, albeit the inter-individual variability increased as the training period progressed. Lastly, in the FSST task (see Fig. 10C), the learning curve was steeper at the onset of training, after which it stabilized towards the end of the training period. The growth curve analyses also showed that there was a significant performance improvement over time in all training tasks.
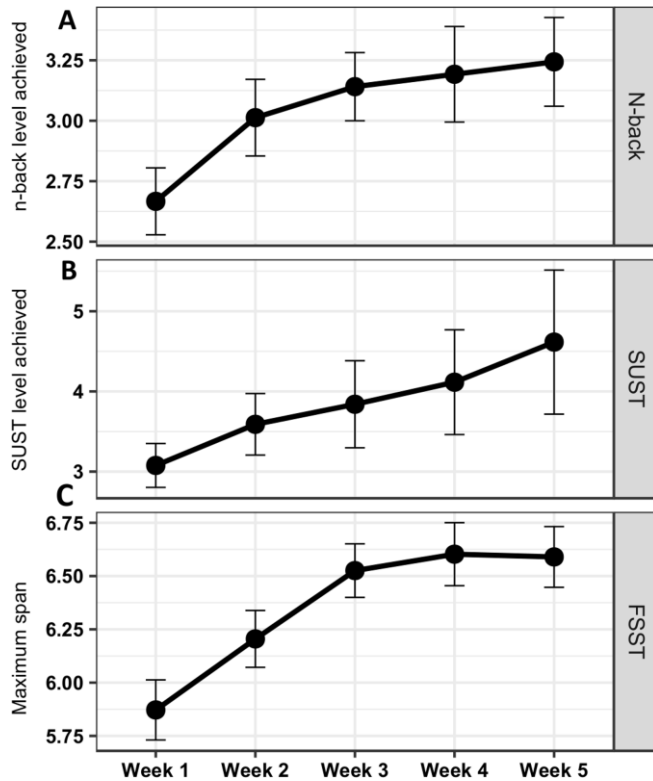

Figure 10. Study III: Training improvement for the WM training group during intervention. (A) Level increases in the SUST task over the 5-week training period. (B) Level increases in the n-back training task over the 5-week training period. (C) Maximum span performance in the FSST task over the 5-week training period. Error bars represent standard error of means.

The mean performance in the pre- and posttest tasks for the WM training group and the active control group are illustrated in Figure 11. ANCOVAs on the criterion tasks showed a significant main effect of group in the 1-back task with digits (see Fig. 11A), 2-back with digits (see Fig. 11B), and the SUS task (see Fig. 11C), indicating that the WM training group performed more accurately in these tasks at posttest compared to the active control group. However, no main effect of group was seen at posttest in the trained task tapping WM maintenance (see Fig. 11D). ANCOVA on the task-specific transfer composite score showed a main effect of group (see Fig. 11E), with the WM training group performing significantly better at posttest compared with the active control group. No main effect of group was observed in the task-general near transfer composite score (see Fig. 11F), or in the far transfer composite scores (se Fig. 11G, and Fig 11H).

As regards the self-assessment measures, ANCOVA did not reveal any main effects of group in the WMQ or the BRIEF-A. However, on the GDS-30, ANCOVA showed a main effect of group at posttest (See Fig 11I), with the WM training group reporting decreased depression scores compared with the active controls.
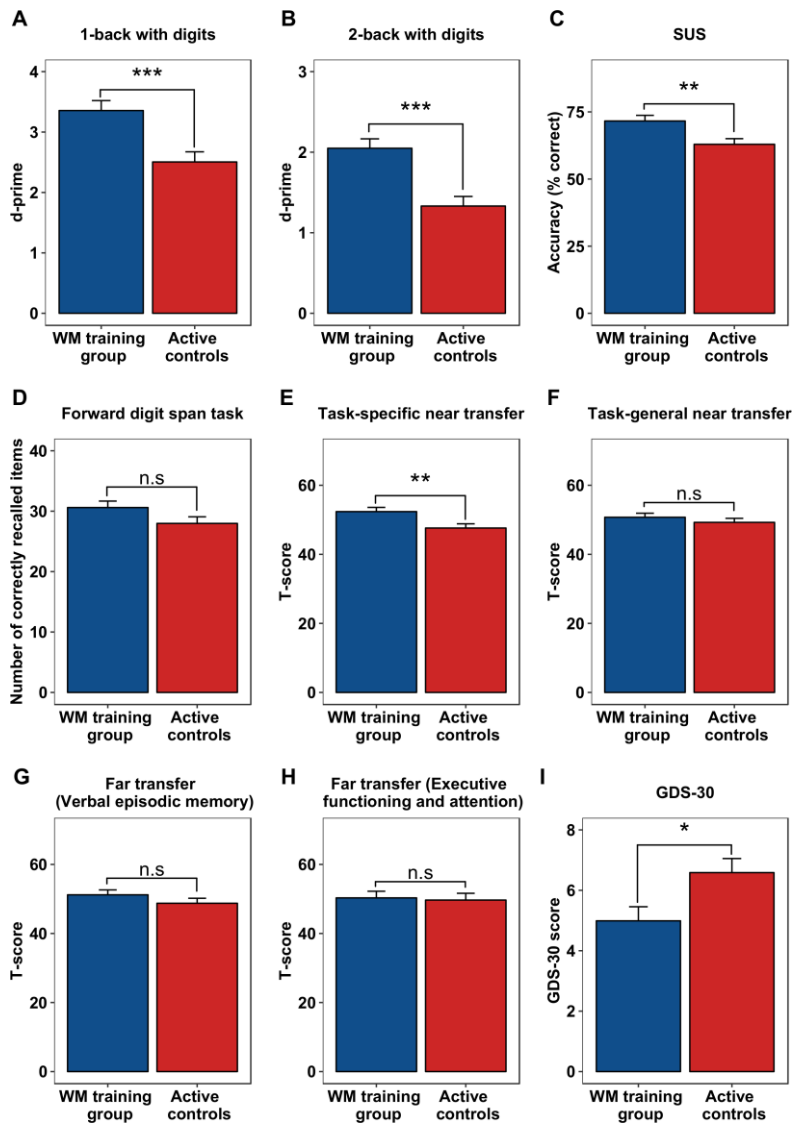
Figure 11. Study III: *Posttest performance (adjusted for pretest performance in ANCOVA) of the WM training group (represented in blue bars) and the active control group (represented in red bars).* ***(A)*** *1-back task with digits;* ***(B)*** *2-back task with digits;* ***(C)*** *Selective updating of sentences task;* ***(D)*** *Forward digit span task;* ***(E)*** *Task-specific near transfer composite score;* ***(F)*** *Task-general near transfer composite score;* ***(G)*** *Verbal episodic memory composite score;* ***(H)*** *Attention and executive functioning composite score, and* ***(I)*** *Geriatric Depression Scale-30 (GDS-30). n.s indicates not significant; * indicates p < .05; ** indicates p < .01; *** indicates p < .001. Error bars represent standard error of means. The y-axis for the composite scores (i.e., panels E-H) illustrates t-standardized scores. In panel I, higher scores indicate higher rates of depression.*

5.4 Study IV

For the criterion task (the trained adaptive n-back task with digits), using the maximum n-back level achieved as the dependent variable, ANCOVA revealed a significant main effect of group. Planned contrasts showed that the strategy training group obtained a significantly greater adjusted posttest mean score compared to those of the active control group and the active control group (see Fig. 12A). In line with the maximum n-back level achieved, ANCOVA showed a main effect of group when employing the average n-back level achieved as the dependent variable. Further examination with planned contrasts revealed a significantly higher posttest mean for the strategy training group compared to those of the active and passive controls.

In the untrained n-back task with letters using d-prime (i.e., accuracy) as the outcome measure, ANCOVA revealed a main effect of group in the more demanding 3-back condition. Planned contrasts showed that the strategy training group reached a significantly higher adjusted posttest mean accuracy compared with the active and passive controls (see Fig. 12B). As regards the accuracy rates in the untrained n-back task with colors, ANCOVA showed a main effect of group in the 3-back condition. Accordingly, planned contrasts showed that the adjusted posttest mean accuracy was significantly higher in the strategy training group compared to those of the active and passive controls (see Fig. 12C). ANCOVA did not show any significant main effect of group in the Forward digit span task (see Fig. 12D), the Running memory task (see Fig. 12E), or the SUD task (see Fig. 12F).

Regarding the relationship between spontaneously generated strategies and WM performance, the simple regression analysis showed that the level of detail in the two control groups explained 45.6% of the variance in n-back performance at posttest (see Fig. 13A). Moreover, ANOVAs on the strategy type showed that the participants using Grouping and comparison outperformed those not using any strategies in all n-back tasks. Also other strategies were related to better n-back performance. For example, the participants using Updating performed significantly better in the digit and letter n-back tasks than the No strategy group, whereas the ones using Grouping outperformed the No strategy group in the digit n-back task (see Fig. 13B, 13C, and 13D).

