Article:

De Simoni, C. and 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). pp. 829-858. ISSN 0096-3445

https://doi.org/10.1037/xge0000453

Working Memory Updating and Binding Training: Bayesian Evidence Supporting the Absence of Transfer

Carla De Simoni

University of Zurich

Claudia C. von Bastian

University of Sheffield

University of Zurich

Author Note

Carla De Simoni, Department of Psychology, University of Zurich; Claudia C. von Bastian, Department of Psychology, University of Sheffield, United Kingdom.

This work was supported by a grant from the Swiss National Science Foundation (No. 100014_146074) to Klaus Oberauer, and by the URPP "Dynamics of Healthy Aging" of the University of Zurich. This project was presented at the 14th biannual congress of the Swiss Psychological Society (SSP/SGP) in Geneva, Switzerland, 2015, at the 57th annual meeting of the Psychonomic Society in Boston, MA, 2016, and at the Conference of the European Society for Cognitive Psychology in Potsdam, Germany, 2017. We thank Clara Overkott, Samuel Winiger, Aline Gaillard, Doris Pfister, Monalisa Stiefel, and Aylin Yantaz for their assistance with collecting the data. We also thank Emma Blakey for her valuable comments on a previous version of this article. The data and analysis scripts are available on the Open Science Framework (https://osf.io/fy5ku). Some of these data have been used in a prior publication (Guye, De Simoni, & von Bastian, 2017).

Correspondence concerning this article should be addressed to Claudia C. von Bastian, Department of Psychology, Cathedral Court, 1 Vicar Lane, Sheffield S1 1HD. E-mail: c.c.vonbastian@sheffield.ac.uk

Abstract

As working memory (WM) predicts a wide range of other abilities, it has become a popular target for training interventions. However, its effectiveness to elicit generalized cognitive benefits is still under debate. Previous research yielded inconsistent findings and focused only little on the mechanisms underlying transfer effects. To disentangle training effects on WM capacity and efficiency, we evaluated near transfer to untrained, structurally different WM tasks and far transfer to closely related abilities (i.e., reasoning, processing speed, task switching, and inhibitory control) in addition to process-specific effects on three WM mechanisms (i.e., focus switching, removal of WM contents, and interference resolution). We randomly assigned 197 young adults to one of two experimental groups (updating or item-to-context binding) or to an active control group practicing visual search tasks. Before and after five weeks of adaptive training, performance was assessed measuring each of the cognitive processes and abilities of interest with four tasks covering verbal-numerical and visual-spatial materials. Despite the relatively large sample size, large practice effects in the trained tasks, and at least moderate correlations between WM training tasks and transfer measures, we found consistent evidence for the absence of any training-induced improvements across all ranges of transfer and mechanisms. Instead, additional analyses of error patterns and self-reported strategy use indicated that WM training encouraged the development of stimulus-specific expertise and use of paradigm-specific strategies. Thus, the results suggest that the WM training interventions examined here enhanced neither WM capacity nor the WM mechanisms assumed to underlie transfer.

Keywords: working memory capacity, updating, binding, cognitive training and transfer

Working Memory Updating and Binding Training: Bayesian Evidence Supporting the Absence of Transfer

Throughout the 20th century, the prospect of broadly enhancing cognitive abilities through repetitively practicing a set of tasks has been of central interest for commercial applications (e.g., the ‘mind training program’ by the Pelmanism Institute of America, 1920) and academic research (e.g., Ball et al., 2002; Judd, 1908; Kramer, Larish, & Strayer, 1995; Logie, Baddeley, Mané, Donchin, & Sheptak, 1989; Willis, Cornelius, Blow, & Baltes, 1983). Over the past 15 years, computerized working memory (WM) training has become a particularly popular target for training interventions. WM is a capacity-limited system that provides temporary access to representations needed for ongoing cognitive processes, thereby building the basis for complex cognition. WM capacity strongly correlates with a wide range of higher-order cognitive abilities such as reasoning (e.g., Engle, Tuholski, Laughlin, & Conway, 1999; Friedman, Miyake, Schmeichel, & Tang, 2006; Kyllonen & Christal, 1990; Oberauer, Süß, Wilhelm, & Wittmann, 2008; Süß, Oberauer, Wittman, Wilhelm, & Schulze, 2002), executive functions (e.g., Friedman et al., 2008; Miyake et al., 2000; Miyake & Friedman, 2012), and academic achievement (e.g., St. Clair-Thompson & Gathercole, 2006). Based on the assumption that those correlations reflect overlapping cognitive processes (e.g., Kovacs & Conway, 2016), it has been hypothesized that training WM may not only improve WM capacity but enhances performance also in tasks measuring those related abilities (e.g., Klingberg, Forssberg, & Westerberg, 2002; Klingberg et al., 2005). Early studies indeed reported promising empirical evidence for the transfer of training gains to untrained WM tasks (i.e., “near transfer”) and even to related abilities (i.e., “far transfer” such as to fluid intelligence, e.g., Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Klingberg, Forssberg, & Westerberg, 2002), sparking enthusiasm that “fluid intelligence is trainable to a significant and meaningful degree" (Sternberg, 2008, p. 6791). More than a decade later, however, meta-analyses report inconsistent findings, with sometimes significant, sometimes nonsignificant small far transfer effects (e.g., Au et al., 2015; Melby-Lervåg, Redick, & Hulme, 2016; Schwaighofer, Fischer, & Bühner, 2015; Soveri, Antfolk, Karlsson, Salo, & Laine, 2017; Weicker, Villringer, & Thöne-Otto, 2015).

It is yet unclear when and under which circumstances WM training elicits far transfer effects, as the mechanisms underlying potential transfer effects have been largely neglected in much of past research (cf. von Bastian & Oberauer, 2014; for notable exceptions, see Gibson, Gondoli, Johnson, Steeger, & Morrissey, 2012; Hussey et al., 2016; Lilienthal, Tamez, Shelton, Myerson, & Hale, 2013; Ralph et al., 2017; Waris, Soveri, & Laine, 2015). Moreover, any meta-analytic findings must be interpreted with caution, as they are based on studies that often suffered from methodological issues such as evaluating WM training effects relative to no-contact control groups, using single transfer measures, or low statistical power (cf. Bogg & Lasecki, 2015; Moreau, Kirk, & Waldie, 2016; Shipstead, Redick, & Engle, 2013). The present study aims to fill this gap by using methodological rigor to systematically investigate whether and, if so, how WM training affects theoretically derived indicators of WM capacity and efficiency.

Mechanisms of Transfer

Generally, training-induced broad cognitive improvements can be caused by either increased WM capacity or enhanced WM efficiency, or a combination of both (for a more detailed discussion, see von Bastian & Oberauer, 2014). Increased capacity is reflected by structural changes and should therefore lead to broad transfer manifested in improved performance across tasks drawing on this capacity-limit. Enhanced efficiency refers to a better exploitation of the capacity available, for example through the use of strategies or a higher level of automatization of WM processes. Different to improvements through enhanced capacity, the range of transfer that follows from enhanced efficiency depends on how efficiency was increased: the less material- or paradigm-specific the improvement, the more transfer should be observed to tasks that draw on the same WM mechanisms. For example, whereas paradigm-specific strategy-use can be expected to transfer to very similar tasks only (for reviews, see Lustig, Shah, Seidler, & Reuter-Lorenz, 2009; Morrison & Chein, 2011), a higher level of automatization should result in transfer to tasks that draw on the same WM mechanisms (cf. von Bastian & Oberauer, 2014).

To disentangle the effects of enhanced capacity and enhanced efficiency, we based the present study on the three-embedded-components model of working memory (Oberauer, 2009; Oberauer & Hein, 2012), which is an extension of the model proposed by Cowan (1995). It assumes three functional levels of information selection, namely the activated part of long-term memory (aLTM), the region of direct access (RDA), and the focus of attention (FoA). The aLTM reflects all representations needed for a current task, activated through perceptual input or spread of activation. In the RDA, a subset of the activated representations is temporarily bound into new structures. In contrast to the aLTM, the capacity of the RDA is limited due to interference between simultaneously maintained bindings (Oberauer, 2005). Lastly, the FoA selects the one item of the RDA that is processed next. According to the three-embedded components model, increased WM capacity would result in an increased number of bindings that can be maintained at a time. Furthermore, the capacity to build and maintain temporary bindings has been hypothesized to be the common limiting factor that explains the strong correlation between working memory and reasoning ("binding hypothesis", cf. Oberauer, Süss, Wilhelm, & Sander, 2007). Thus, increased WM capacity in terms of a larger number of simultaneously maintained bindings should be reflected in broad transfer to other WM tasks, reasoning, and other cognitive abilities (e.g., executive functions) that draw on the same ability of simultaneously maintaining bindings.