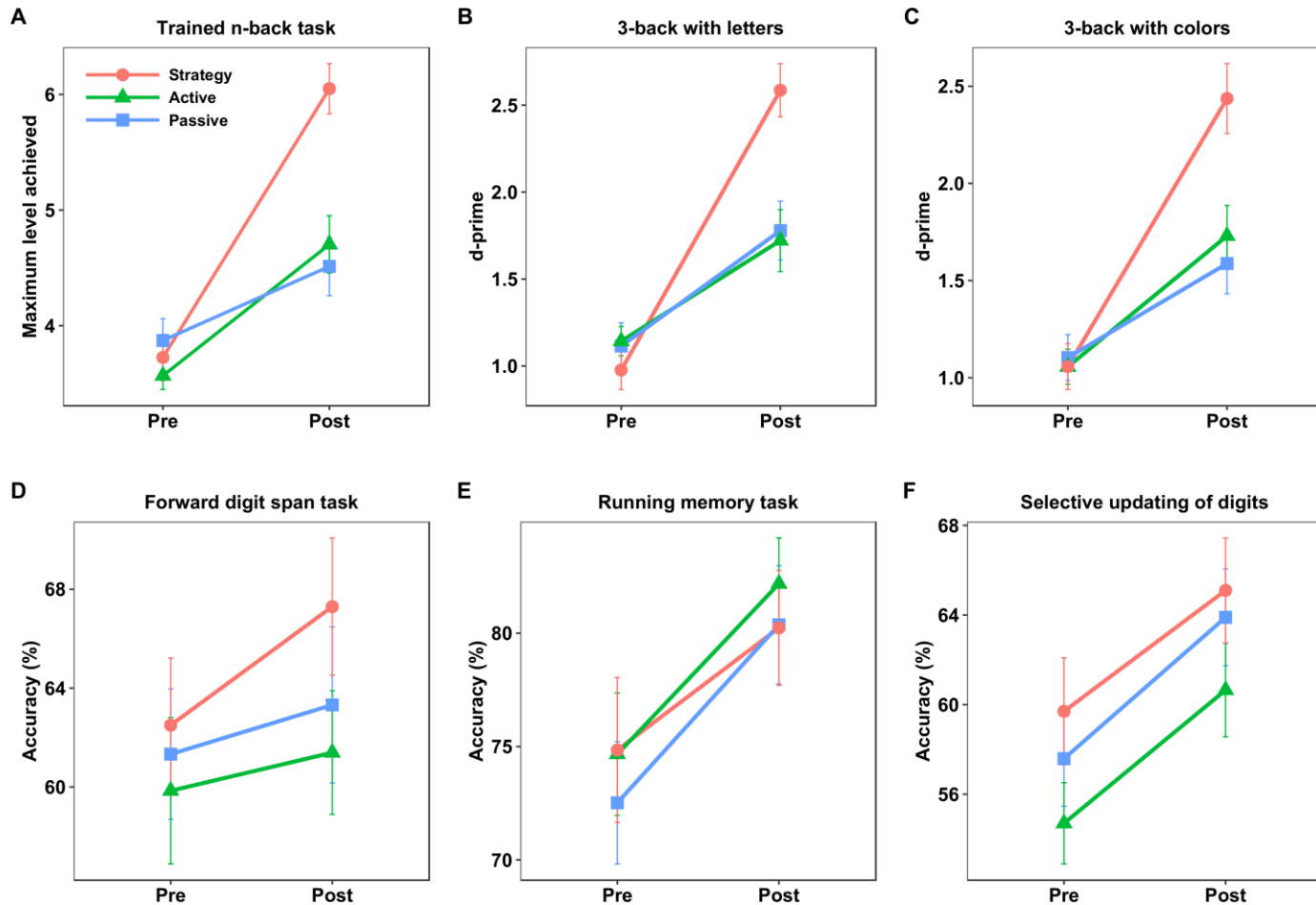**Figure 12. Study IV:** *Mean performance at pre- and posttest for the strategy training group (red lines), active control group (green lines), and the passive control group (blue lines). **(A)** The trained n-back task with digits; **(B)** 3-back with letters; **(C)** 3-back with colors; **(D)** Forward digit span task; **(E)** Running memory task; **(F)** Selective updating of digits. Error bars represent standard error of means.*
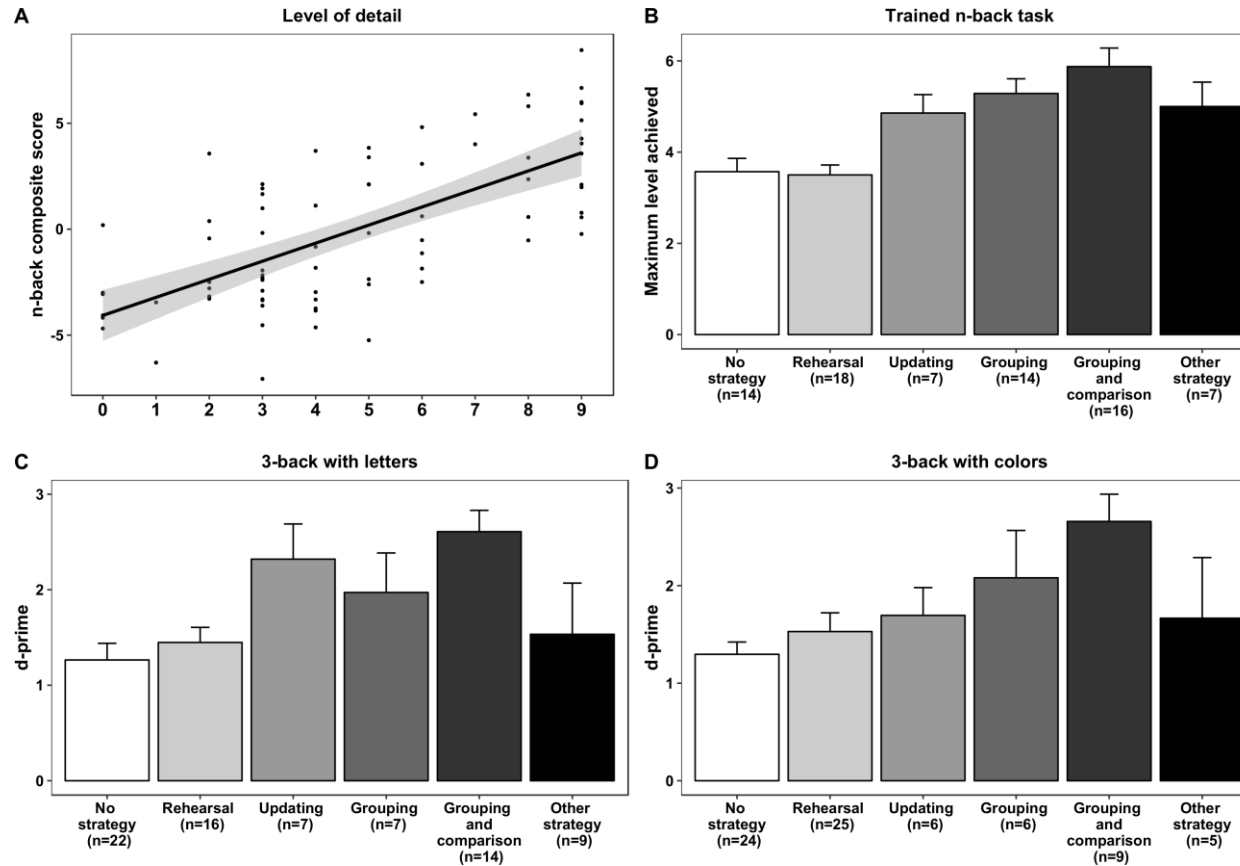
Figure 13. Study IV: (A) *The relationship between level of detail in reported strategies at posttest (x-axis) and n-back performance at posttest (y-axis). The level of detail was summed up from all three n-back tasks (range 0-9). The n-back composite score comprised of the z-transformations of the average and maximum n-back level reached in the trained n-back task with digits and the d-prime values from the two untrained 3-back tasks. The grey shaded areas represent 95% confidence intervals for the slope.* **(B)** *Posttest performance by strategy type subgroups in the trained n-back task with digits,* **(C)** *in 3-back with letters, and* **(D)** *in 3-back with colors. The whiskers in 8B, 8C, and 8D illustrate standard error of means.*

6 DISCUSSION

The present thesis had two main aims. The first aim was to examine the effects of core training targeting verbal WM. The focus on verbal WM was motivated by its pivotal role in human cognition, including verbal communication and language processing. The second aim was to address the cognitive mechanisms underlying WM training effects by testing the viability of the Strategy Mediation hypothesis in accounting for training-induced changes in WM. Here the results of the individual studies are discussed in more detail. This is followed by considerations on the Strategy Mediation vs WM Capacity hypotheses, as well as on methodological considerations, limitations, and future directions.