Better WM performance could also reflect a more efficient use of basic WM processes that enhance the encoding, maintenance or retrieval of an otherwise stable number of bindings. Here, we focus on three WM mechanisms: enhanced focus switching, removal of no longer relevant information, and interference resolution. Focus switching refers to the ability of the FoA to flexibly shift attention between single items held in the RDA. Reducing the time needed to move the FoA from one item to another increases the time to refresh and, thus, maintain memoranda (Barrouillet, Bernardin, & Camos, 2004), thereby potentially enhancing recall performance. There is some evidence that training can indeed reduce the cost in reaction times (RT) caused by focus switching (e.g., Dorbath, Hasselhorn, & Titz, 2011; Oberauer, 2006; Verhaeghen, Cerella, & Basak, 2004), but it is yet unclear how these improvements relate to transfer effects.

Removal of no longer relevant information is the “unlearning or unbinding of an item from its context” (Ecker, Lewandowsky, & Oberauer, 2014, p. 3). For the proper functioning of working memory, especially the building of new bindings in the RDA, it is essential that outdated information is removed because it would otherwise strongly interfere with the information that is relevant for a current task (Oberauer & Lewandowsky, 2016). More efficient removal is marked by a reduction in the time it takes to remove the no longer relevant information.

Finally, interference resolution is the ability to overcome interference among bindings held in the RDA and is often assumed to be one of the most central mechanisms of transfer (Au et al., 2015; Hussey et al., 2016; Oelhafen et al., 2013). Interference resolution becomes important whenever a conflict in information processing occurs. For example, in a recognition task, conflict occurs when participants are presented with a recent item that, however, was not part of the current memory set; hence the material is highly familiar yet presently irrelevant. This familiarity causes a strong tendency to wrongly identify the information as part of the memory array (cf. Oberauer, 2008). Resolving this interference requires the recollection of the item and its context (i.e., the whole binding, Oberauer, 2005; see also Szmalec, Verbruggen, Vandierendonck, & Kemps, 2011). Thus, enhanced interference resolution is reflected by improved recollection performance.

How Strong is the Evidence for Transfer?

Although methodological issues have been discussed extensively elsewhere (e.g., Shipstead et al., 2013), several problems persist in the training literature, such as the inclusion of no-contact control groups and the measurement of abilities through single tasks (cf. von Bastian, Guye, & De Simoni, 2018; Guye, Röcke, Mérillat, von Bastian, & Martin, 2016). An additional pervasive issue is the low power of most WM training studies due to small sample sizes, with an average group size of n = 20 (Melby-Lervåg et al., 2016). Low statistical power not only increases the probability of false-negative and false-positive findings (e.g., Button et al., 2013), but can also inflate effect sizes. For example, in their simulation study, Halsey, Curran-Everett, Vowler, and Drummond (2015) showed that attempts to detect a true medium effect (Cohen's d = 0.50) with low statistical power (n = 30, theoretical power = 48%) yielded 97% of inflated effects sizes, with the significant effect sizes ranging from d = 0.44 to d = 1.23. As the true size of transfer effects is unknown, we can only speculate about the number of inflated effect sizes in training research; however, meta-analytic effect size estimates are likely to be overestimated (cf. Bogg & Lasecki, 2015; Dougherty, Hamovitz, & Tidwell, 2015; see also von Bastian et al., 2018).

A suitable alternative approach to evaluate the evidence of training and transfer effects is the use of Bayesian inference, where the strength of evidence is expressed by the Bayes factor (BF). The BF is the likelihood of the data under one hypothesis (usually the alternative hypothesis, H_1) relative to the likelihood of the data under the other hypothesis (usually the null hypothesis, H_0, cf. Jeffreys, 1961). In contrast to null hypothesis statistical testing, BFs allow for quantifying evidence not only for the alternative hypothesis (i.e., the presence of training and transfer effects) but also for the null hypothesis (i.e., absence of training and transfer effects). Consequently, BFs are increasingly popular in cognitive training research (e.g., Dougherty et al., 2015; Goghari & Lawlor-Savage, 2017; Guye, De Simoni, & von Bastian, 2017; Guye & von Bastian, 2017; Sprenger et al., 2013; von Bastian & Oberauer, 2013). For example, Dougherty and colleagues (2015) recently reevaluated the 20 n-back training studies included in the meta-analysis by Au et al. (2015) with a Bayesian approach. Out of the 24 comparisons, 11 (i.e., 46%) contributed only ambiguous evidence (BF < 3), indicating that the data from these studies were not sensitive enough to support either hypothesis. Given that the average group size in the included studies was only n = 20, the ambiguity of the results was probably due to low power. Hence, evidence regarding transfer effects of WM training is still inconclusive, with large-scale WM training studies contributing stronger evidence still being needed.