6.1 The Selective updating of sentences paradigm

Study I sought to develop a novel sentence-level WM task, coined Selective Updating of Sentences (SUS), which was motivated by concerns that most of the commonly used WM tasks include rather artificial materials (e.g., Holmes, 2011; Holmes & Gathercole, 2014; Klingberg, 2010; Moreau & Conway, 2014; Shipstead, Redick, & Engle, 2010). The SUS task operates with semantically and syntactically well-formed sentences, thus adding naturalness to the task. Two experiments provided evidence that the SUS task has adequate psychometric properties. Experiment 1 indicated that the SUS task exhibited adequate internal consistency, and that it demonstrated concurrent validity by correlating significantly with well-established WM tasks (three complex span tasks). Experiment 2 replicated the findings of Experiment 1, and further showed that SUS task performance was consistent over time by showing an adequate level of test-retest reliability. Moreover, it correlated positively with all verbal episodic memory tasks, and shared some unique variance with the Sentence recall tasks over and above the complex span tasks. **Studies II-III** showed that as such, the Selective Updating of Sentences Training (SUST) paradigm worked as a WM training task because steady performance increases during training were observed in both healthy younger adults as well as in patients with PD. Taken together, these findings indicate that the SUS task is an adequate task for measuring verbal WM processes, and is as such applicable to WM training in both healthy and clinical populations.

6.2 Limited transfer following verbal WM training

Study II investigated the transfer effects of home-based sentence-level core WM training in healthy young adults. The employment of a sentence-level training paradigm was motivated by a recent study (Payne & Stine-Morrow, 2017) indicating

far transfer effects with a partly similar training paradigm as in our study. The results of **Study II** showed no statistically significant near transfer effects to untrained WM tasks or far transfer effects to measures tapping on verbal episodic memory and word fluency. While the absence of far transfer was less surprising (e.g., Melby-Lervåg et al., 2016), more so was the total lack of near transfer effects as WM training typically yields improvements particularly in tasks that are based on the same task paradigm as the trained ones (Soveri et al., 2017). In our pre- and posttest battery, we administered three near transfer tasks, out of which two were task-general transfer measures (i.e., the Alphabet WM task, and the Minus 2 span task), and one a task-specific transfer measure (i.e., the Operation span task). Considering recent meta-analytical evidence (Soveri et al., 2017), one would have expected an improved performance at least in the Operation span task that shared the same task paradigm with one of the tasks practiced during training (i.e., the SRST paradigm). However, given the absence of group differences also in the criterion-related Reading span task at posttest, generalization to an untrained variant of the same task paradigm would have been doubtful and theoretically difficult to explain, because task-specific practice effects are considered to be necessary for transfer to occur (Shipstead, Hicks, & Engle, 2012).

Besides the scarce generalization effects, perhaps a more surprising finding was the weak criterion-related training effects. Here, the WM training group improved significantly only on the SUS task after excluding one participant that was close for being a multivariate outlier. The weak training effects cannot be explained by lack of improvement during the training period: here, both the SUST and the SRST paradigms showed increased performances on most of the training sessions. Neither can it be explained by motivational factors (which we controlled for), or expectations regarding the improvements at posttest as a function of training. In fact, the participants in the WM training group reported significantly higher subjective improvements in most of the pre- and posttest tasks compared to the active control group. Instead, we suggest that the lack of criterion-related training effects and the absence of transfer to task-specific WM measures might reflect the automaticity of WM systems related to sentence processing. Reading texts (i.e., via Internet, newspapers, or books) is for most people a natural part of everyday life, and the skills related to reading are for most literates highly overlearned following years of practice (Fischler & Bloom, 1980). Thus, practice with tasks tapping processes that are highly automatized might elicit less training effects compared with paradigms using more artificial stimulus sequences that occur less often in daily life.

Study III examined the feasibility and efficacy of home-based core training targeting verbal WM updating in patients with PD, and the extent of transfer to other tasks. This study was designed for two different reasons. The first one concerned the scarceness of previously conducted randomized controlled trials of WM training in

PD patients (Glizer & MacDonald, 2016; Walton, Naismith, Lampit, Mowszowski, & Lewis, 2017), leaving it open as to what extent this particular group would benefit from core training targeting verbal WM updating. The second reason was the lack of knowledge on the feasibility of administering cognitive training in its entirety at home without any face-to-face contact in PD patients (but see Edwards et al., 2013).

Regarding the improvements following training, the WM training group showed statistically significant gains on the two trained tasks tapping on WM updating, and a close-to-significant improvement on the Forward digit span task that tapped on WM maintenance. Moreover, statistically significant improvements were observed on the task-specific near transfer composite measure comprising of untrained variants of the trained WM tasks. No task-general near transfer was seen on the composite score containing WM measures with different task paradigms, or on the two far transfer composite measures tapping on verbal episodic memory as well as executive functioning and attention. Thus, the pattern of transfer effects observed in the present study concurs with the most recent relevant meta-analysis (Soveri et al., 2017), suggesting that training-related transfer in PD patients yields a similar transfer pattern to untrained WM tasks as has been observed in healthy young adults.

As regards the self-assessment measures, WM training did not yield any subjective changes in everyday executive functioning or WM functioning. However, we did find a significant post-training decrease in the depression scores for the WM training group. In a previous cognitive training study on PD by Peña et al. (2014), the results also indicated a nearly significant trend effect for lowered depression scores following training, and a recent meta-analysis suggests that cognitive training is associated with reduced depression rates in patients with major depressive disorder (Motter et al., 2016). Thus, the reduction in the depression scores observed in the present study is broadly in line with previous findings, suggesting that WM training might have some mood-related benefits on PD patients even in the presence of only limited cognitive transfer.

The fully home-based WM training regime in **Study III** proved to be feasible to PD patients. Except for some issues regarding the technical aspects of our experiment platform, such as instructions how to access and login to their individual accounts, the enrolled PD patients were capable to quite independently complete the pre- and posttests as well as the training program without any greater assistance by telephone or email. Moreover, the high rate of motivation (WM training group $M_{motivation}$ = 4.2/5.0, active control group $M_{motivation}$ = 3.9/5.0), the low dropout rate (only two participants from the active control group withdrew after being enrolled to training), the several requests for participating in a possible follow-up study, and the positive feedback we received retrospectively, suggest that home-based training is an easily

administered and a highly cost-effective alternative for future cognitive training studies in clinical contexts such as in patients with PD.

6.3 Task-specific transfer following a single session of strategy-based WM training

The pre- and posttest analysis in **Study IV** that probed for the effects of an externally provided strategy showed that it led to a significantly higher n-back level in the trained n-back task at posttest compared to the active and passive controls. Moreover, compared to the two control groups, the strategy training group showed higher accuracy rates in the 3-back task with letters, and the 3-back task with colors following training. The two untrained n-back tasks also enabled reaction time measures, but here no group differences were seen at posttest. As regards the untrained structurally different WM tasks, no group differences were seen on either the SUD task, the Forward simple span task, or the Verbal running memory task. Hence, the transfer effects obtained in the present study after just 30 minutes of supervised adaptive WM practice echo the ones typically seen after 4-6 weeks of unsupervised adaptive WM training (Soveri et al., 2017).

The role of self-generated strategies was investigated by examining whether the level of detail and the strategy type of the control participants' self-reported strategies at posttest were related to WM performance in the n-back tasks at posttest. The results showed that the level of detail in self-reported strategies explained almost half of the variance in n-back performance at posttest. Moreover, the strategy type analyses indicated that several self-generated strategies (especially strategies similar to the one provided for the strategy training group) were related to higher task performance on the n-back tasks, irrespective of task stimuli. This finding is in line with previous WM strategy studies (e.g., McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003), and highlights the impact efficient strategies have on WM performance, be they external or internal.

6.4 The WM Capacity hypothesis vs. the Strategy Mediation hypothesis

Over the past decade or so, WM training and its generalization to untrained tasks have stirred considerable interest. As the aim with WM training has been to elicit transfer to other cognitive domains, most of the studies have relied on the WM Capacity hypothesis either explicitly or implicitly, presupposing that WM can be expanded with *core training* that entails an intensive practice period with adaptive and challenging WM tasks close to individual maximum performance (Shipstead et al., 2012). Since WM is related to a wide range of other cognitive abilities (e.g., Daneman

& Merikle, 1996; Redick & Lindsey, 2013), an increased WM capacity should have widespread effects in the form of positive transfer to other cognitive domains (Morrison & Chein, 2011; von Bastian & Oberauer, 2014).