Present Study

In the present study, our goals were to examine (1) the effectiveness of WM training in eliciting near and far transfer effects and, (2), the specific mechanisms underlying training-induced improvements in cognitive performance. We compared two WM training interventions – memory updating and associative binding – to an active control group practicing visual search. Updating training is amongst the most widely used WM training interventions (cf. von Bastian & Oberauer, 2014; e.g., n-back tasks: Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Lilienthal et al., 2013; Redick et al., 2013, keep-track tasks: Dahlin, Neely, Larsson, Backman, & Nyberg, 2008; Dahlin, Nyberg, Bäckman, & Neely, 2008, running-memory tasks: Waris et al., 2015, and memory updating tasks: Linares, Borella, Lechuga, Carretti, & Pelegrina, 2017; Schmiedek, Lövdén, & Lindenberger, 2010). In contrast, there are only few process-based associative binding training interventions, with the few existing studies having focused on older adults (Bellander et al., 2017; Zimmermann, von Bastian, Röcke, Martin, & Eschen, 2016). However, according to the binding hypothesis (Oberauer et al., 2007), a training regimen targeting the ability to simultaneously maintain multiple bindings directly should maximize chance to observe broad transfer.

To evaluate changes in capacity, we assessed near transfer to the respective other set of tasks (i.e., binding tasks served as near transfer tasks for updating training and vice versa), and far transfer to abilities that have been shown to strongly correlate with WM, such as reasoning (e.g., Friedman, Miyake, Schmeichel, & Tang, 2006; Kyllonen & Christal, 1990; Oberauer, Süß, Wilhelm, & Wittmann, 2008; Süß, Oberauer, Wittman, Wilhelm, & Schulze, 2002), shifting and inhibition (e.g., Friedman et al., 2009; Miyake et al., 2000; Miyake & Friedman, 2012), and processing speed (e.g., McAuley & White, 2011; Schmiedek, Oberauer, Wilhelm, Süss, & Wittmann, 2007). Although both WM training paradigms could increase the efficiency of all three WM processes we assessed, we assumed that updating training would put particularly strong demands on focus switching (cf. Oberauer, 2006), followed by removal of outdated information (cf. Ecker et al., 2014; Ecker, Lewandowsky, Oberauer, & Chee, 2010), whereas binding training would emphasize more on interference resolution (e.g., Oberauer, 2005), followed by removal of outdated information. In addition, we explored two alternative mechanisms of change that would boost performance in the trained paradigms, but not transfer to other abilities: systematic shifts in bias towards familiarity-based processing, and the use of paradigm-specific strategies.

We followed the best practice recommendations for training interventions provided by recent reviews (e.g., Noack, Lövdén, & Schmiedek, 2014; Shipstead et al., 2012; Simons et al., 2016; von Bastian & Oberauer, 2014), such as the inclusion of an active control group, the usage of multiple indicators to measure cognitive abilities, and an adequate sample size. First, we included an active control group to differentiate between effects that emerge due to the training intervention and those caused by participating in a study (cf. Klingberg, 2010). Participants of the control group trained visual search tasks, which demand only little WM (e.g., Kane, Poole, Tuholski, & Engle, 2006; Sobel, Gerrie, Poole, & Kane, 2007) and were successfully used in previous training studies (e.g., Foster et al., 2017; Harrison et al., 2013; Redick et al., 2013; von Bastian, Langer, Jäncke, & Oberauer, 2013). To check whether expectations were similar across the experimental and control training groups (Boot, Simons, Stothart, & Stutts, 2013), we asked participants to rate their subjective cognitive improvement. Second, we assessed each cognitive function with four indicators to prevent task-specific features being responsible for the observation of potential transfer effects (cf. Noack et al., 2014; Shipstead et al., 2012). Third, our training groups comprised between 59 and 72 participants, thus our group sample sizes were about three times as large as the size of average treatment groups in WM training research (i.e., n = 20, cf. Melby-Lervåg et al., 2016).