Despite the promising early findings from studies employing the WM core training approach (e.g., Borella et al., 2010; Holmes et al., 2009; Jaeggi et al., 2008), more recently published and methodologically more stringent studies have shown generalization mostly within the WM domain (e.g., Chooi & Thompson, 2012; Colom et al., 2013; Lilienthal et al., 2013; Salminen et al., 2012; Thompson et al., 2013). The results from the core training studies in the present thesis indicated even more limited transfer effects. Here, training-related changes were observed only in the trained WM tasks (**Study III,** partly **Study II)** and in untrained WM tasks that were structurally similar to the trained ones (only **Study III)**. Generalization to WM tasks based on untrained paradigms, or to tasks measuring other cognitive domains was non-existent. Even though these results may seem particularly meager, one should point out that most of the previous training studies have focused on the far transfer effects, and not systematically taken into consideration how much the untrained WM tasks resembled the practiced task(s). However, the few studies that have more carefully examined the transfer pattern within the WM domain indicate an absence of near transfer to WM tasks that are structurally dissimilar to the trained ones (De Simoni & von Bastian, 2018; Guye & von Bastian, 2017; von Bastian & Eschen, 2016), thus concurring with the results in the present thesis. Our weak and limited transfer effects also concur with the evidence from recent meta-analyses (Melby-Lervåg et al., 2016; Soveri et al., 2017). For example, Melby-Lervåg et al. (2016) evaluated 87 WM training studies concerning transfer effects to WM, intelligence, and several cognitive abilities relevant in daily life. The results indicated that WM training yields reliable improvements only on verbal and nonverbal WM tasks identical or similar to the trained tasks. In the meta-analysis by Soveri et al. (2017), altogether 33 randomized WM training studies employing n-back training were included in the analyses. The results were highly similar to the results in Melby-Lervåg et al. (2016): substantial improvements were observed only in untrained variants of the trained tasks, whereas transfer effects to untrained structurally dissimilar WM tasks, fluid intelligence, and cognitive control were of similar size and small.

Relating these rather bleak transfer effects to the hotly debated issue of the mechanisms underlying WM training, evidence favoring the WM Capacity hypothesis appears weak. Rather, the typically seen task-specific transfer effects could speak for the alternative view, namely the Strategy Mediation hypothesis (Dunning & Holmes, 2014; Peng & Fuchs, 2017). According to this hypothesis, WM training does not lead to an expansion of the WM capacity, but to a more effective use of existing WM resources, driven by employment of successful strategies. Strategies are mentally

effortful (Dunning & Holmes, 2014) and have been shown to have an substantial impact on WM performance (Carretti et al., 2007; Dunning & Holmes, 2014; McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003). The Strategy Mediation hypothesis was directly tested in **Study IV**. The results provided unequivocal support for the Strategy Mediation hypothesis: strategy instruction coupled with just a single adaptive WM training session yielded task-specific transfer typically seen following 4-6 week of unsupervised WM training (Soveri et al., 2017), and type and level of self-generated strategies had a strong impact on WM performance.

The very limited transfer following WM training could appear even paradoxical given the strong associations between WM performance and other higher-order cognitive abilities (e.g., Conway et al., 2003; Gathercole & Pickering, 2001; Poole & Kane, 2009). However, the Strategy Mediation hypothesis implies that this paradox is more apparent than real: after training, the role of WM changes only for the specific task type that was being trained. Moreover, the association between WM and, for instance, fluid intelligence may in the first place reflect fast and effective adaptation to novel tasks (e.g., Salthouse & Pink, 2008; Unsworth & Engle, 2005), perhaps not separate from an individual's proneness to create strategies.

To sum up, what does the current evidence suggest concerning the malleability of WM to cognitive training? Firstly, the WM system appears to be relatively fixed, with individual differences in WM functioning modulating the efficiency with which the WM capacity is used (Daneman & Carpenter, 1980; Peng & Fuchs, 2017). Indeed, this inter-individual variation in the utilization of this limited capacity system is large, and a major part of this variation has been attributed to hereditary factors. For example, twin- and family estimates have shown an heritability spanning from 31 to 65% on WM performance (as measured with various tasks tapping on WM) (e.g., Ando, Ono, & Wright, 2001; Blokland et al., 2011; Vogler et al., 2014), indicating that the WM capacity is to a significant part predetermined. Secondly, a substantial part of the improvements in WM efficiency appears to be based on adoption of effective strategies which would be expected to yield transfer to tasks based on either the same class of materials or task settings where the strategy in question can be applied (von Bastian & Oberauer, 2014). As the recent evidence mostly pinpoints to highly limited transfer effects (Guye & von Bastian, 2017; Melby-Lervåg et al., 2016; Soveri et al., 2017), it seems to be time to challenge the widely used "memory as a muscle" metaphor, according to which WM capacity can be enhanced through repeated and demanding training (Melby-Lervåg et al., 2016; Morra & Borella, 2015).

6.5 Methodological considerations

As in all studies, there are methodological considerations that should be taken into account when interpreting the results. One rather novel aspect of **Study II-IV** is that

the intervention periods were performed online in non-laboratory settings. Such a procedure eliminates the burden of repeated daily lab visits during the training period (Payne & Stine-Morrow, 2017) and enables a more efficient data collection. On the other hand, its disadvantages lie in the lack of control over the testing environment (Ford, 2017), potential issues with careless responding, and unreliable effort (Feitosa, Joseph, & Newman, 2015; Smith, Roster, Golden, & Albaum, 2016). Thus, despite carefully controlling for confounds (such as motivation and alertness) during the training period and providing meticulous instructions how to perform the tasks, we do not know whether all participants performed the training sessions in the expected way. Consequently, there is some evidence indicating that lab-based training might be more effective as compared with training at home (Lampit, Hallock, & Valenzuela, 2014; Schwaighofer et al., 2015). Thus, one cannot dismiss the possibility that the outcomes would have been different in **Studies II**-**III** if training would have taken place in the lab.

As WM training studies are both time-consuming and costly, most studies have had small sample sizes, typically spanning from 20 to 30 participants. This was also the case in **Studies II**-**IV** even though they can be considered to belong to the more rigorous ones (especially **Study IV**) in the context of WM training. Due to low statistical power, and thereby an increased likelihood of making type II errors, statistical analyses conducted with small samples should always be interpreted cautiously (Faber & Fonseca, 2014). Another methodological consideration concerns the statistical analyses employed in the present training studies. For investigating the pre- and posttest change in the cognitive tasks, we used ANCOVA analyses with pretest performance as the covariate, the posttest performance as the dependent variable, and group as the fixed factor. Such analyses require complete data from each time point, that is, the participants can only be included in the analysis if performance from both time points is available. In case of a lot of missing data (e.g., due to technical issues, and high dropout rates), it can lead to a substantial loss of power, and in the worst case, to biased outcomes (Enders & Bandalos, 2001; Roth, 1994). Thus, a viable analytical alternative would have been to employ linear mixed models (LMM). LMMs can handle missing data points, and conversely to the ANCOVA, it enables one to account for all systematic variance in the data in one and the same model (Krueger & Tian, 2004). Moreover, the sources of variance can be modeled either as fixed effects (e.g., group) or random effects (e.g, subjects or items) (Baayen, Davidson, & Bates, 2008). Ideally, with sufficiently large sample sizes, analyses at a latent level would clearly be preferable, as such analyses allow researchers not only handle missing data, but also to interpret the pre- and posttest outcomes at a construct level rather than at task level (e.g., Nilsson, Lebedev, Rydström, & Lövdén, 2017; Schmiedek, Lövden, & Lindenberger, 2010).