Method

Participants

We recruited young adults between 18 and 36 years for a "cognitive training study" through the participant pool of the Department of Psychology of the University of Zurich, postings at the university campus, and short study presentations in lectures. Inclusion criteria were German native speaker or high proficiency in German, normal or corrected-to-normal vision, no color blindness, no current psychiatric or neurological disorders, and no psychotropic drug use. Participants were reimbursed after posttest completion with CHF 120 (approx. USD 125), or 10 course credits and CHF 20 (approx. USD 21). Moreover, participants received a bonus of up to CHF 50 (approx. USD 52) depending on the level of difficulty they achieved during training (cf. von Bastian & Oberauer, 2013). The experimental protocol was approved by the institutional review board at the University of Zurich in compliance with the Helsinki Declaration. All participants gave written consent to taking part in the study.

As previous studies were likely severely underpowered (cf. Bogg & Lasecki, 2015) and, hence, probably yielded inflated effect size estimates (Halsey et al., 2015), we refrained to use those estimates for power analyses. Instead, we aimed to recruit at least three times as many participants than previous studies (i.e., n = 60 per group). We managed to recruit 241 participants who were randomly assigned to one of the three training groups. The study followed a double-blind design, hence neither the participants nor the experimenter conducting pre- and posttests knew to which group participants were assigned to. As illustrated in Figure 1, N = 233 participants begun with the training intervention of which 216 completed the study. Reasons for the 17 dropouts were lack of time (4), technical problems (8), personal reasons (3), health issues (1), or loss of interest (1). We had to exclude another 19 participants due to a programming error in the updating training intervention (10) or lack of compliance as evidenced by performance below chance level in more than five (i.e., 25%) training sessions (9). The final sample we analyzed consisted of 197 participants. Table 1 lists the descriptive statistics of the demographic variables. The WM training groups were comparable to the active control in terms of gender (updating: $BF_{H1} = 1/4.37 \pm 0.00\%$, binding: $BF_{H1} = 1/4.46 \pm 0.00\%$). Evidence also supported the absence of age differences for the binding relative to the active control group, $BF_{H1} = 1/3.29 \pm 0.00\%$. It was ambiguous though for the updating group, $BF_{H1} = 1/1.13 \pm 0.00\%$, with participants in the updating group being, on average, 1 year younger than participants in the active control group.