In **Studies II-III**, a methodological consideration related to the SUST paradigm should be pointed out. Although a steady rate of SUST improvement across the training period was observed in both studies, it is evident that the healthy younger adults in **Study II** reached substantially higher SUST levels as compared with the PD group in **Study III** (see Figure 14 for an illustration). This difference could partly be explained by an adjustment we made in the SUST task for the PD group in **Study III**. Here, we prevented adjacent words in a sentence to be updated simultaneously, whereas the words in **Study II** were allowed to be updated in any position. It might be that the updates located further apart from each other in a sentence would have elicited more interference due to distractions from the in-between words (e.g., Gibson, Desmet, Grodner, Watson, & Ko, 2005; Lewis & Vasishth, 2005; Nicenboim, Vasishth, Gattei, Sigman, & Kliegl, 2015), thereby creating more demands on WM in the SUST task employed in **Study III**. Also other factors, such as the nature of the level adjustment in the SUST task (see the Methods sections in the original articles), discrepancies in the time spent on the SUST task per session (healthy adults 15 min/session, PD group 10 min/session), the cognitive decline stemming from both an increased age (e.g., Salthouse, Atkinson, & Berish, 2003; Schroeder & Salthouse, 2004) and from the progressive neurological disease itself (Aarsland et al., 2017; Dubbelink et al., 2013) might have contributed to higher SUST levels during the training period for the healthy young adults.
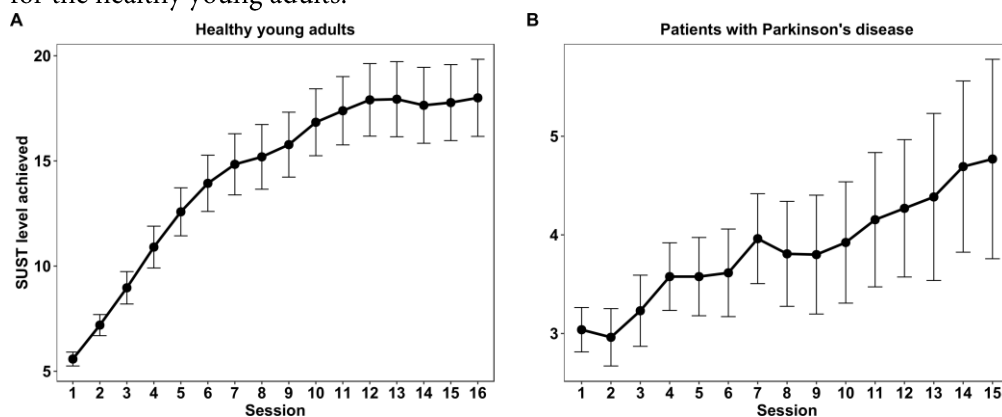
Figure 14. Studies II-III: Training improvement in the SUST task for (A) the healthy younger adults across the 16 training sessions, and for (B) the patients with Parkinson's disease across the 15 training sessions. Error bars represent standard error of means.

6.6 Limitations and future directions

Some limitations related to **Study I** should be pointed out. Firstly, our examination of the psychometric properties in the SUS task did not address its

discriminant validity that would have been worthwhile to test. Secondly, albeit the basic idea of **Study I** was to put forth a more ecologically valid WM task, the SUS task is artificial and does not simulate everyday verbal WM functioning per se. To probe for its ecological relevance, future studies could compare the SUS task against more ecologically valid measures of verbal ability such as verbal academic performance. Thirdly, a potential shortcoming concerned the training regimes employed in **Studies II-III,** both of which encompassed multiple WM training tasks. In case of substantial transfer, it would have been difficult to determine from which training task the generalization effects might have stemmed from. Thus, a carefully operationalized transfer model based on a theoretical framework of WM (e.g., Guye & von Bastian, 2017), a priori power analyses for determining the required sample sizes, and use of a single training paradigm would be needed to enable more specific and more valid interpretations of the mechanisms of eventual transfer.

Limitations specifically related to **Study III** concern the homogeneity of the present PD patients. As we were interested to investigate the cognitive level of the patients at onset, we compared their WM performance at baseline against those of healthy controls with comparable age and education. The results showed that the cognitive status of the present PD patients was well-preserved as their performance as a group was on par with the healthy controls. Thus, one cannot generalize our findings to PD patients with clear-cut cognitive deficits, and it is possible that the outcomes might have been different in case such deficits would have been present in our patient group as a whole (Vermeij et al., 2017). Another limitation concerns the screening procedure that we employed at the prescreening phase. Due to the fully online study design, we did not administer any motor examination for the enrolled participants. Thus, we do not know in detail how severe the PD patients' motor symptoms were. Moreover, it is possible that the telephone-based interview did not detect all relevant aspects of the PD symptomatology. In future home-based studies, video recordings could provide more information regarding the severity of the motor symptoms.

In **Study IV**, we concluded that employment of efficient strategies can be an important mechanism for training-induced changes following WM training. However, there are several limitations that should be taken into account when interpreting the results. Perhaps the most notable issue concerns the shortness of the study: it can only address the initial strategy-related effects in WM training. Accordingly, it raises the intriguing question on strategy evolvement during the typical 4-6 week WM memory training. At a more general level, in future studies it would be useful to study WM training mechanisms in the context of general cognitive skill learning. The *theory of triarchic learning* by Chein & Schneider (2012) suggests that specific cognitive skills develop through three hierarchically organized learning systems, starting from the *Formation phase*, followed by *Controlled execution*, and

proceeding into *Automatic execution.* During the formation phase, metacognitive skills (e.g., generation of new behavioral routines and performance monitoring) are highly activated, allowing one to consider effective alternative strategies (e.g., grouping or rehearsal in the context of WM tasks) when performing a novel task for the first time. This phase is followed by controlled execution during which the cognitive control system is engaged: at this point, suitable strategies are adopted and one have gained more acquaintance with the task. Lastly, learning gradually shifts to more automatic execution. This entails a consolidating phase where the learned representations and procedures slowly strengthen specific input-output associations that underlie the execution of the task. Consequently, the role of the cognitive control system is gradually reduced, eventually leading to more or less automatic task performance. Applying the triarchic theory to WM training, a tentative speculation is that several weeks of uninstructed WM training gives ample opportunity for one to firstly elaborate how to solve the training task, to develop a efficient strategy for the WM task at hand, retain that strategy for the rest of the training period, and show a gradually stabilizing learning curve towards the end of intervention. To test the Strategy Mediation hypothesis in the more general skill learning context, future studies should investigate the development of strategies and the transfer effects following a full period of strategy-based WM training.

Another limitation in **Study IV** was that the strategy-based training was administered only on one task paradigm, namely the single n-back task. Although the single n-back task has been extensively used as a training task (e.g., Bürki, Ludwig, Chicherio, & de Ribaupierre, 2014; Küper & Karbach, 2016; Schwarb, Nail, & Schumacher, 2016), it should be noted that also other paradigms, such as the dual n-back task (e.g., Jaeggi et al., 2008; Lilienthal et al., 2013; Redick et al., 2013), and the complex span tasks (e.g., Chein & Morrison, 2010; Payne & Stine-Morrow, 2017) are commonly used for WM training. Thus, our results should be interpreted only as evidence for a strategy effect under a specific set of circumstances, and the generalizability of our findings to other tasks and paradigms remains to be seen. Thus, future studies should investigate the transfer effects of strategy-based WM training also with other paradigms than the single n-back task.

Given that the most substantial transfer effects following WM training appear to be limited to task-specific transfer measures (Soveri et al., 2017), one could take either an optimistic or pessimistic view on such improvements. If one chooses to take an optimistic view, it could be argued that not all strategies that might underlie task-specific transfer have to be task-specific, and the strategies employed for a certain task would have the potential to transfer to other task paradigms where the same strategy can be useful (Sprenger et al., 2013; von Bastian & Oberauer, 2014). There is, for example, some evidence indicating that strategies such as grouping, imagery, and

sentence generation carry the potential to transfer over task paradigms (Dunlosky & Kane, 2007). If one chooses to take a pessimistic view of strategy use, one could emphasize more the task-specific aspects of strategies. Research on mnemonics has shown the limited scope of strategies skilled memorizers utilize for remembering task-relevant information (Dresler et al., 2017; Maguire et al., 2003). Mnemonic strategies enable amazing feats: in an influential study by Chase and Ericsson (1982), a participant was able to recall a span list of 80 consecutive digits, but the drawback is the weak generalization, typically not extending further than to the task that has been practiced with. Nevertheless, as research on strategy use in the context of WM training is still in its infancy, it seems worthwhile to pursue this further to better understand the evolvement and effects of strategy use in WM performance across tasks and individuals (Morrison, Rosenbaum, Fair, & Chein, 2016).

As most of the available evidence indicates highly limited transfer effects of WM training, one can ask whether it is worthwhile to continue with very time-consuming and expensive training studies chasing for widespread generalization effects that – as it currently seems – do not exist. First of all, it should be pointed out that albeit the recently conducted meta-analyses show only small effect sizes for far transfer, most of them still indicate that these effects are statistically significant (e.g., Au et al., 2015; Soveri et al., 2017; Weicker et al., 2016). It is debatable whether such improvements could bear some positive benefits also to cognitive functioning in real-life settings. Would, for example, an increase of a few IQ points make a difference in everyday life and how long-lasting would such an effect be? Secondly, even though WM training might not have lived up to its expectations, it does not necessarily mean that it would be impossible to come up with novel complex training protocols that could elicit more far transfer effects. One could, for instance, argue that less explored sub-components of the WM construct could have more potential to elicit transfer. However, the prospects for finding such effects appear slight, since most of the current evidence suggests weak effects irrespective whether training taps on processes such as switching (Nilsson et al., 2017), updating (Waris et al., 2015), or binding (De Simoni & von Bastian, 2018; Guye & von Bastian, 2017). Thus, the present evidence suggests that that whatever WM component is being trained, the present form of WM training simply fails to elicit generalization. Instead of administering repetive and highly artificial training tasks on a computer screen, perhaps a more holistic training approach would have more substantial effects extending across cognitive domains. An interesting avenue of research would for instance be to build a training platform using virtual reality techniques (e.g., Gamito et al., 2015) encompassing tasks that are both varied, multi-faceted, WM-demanding, and ecologically valid, thereby tapping cognitive demands that we meet in everyday life as well.