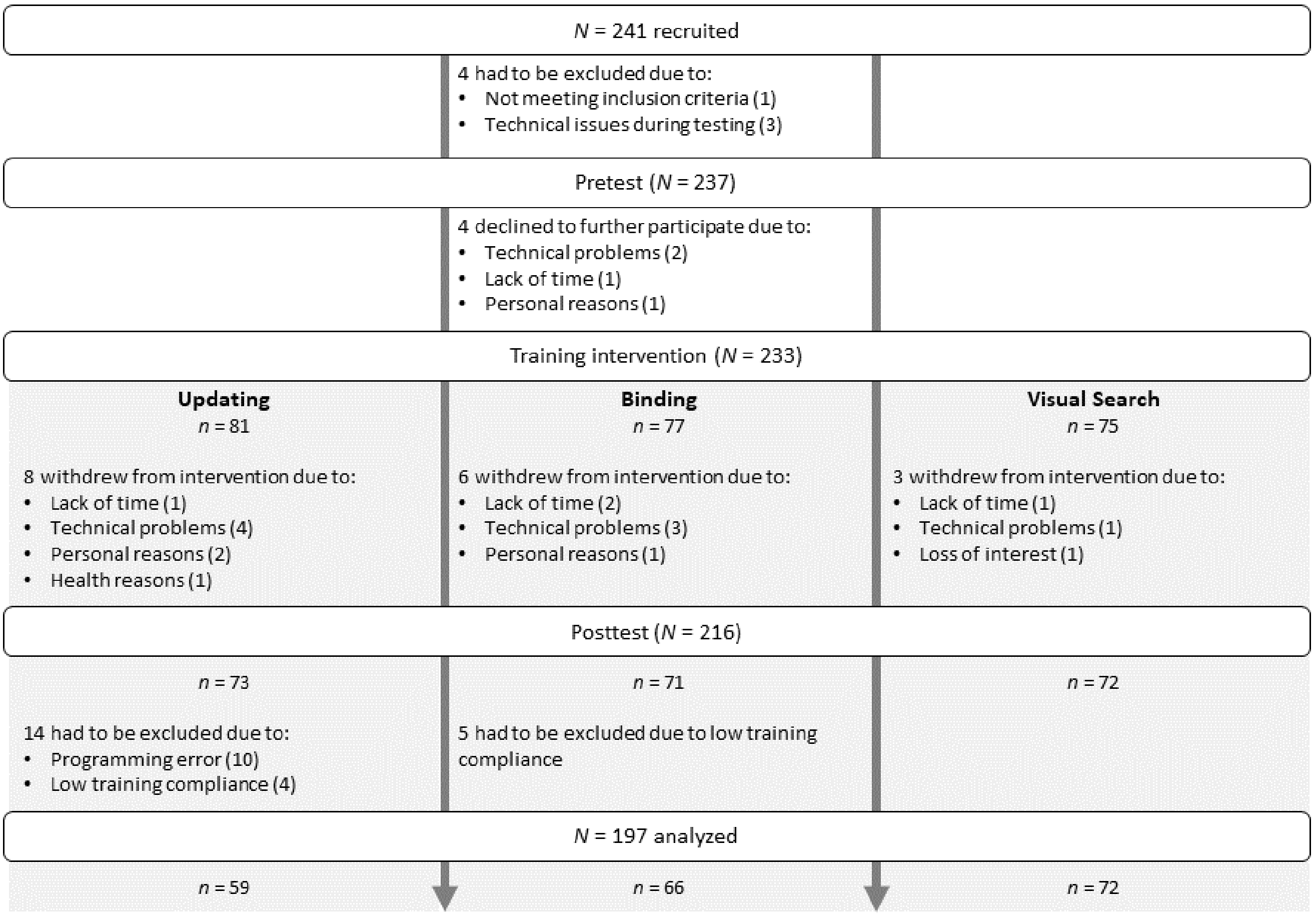

Figure 1. Flowchart of participant recruitment and study completion.

Table 1
Participant Demographics

Measure	Group: Updating	Binding	Active Control
Sample size (n)	59	66	72
Gender (f/m)	40/19	45/21	49/23
Age (M, SD)	22.61 (2.97)	24.55 (4.05)	23.81 (4.16)

Note. BFs indicated support for the null hypothesis that there were no group differences as determined by Bayesian Pearson chi-square test (gender) and Bayesian two-tailed t-tests (age).

Design and Materials

Participants completed 20 training sessions of extensive cognitive training over the course of 5 weeks. Training and transfer effects were assessed by administering a test battery before and after training.

Training. Participants completed training at their own computer or laptop at home using Tatool (von Bastian, Locher, & Ruflin, 2012), a Java-based open-source training and testing tool (www.tatool.ch). After each session, data were automatically uploaded to a web server running Tatool Online. An experimenter, who was not involved in the collection of outcome measures, monitored participants' training performance and served as contact person during training. As in previous work (von Bastian & Oberauer, 2013; von Bastian et al., 2013; von Bastian & Eschen, 2016), we aimed to maximize experimental control through automated online analyses to detect irregularities (e.g., performance below chance level). To increase individual commitment, participants (1) signed a participation agreement at the beginning of the study, (2) were made aware of their progress being constantly monitored, (3) received regular emails (i.e., after 2 and 4 weeks of training) on their training progress, and (4) were reminded to practice when falling behind their training schedule (i.e., less than four sessions completed per week).

Each group practiced four tasks with varying material (numerical, verbal, visual, and spatial, see Figure 2 for an illustration and Table 2 for details) for approximately 10 min each per training session. The order of task administration was randomized for each training session. For all groups, task duration was restricted to a maximum of 11.25 min so that a training session did not exceed 45 min. The updating group completed up to 12 trials, the binding group up to 24 trials, and the active control group up to 100 trials per task and session.

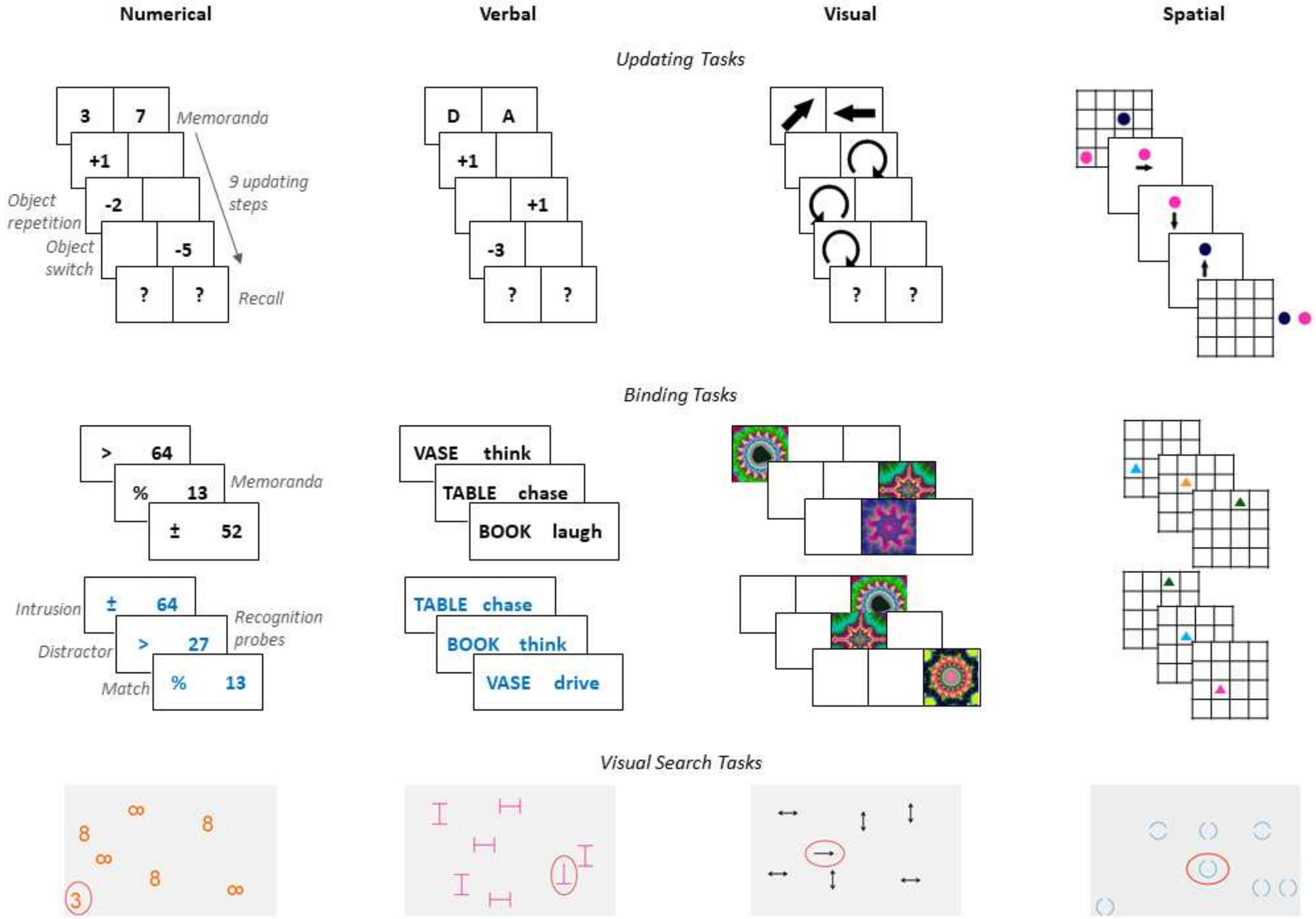

Figure 2. Schematic depiction of the four training tasks administered in each training group. In the updating tasks (top row), participants had to memorize an initial set of memoranda, update those memoranda according to specific operations, and then maintain the new result. Updating steps could require updating the same memory object as in the preceding step (object repetition) or shifting the focus of attention to another

object (object switch). In the binding tasks (middle row), participants had to recognize previously memorized items and their current context. After a set of memoranda, participants were presented recognition probes with exactly the same context (matches), a different context (e.g., at a wrong location, intrusions), or that were entirely new (distractors). In the visual search tasks (bottom row), participants had to identify the odd-one-out object (highlighted in red for illustration purposes).