SUMMARY AND CONCLUSIONS

The present thesis investigated the effects of core training targeting verbal WM and examined the viability of the Strategy Mediation hypothesis as an account for training-induced changes in WM. The key findings and conclusions of the thesis are listed below:

1. A novel sentence-level WM updating task proved to be a adequate task for capturing verbal WM processes (**Study I**), and showed to work as a WM training task as well (**Studies II-III**).

2. Core training targeting verbal WM at sentence-level showed no transfer effects in healthy young adults (**Study II**).

3. Core training targeting verbal WM updating resulted in task-specific near transfer, but not task-general near transfer or far transfer effects in patients with PD (**Study III**).

4. A single session of strategy-based WM training was enough to elicit task-specific near transfer typically seen after several weeks of training. Moreover, type and level of detail of self-generated strategies was associated with an improved WM performance (**Study IV**).

5. Overall, the results of the present thesis suggest that core WM training appears to yield improvements mostly in the trained WM tasks and in its untrained variants. These improvements might be mediated by development of task-specific strategies that participants develop during training. Thus, the findings from the present thesis are in line with the Strategy Mediation hypothesis rather than the WM Capacity hypothesis (**Studies II-IV**).

REFERENCES

Aarsland, D., Creese, B., Politis, M., Chaudhuri, K. R., Ffytche, D. H., Weintraub, D., & Ballard, C. (2017). Cognitive decline in Parkinson disease. *Nature Reviews Neurology*, *13*(4), 217–231. doi:10.1038/nrneurol.2017.27

Ando, J., Ono, Y., & Wright, M. J. (2001). Genetic structure of spatial and verbal working memory. *Behavior Genetics*, *31*(6), 615–624. doi:10.1016/j.neurobiolaging.2009.12.023

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 and Review*, *22*(2), 366–377. doi:10.3758/s13423-014-0699-x

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. doi:10.1016/j.jml.2007.12.005

Bäckman, L., Waris, O., Johansson, J., Andersson, M., Rinne, J. O., Alakurtti, K., … Nyberg, L. (2017). Increased dopamine release after working-memory updating training: Neurochemical correlates of transfer. *Scientific Reports*, *7*. doi:10.1038/s41598-017-07577-y

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. doi:10.1016/S1364-6613(00)01538-2

Bailey, H. R., Dunlosky, J., & Hertzog, C. (2014). Does strategy training reduce age-related deficits in working memory? *Gerontology*, *60*, 346–356. doi:10.1159/000356699

Baniqued, P. L., Allen, C. M., Kranz, M. B., Johnson, K., Sipolins, A., Dickens, C., … Kramer, A. F. (2015). Working memory, reasoning, and task switching training: Transfer effects, limitations, and great expectations? *PLoS ONE*, *10*(11). doi:10.1371/journal.pone.0142169

Bartsch, A., & Kothe, E. (2016). The effect of non-specific response inhibition training on alcohol consumption: an intervention. *Journal of Addiction Research & Therapy*, *07*(01), 260. doi:10.4172/2155-6105.1000260

Beck, A., Steer, R., & Brown, G. (2004). *BDI-II Käsikirja [Manual of the BDI-II Finnish version]*. Helsinki: Psykologien Kustannus; 2004.

Benton, A., & Hamsher, K. (1978). *Multilingual Aphasia Examination. University of Iowa, Iowa* (Rev.). Iowa City: Department of Neurology, University of Iowa Hospitals.

Bergman-Nutley, S., & Klingberg, T. (2014). Effect of working memory training on working memory, arithmetic and following instructions. *Psychological Research, 78*(6), 869–877. doi:10.1007/s00426-014-0614-0

Blokland, G. A., McMahon, K. L., Thompson, P. M., Martin, N. G., de Zubicaray, G. I., & Wright, M. J. (2011). Heritability of working memory brain activation. *Journal of Neuroscience, 31*(30), 10882–10890. doi:10.1523/JNEUROSCI.5334-10.2011

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science, 8*(4), 445–454.

doi:10.1177/1745691613491271

Borella, E., Carretti, B., Riboldi, F., & De Beni, R. (2010). Working memory training in older adults: evidence of transfer and maintenance effects. *Psychology and Aging, 25*(4), 767–778. doi:10.1037/a0020683

Brehmer, Y., Westerberg, H., & Bäckman, L. (2012). Working-memory training in younger and older adults: training gains, transfer, and maintenance. *Frontiers in Human Neuroscience, 6*, 63. doi:10.3389/fnhum.2012.00063

Bürki, C. N., Ludwig, C., Chicherio, C., & de Ribaupierre, A. (2014). Individual differences in cognitive plasticity: an investigation of training curves in younger and older adults. *Psychological Research, 78*(6), 821–835. doi:10.1007/s00426-014-0559-3

Carretti, B., Borella, E., & De Beni, R. (2007). Does strategic memory training improve the working memory performance of younger and older adults? *Experimental Psychology, 54*(4), 311–320. doi:10.1027/1618-3169.54.4.311

Carretti, B., Borella, E., Zavagnin, M., & De Beni, R. (2013). Gains in language comprehension

relating to working memory training in healthy older adults. *International Journal of Geriatric Psychiatry*, *28*(5), 539–546. doi:10.1002/gps.3859

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. *Psychology of Learning and Motivation*, *16*(C), 1–58. doi:10.1016/S0079-7421(08)60546-0

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. doi:10.3758/PBR.17.2.193

Chein, J. M., & Schneider, W. (2012). The brain's learning and control architecture. *Current Directions in Psychological Science*, *21*(2), 78–84. doi:10.1177/0963721411434977

Chooi, W.-T., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence*, *40*(6), 531–542. doi:10.1016/j.intell.2012.07.004

Clark, C. M., Lawlor-Savage, L., & Goghari, V. M. (2017). Working memory training in healthy young adults: Support for the null from a randomized comparison to active and passive control groups. *PLoS ONE*, *12*(5). doi:10.1371/journal.pone.0177707

Colom, R., Román, F. J., Abad, F. J., Shih, P. C., Privado, J., Froufe, M., … Jaeggi, S. M. (2013). Adaptive n-back training does not improve fluid intelligence at the construct level: Gains on individual tests suggest that training may enhance visuospatial processing. *Intelligence*, *41*(5), 712–727. doi:10.1016/j.intell.2013.09.002

Conners, C. K., Epstein, J. N., Angold, A., & Klaric, J. (2003). Continuous performance test performance in a normative epidemiological sample. *Journal of Abnormal Child Psychology*, *31*(5), 555–562. doi:10.1023/A:10254573

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769–786.

Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–

552. doi:S1364661303002833

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. doi:10.1017/S0140525X0100392 2

Cowan, N. (2014). Working nemory underpins cognitive development, learning, and education. *Educational Psychology Review*, *26*(2), 197– 223. doi:10.1007/s10648-013-9246-y

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42–100. doi:http://dx.doi.org/10.1016/j.c ogpsych.2004.12.001

Dahlin, E., Bäckman, L., Neely, A. S., & Nyberg, L. (2009). Training of the executive component of working memory: Subcortical areas mediate transfer effects. *Restorative Neurology and Neuroscience*, *27*(5), 405–419. doi:10.3233/RNN-2009-0492

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

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422–433. doi:10.3758/BF03214546.

De Simoni, C., & von Bastian, C. C. (2018). Working memory updating and binding training: Bayesian evidence supporting the absence of transfer. *Journal of Experimental Psychology: General*, *147*(6), 829–858. doi:10.1037/xge0000453

Dehn, M. J. (2015). *Essentials of working memory assessment and intervention*. Hoboken, NJ: Wiley & Sons.

Dresler, M., Shirer, W. R., Konrad, B. N., Müller, N. C. J., Wagner, I. C., Fernández, G., … Greicius, M. D. (2017). Mnemonic Training Reshapes Brain Networks to Support Superior Memory. *Neuron*, *93*(5), 1227–1235.e6. doi:10.1016/j.neuron.2017.02.00 3

Dubbelink, K. T. E. O., Stoffers, D., Deijen, J. B., Twisk, J. W. R., Stam, C. J., & Berendse, H. W.

(2013). Cognitive decline in Parkinson's disease is associated with slowing of resting-state brain activity: A longitudinal study. *Neurobiology of Aging*, *34*(2), 408–418. doi:10.1016/j.neurobiolaging.2012.02.029

Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *Quarterly Journal of Experimental Psychology*, *60*(9), 1227–1245. doi:10.1080/17470210600926075

Dunning, D. L., & Holmes, J. (2014). Does working memory training promote the use of strategies on untrained working memory tasks? *Memory and Cognition*, *42*(6), 854–862. doi:10.3758/s13421-014-0410-5

Edwards, J. D., Hauser, R. A., O'Connor, M. L., Valdes, E. G., Zesiewicz, T. A., & Uc, E. Y. (2013). Randomized trial of cognitive speed of processing training in Parkinson disease. *Neurology*, *81*(15), 1284–1290. doi:10.1212/WNL.0b013e3182a823ba

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*(3), 430–457. doi:10.1207/S15328007SEM0803_5

Engle, R. W., Laughlin, J. E., Tuholski, S. W., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331. doi:10.1037/0096-3445.128.3.309

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245. doi:10.1037/0033-295X.102.2.211

Faber, J., & Fonseca, L. M. (2014). How sample size influences research outcomes. *Dental Press Journal of Orthodontics*, *19*(4), 27–29. doi:10.1590/2176-9451.19.4.027-029.ebo

Feitosa, J., Joseph, D. L., & Newman, D. A. (2015). Crowdsourcing and personality measurement equivalence: A warning about countries whose primary language is not English. *Personality and Individual Differences*, *75*, 47–52. doi:10.1016/j.paid.2014.11.017

Fischler, I., & Bloom, P. A. (1980). Rapid processing of the meaning

of sentences. *Memory & Cognition*, *8*(3), 216–225. doi:10.3758/BF03197609

Ford, J. B. (2017). Amazon's Mechanical Turk: A Comment. *Journal of Advertising*, *46*(1), 156–158. doi:10.1080/00913367.2016.1277380

Gamito, P., Oliveira, J., Coelho, C., Morais, D., Lopes, P., Pacheco, J., … Barata, A. F. (2017). Cognitive training on stroke patients via virtual reality-based serious games. *Disability and Rehabilitation*, *39*(4), 385–388. doi:10.3109/09638288.2014.934925

Gathercole, S. E., & Pickering, S. (2001). Working memory deficits in children with special educational needs. *British Journal of Special Education*, *28*(2), 89–97. doi:10.1111/1467-8527.00225

Gibson, E., Desmet, T., Grodner, D., Watson, D., & Ko, K. (2005). Reading relative clauses in English. *Cognitive Linguistics*, *16*(2), 313–353. doi:10.1515/cogl.2005.16.2.313

Glizer, D., & MacDonald, P. A. (2016). Cognitive training in Parkinson's disease: A review of studies from 2000 to 2014. *Parkinson's Disease*, *2016*, 1–19.

doi:10.1155/2016/9291713

Guye, S., & von Bastian, C. C. (2017). Working memory training in older adults: Bayesian evidence supporting the absence of transfer. *Psychology and Aging*, *32*(8), 732–746. doi:10.1037/pag0000206

Hambrick, D. Z., Oswald, F. L., Darowski, E. S., Rench, T. A., & Brou, R. (2010). Predictors of multitasking performance in a synthetic work paradigm. *Applied Cognitive Psychology*, *24*(8), 1149–1167. doi:10.1002/acp.1624

Holmes, J. (2011). Baby brain: Training executive control in infancy. *Current Biology*, *21*(18), R684–R685. doi:10.1016/j.cub.2011.08.026

Holmes, J., & Gathercole, S. E. (2014). Taking working memory training from the laboratory into schools. *Educational Psychology*, *34*(4), 440–450. doi:10.1080/01443410.2013.797338

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

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. doi:10.1073/pnas.0801268105

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. doi:10.3758/s13421-013-0364-z

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning— implications for training and transfer. *Intelligence, 38*(6), 625–635. doi:10.1016/j.intell.2010.09.001

Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when: an experience-sampling study of working memory and executive control in daily life. *Psychological Science, 18*(7), 614–621. doi:10.1111/j.1467-9280.2007.01948.x

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. doi:10.1037/0096-3445.133.2.189

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. doi:10.1177/0956797614548725

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55*(4), 352–358. doi:10.1037/h0043688

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences, 14*(7), 317–324. doi:10.1016/j.tics.2010.05.002

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.

König, C. J., Bühner, M., & Mürling,

G. (2005). Working memory, fluid intelligence, and attention are predictors of multitasking performance, but polychronicity and extraversion are not. *Human Performance, 18*(3), 243–266. doi:10.1207/s15327043hup1803_3

Krueger, C., & Tian, L. (2004). A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. *Biological Research for Nursing, 6*(2), 151–157. doi:10.1177/1099800404267682

Küper, K., & Karbach, J. (2016). Increased training complexity reduces the effectiveness of brief working memory training: Evidence from short-term single and dual n-back training interventions. *Journal of Cognitive Psychology, 28*(2), 199–208. doi:10.1080/20445911.2015.1118106

Lampit, A., Hallock, H., & Valenzuela, M. (2014). Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers. *PLoS Medicine, 11*(11). doi:10.1371/journal.pmed.1001756

Lawlor-Savage, L., & Goghari, V. M. (2016). Dual N-back working memory training in healthy adults: A randomized comparison to processing speed training. *PLoS ONE, 11*(4). doi:10.1371/journal.pone.0151817

Lewis, R., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*(3), 375–419. doi:10.1207/s15516709cog0000_25

Lewis, R., Vasishth, S., & Van Dyke, J. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences, 10*(10), 447–454. doi:10.1016/j.tics.2006.08.007

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. doi:10.3758/s13423-012-0335-6

Lustig, C., Shah, P., Seidler, R., & Reuter-Lorenz, P. A. (2009). Aging, training, and the brain: A review and future directions. *Neuropsychology Review, 19*(4), 504–522. doi:10.1007/s11065-009-9119-9

Maguire, E. A., Valentine, E. R., Wilding, J. M., & Kapur, N. (2003). Routes to remembering: The brains behind superior memory. *Nature Neuroscience, 6*(1), 90–95. doi:10.1038/nn988

McCabe, J. A., Redick, T. S., & Engle, R. W. (2016). Brain-training pessimism, but applied-memory optimism. *Psychological Science in the Public Interest, 17*(3), 187–191. doi:10.1177/1529100616664716

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition, 29*(1), 10–17. doi:10.3758/BF03195736

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49*(2), 270–291. doi:10.1037/a0028228

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer": evidence from a meta-analytic review. *Perspectives on Psychological Science, 11*(4), 512–534. doi:10.1177/1745691616635612

Mirman, D. (2014). *Growth curve analysis and visualization using R.* New York, NY: CRC Press.

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. doi:10.1006/cogp.1999.0734

Moreau, D., & Conway, A. R. A. (2014). The case for an ecological approach to cognitive training. *Trends in Cognitive Sciences, 18*, 334–336. doi:10.1016/j.tics.2014.03.009

Morra, S., & Borella, E. (2015). Working memory training: from metaphors to models. *Frontiers in Psychology, 6*, 1097. doi:10.3389/fpsyg.2015.01097

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. doi:10.3758/s13423-010-0034-0

Morrison, A. B., Rosenbaum, G. M., Fair, D., & Chein, J. M. (2016). Variation in strategy use across measures of verbal working memory. *Memory & Cognition,*

44(6), 922–936. doi:10.3758/s13421-016-0608-9

Motter, J. N., Pimontel, M. A., Rindskopf, D., Devanand, D. P., Doraiswamy, P. M., & Sneed, J. R. (2016). Computerized cognitive training and functional recovery in major depressive disorder: A meta-analysis. *Journal of Affective Disorders*, *189*(1), 184–191. doi:10.1016/j.jad.2015.09.022

Murty, V. P., Sambataro, F., Radulescu, E., Altamura, M., Iudicello, J., Zoltick, B., … Mattay, V. S. (2011). Selective updating of working memory content modulates meso-cortico-striatal activity. *NeuroImage*, *57*(3), 1264–1272. doi:10.1016/j.neuroimage.2011.05.006

Nicenboim, B., Vasishth, S., Gattei, C., Sigman, M., & Kliegl, R. (2015). Working memory differences in long-distance dependency resolution. *Frontiers in Psychology*, *6*, 312. doi:10.3389/fpsyg.2015.00312

Nilsson, J., Lebedev, A. V, Rydström, A., & Lövdén, M. (2017). Direct-current stimulation does little to improve the outcome of working memory training in older adults. *Psychological Science*, *28*(7), 907–920. doi:10.1177/0956797617698139

Oberauer, K., Süß, H. M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, *36*(6), 641–652. doi:10.1016/j.intell.2008.01.007

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., … Ballard, C. G. (2010). Putting brain training to the test. *Nature*, *465*(7299), 775–778. doi:10.1038/nature09042

Payne, B. R., & Stine-Morrow, E. A. L. (2017). The effects of home-based cognitive training on verbal working memory and language comprehension in older adulthood. *Frontiers in Aging Neuroscience*, *9*, 256. doi:10.3389/fnagi.2017.00256

Peña, J., Ibarretxe-Bilbao, N., Garcia-Gorostiaga, I., Gomez-Beldarrain, M. A., Diez-Cirarda, M., & Ojeda, N. (2014). Improving functional disability and cognition in Parkinson disease: randomized controlled trial. *Neurology*, *83*(23), 2167–2174. doi:10.1212/WNL.0000000000001043

Peng, P., & Fuchs, D. (2017). A randomized control trial of working memory training with and without strategy instruction: Effects on young children's working memory and

comprehension. *Journal of Learning Disabilities*, *50*(1), 62–80. doi:10.1177/0022219415594609

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

Poole, B. J., & Kane, M. J. (2009). Working-memory capacity predicts the executive control of visual search among distractors: The influences of sustained and selective attention. *The Quarterly Journal of Experimental Psychology*, *62*(7), 1430–1454. doi:10.1080/17470210802479329

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, *20*(2), 110–122. doi:10.1016/j.lindif.2009.10.005

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

Redick, T. S., Shipstead, Z., Harrison,

T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., … Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, *142*(2), 359. doi:10.1037/a0029082

Richmond, L. L., Morrison, A. B., Chein, J. M., & Olson, I. R. (2011). Working memory training and transfer in older adults. *Psychology and Aging*, *26*(4), 813–822. doi:10.1037/a0023631

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, *47*(3), 537–560. doi:10.1111/j.1744-6570.1994.tb01736.x

Roth, R. M., Isquith, P. K., & Gioia, G. A. (2005). *BRIEF-A: Behavior Rating Inventory of Executive Function--adult Version*. Psychological Assessment Resources.

Salminen, T., Strobach, T., & Schubert, T. (2012). On the impacts of working memory training on executive functioning. *Frontiers in Human Neuroscience*, *6*. doi:10.3389/fnhum.2012.00166

Salo, R., Henik, A., & Robertson, L. C.

(2001). Interpreting Stroop interference: An analysis of differences between task versions. *Neuropsychology*, *15*(4), 462–471. doi:10.1037/0894-4105.15.4.462

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General*, *132*(4), 566–594. doi:10.1037/0096-3445.132.4.566

Salthouse, T. A., & Pink, J. E. (2008). Why is working memory related to fluid intelligence? *Psychonomic Bulletin and Review*, *15*, 364–371. doi:10.3758/PBR.15.2.364

Schmiedek, F., Lövden, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: findings from the COGITO study. *Frontiers in Aging Neuroscience*, *2*, 27. doi:10.3389/fnagi.2010.00027

Schroeder, D. H., & Salthouse, T. A. (2004). Age-related effects on cognition between 20 and 50 years of age. *Personality and Individual Differences*, *36*(2), 393–404. doi:10.1016/S0191-8869(03)00104-1

Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, *50*(2), 138–166. doi:10.1080/00461520.2015.1036274

Schwarb, H., Nail, J., & Schumacher, E. H. (2016). Working memory training improves visual short-term memory capacity. *Psychological Research*, *80*(1), 128–148. doi:10.1007/s00426-015-0648-y

Shipstead, Z., Hicks, K. L., & Engle, R. W. (2012). Cogmed working memory training: Does the evidence support the claims? *Journal of Applied Research in Memory and Cognition*, *1*(3), 185–193. doi:10.1016/j.jarmac.2012.06.003

Shipstead, Z., Redick, T., & Engle, R. (2010). Does working memory training generalize? *Psychologica Belgica*, *50*, 245–276. doi: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. doi:10.1037/a0027473

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C.

F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest, 17*(3), 103–186. doi:10.1177/1529100616661983

Smith, S. M., Roster, C. A., Golden, L. L., & Albaum, G. S. (2016). A multi-group analysis of online survey respondent data quality: Comparing a regular USA consumer panel to MTurk samples. *Journal of Business Research, 69*(8), 3139–3148. doi:10.1016/j.jbusres.2015.12.002

Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin and Review, 24*(4), 1077–1096. doi:10.3758/s13423-016-1217-0

Soveri, A., Karlsson, E. P. A., Waris, O., Grönholm-Nyman, P., & Laine, M. (2017). Pattern of near transfer effects following working memory training with a dual n-back task. *Experimental Psychology, 64*(4), 240–252. doi:10.1027/1618-3169/a000370

Spitzer, R., Kroenke, K., & Williams, J. B. W. (1999). Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA, 282*(18), 1737–1744. doi:10.1001/jama.282.18.1737

Sprenger, A. M., Atkins, S. M., Bolger, D. J., Harbison, J. I., Novick, J. M., Chrabaszcz, J. S., … Dougherty, M. R. (2013). Training working memory: Limits of transfer. *Intelligence, 41*(5), 638–663. doi:10.1016/j.intell.2013.07.013

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137–149. doi:10.3758/BF03207704

Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence, 30*(3), 261–288. doi:10.1016/S0160-2896(01)00100-3

Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology, 96*(3), 471–491. doi:10.1037/0022-0663.96.3.471

Swanson, H. L., & McMurran, M.

(2018). The impact of working memory training on near and far transfer measures: Is it all about fluid intelligence? *Child Neuropsychology*, *24*(3), 370–395. doi:10.1080/09297049.2017.1280142

Thompson, T. W., Waskom, M. L., Garel, K. L. A., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., … Gabrieli, J. D. E. (2013). Failure of working memory training to enhance cognition or intelligence. *PLoS ONE*, *8*(5). doi:10.1371/journal.pone.0063614

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

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154. doi:10.1016/0749-596X(89)90040-5

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence*, *33*(1), 67–81.

doi:10.1016/j.intell.2004.08.003

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. doi:10.1080/09602011.2012.681110

Vermeij, A., Kessels, R. P. C., Heskamp, L., Simons, E. M. F., Dautzenberg, P. L. J., & Claassen, J. A. H. R. (2017). Prefrontal activation may predict working-memory training gain in normal aging and mild cognitive impairment. *Brain Imaging and Behavior*, *11*(1), 141–154. doi:10.1007/s11682-016-9508-7

Vogler, C., Gschwind, L., Coynel, D., Freytag, V., Milnik, A., Egli, T., … Papassotiropoulos, A. (2014). Substantial SNP-based heritability estimates for working memory performance. *Translational Psychiatry*, *4*(9). doi:10.1038/tp.2014.81

von Bastian, C. C., & Eschen, A. (2016). Does working memory training have to be adaptive? *Psychological Research*, *80*(2), 181–194. doi:10.1007/s00426-015-0655-z

von Bastian, C. C., & Oberauer, K. (2014). Effects and mechanisms of working memory training: a review. *Psychological Research, 78*(6), 803–820. doi:10.1007/s00426-013-0524-6

Walton, C. C., Naismith, S. L., Lampit, A., Mowszowski, L., & Lewis, S. J. G. (2017). Cognitive training in Parkinson's disease: a theoretical perspective. *Neurorehabilitation and Neural Repair, 31*(3), 207–216. doi:10.1177/1545968316680489

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

Was, C. A., Rawson, K. A., Bailey, H., & Dunlosky, J. (2011). Content-embedded tasks beat complex span for predicting comprehension. *Behavior Research Methods, 43*(4), 910–915. doi:10.3758/s13428-011-0112-x

Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers, 35*(4), 550–564. doi:10.3758/BF03195534

Weicker, J., Villringer, A., & Thöne-Otto, A. (2016). Can impaired working memory functioning be improved by training? A meta-analysis with a special focus on brain injured patients. *Neuropsychology, 30*(2), 190–212. doi:10.1037/neu0000227

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1983). Development and validation of a geriatric depression screening scale: a preliminary report. *Journal of Psychiatric Research, 17*(1), 37–49. doi:10.1016/0022-3956(82)90033-4

Zysset, S., Müller, K., Lohmann, G., & Von Cramon, D. Y. (2001). Color-word matching Stroop task: Separating interference and response conflict. *NeuroImage, 13*(1), 29–36. doi:10.1006/nimg.2000.0665