Table 2
Description of the Training and Transfer Tasks

Task	Description	# Trials	Conditions (%)	Set Sizes	Timing
Updating					
Digits (n)	Memorize digits, update by applying simple arithmetic operations (adapted from Lewandowsky et al., 2010).	16	50 switches 50 repetitions	3, 5	Stimulus: 500 ms Cue: 500 ms Operation: ur
Letters (ve)	Memorize letters, update by shifting forward or backward in the alphabet (adapted from Lewandowsky et al., 2010).	16	50 switches 50 repetitions	2, 4	Stimulus: 500 ms Cue: 500 ms Operation: ur
Arrows (vi)	Memorize arrows, update by rotating 45 degrees clockwise or counter-clockwise (adapted from Harrison et al., 2013).	16	50 switches 50 repetitions	2, 4	Stimulus: 500 ms Cue: 500 ms Operation: ur
Locations (s)	Memorize colored circles in a 4 x 4 grid, update by shifting one cell up, down, left, or right (adapted from Lewandowsky et al., 2010).	16	50 switches 50 repetitions	3, 5	Stimulus: 500 ms Cue: 500 ms Operation: ur
Binding					
Symbol-Digit (n)	Recognize pairings of mathematical symbols and two-digit numbers (adapted from Wilhelm et al., 2013).	24	50 matches 25 intrusions 25 distractors	4-6	Stimulus: 900 ms probe: ur
Noun-Verb (ve)	Recognize pairings of nouns and verbs (adapted from Wilhelm et al., 2013).	24	50 matches 25 intrusions 25 distractors	4-6	Stimulus: 1800 ms probe: ur
Fractal-Location (vi)	Recognize pairings of fractals and their location in a row of boxes (adapted from Oberauer, 2005).	24	50 matches 25 intrusions 25 distractors	4-6	Stimulus: 900 ms probe: ur
Color-Location (s)	Recognize pairings of colored triangles and their locations in a 4 x 4 grid (adapted from Oberauer, 2005).	14	50 matches 25 intrusions 25 distractors	4-6	Stimulus: 1800 ms probe: ur
Visual Search					
Numbers (n)	Search for a “3” among horizontally and vertically presented “8”s (adapted from Kane et al., 2006).	30	80 target present 20 target absent	7, 9, 11	ur
Letters (ve)	Search for a “T” among horizontally and vertically presented “I”s (adapted from Harrison et al., 2013).	30	80 target present 20 target absent	7, 9, 11	ur

Arrows (vi)	Search for a single-headed arrow among double-headed arrows (adapted from Kane et al., 2006).	30	80 target present 20 target absent	7, 9, 11	ur
Circles (s)	Search for a circle with one gap among circles with two gaps (von Bastian et al., 2013).	30	80 target present 20 target absent	7, 9, 11	ur
			Removal		
Digits (n)	Memorize and substitute digits (adapted from Ecker et al., 2010).	18	50 long CTI 50 short CTI	3	Stimulus: 500 ms long cue: 1800 ms short cue: 200 ms substitution: ur
Letters (ve)	Memorize and substitute letters (adapted from Ecker et al., 2010).	18	50 long CTI 50 short CTI	3	Stimulus: 500 ms long cue: 1800 ms short cue: 200 ms substitution: ur
Arrows (vi)	Memorize and substitute arrows (adapted from Ecker et al., 2010).	18	50 long CTI 50 short CTI	3	Stimulus: 500 ms long cue: 1800 ms short cue: 200 ms substitution: ur
Locations (s)	Memorize and substitute colored circles in a 4 x 4 grid (adapted from Ecker et al., 2010).	18	50 long CTI 50 short CTI	3	Stimulus: 500 ms long cue: 1800 ms short cue: 200 ms substitution: ur
			Reasoning		
Diagramming Relationships (ve)	Identify the Venn diagram out of five options that best describes the semantic relationship between three nouns (e.g., “animals, cats, and dogs” would be best represented by one circle corresponding to “animals” containing two separate circles for “cats” and “dogs”; Ekstrom, French, Harman, & Dermen, 1976).	30		-	Time limit: 8 min
Letter Sets (ve)	Identify the letter set that deviates from the logical pattern underlying the other four sets (Ekstrom et al., 1976).	30		-	Time limit: 14 min
Locations Test (vi)	Based on the logical pattern underlying the spatial distribution of “x”s on rows of dashes, identify at which of five locations the next “x” has to be placed (Ekstrom et al. 1976).	28		-	Time limit: 12 min
RAPM (s)	Complete a pattern by choosing 1 out of 8 options (Arthur & Day, 1994; see also Raven, 1990).	28		-	Time limit: 15 min
			Shifting		

Parity-Magnitude (n)	Classify (1-9, excluding 5) as odd or even, or smaller or larger than 5 (von Bastian et al., 2016).	Single: 2 x 64 Mixed: 129	50 switches 50 repetitions	-	ur
Animacy-Size (ve)	Classify line-drawings of animals and objects as living or non-living, or smaller or larger than a soccer ball (von Bastian et al., 2016).	Single: 2 x 64 Mixed: 129	50 switches 50 repetitions	-	ur
Color-Shape (vi)	Categorize figures according to their color (blue or green) or their shape (round or angular, von Bastian et al., 2016).	Single: 2 x 64 Mixed: 129	50 switches 50 repetitions	-	ur
Fill-Frame (s)	Classify whether a geometric figure is dotted or striped, or framed or not (von Bastian & Oberauer, 2013).	Single: 2 x 64 Mixed: 129	50 switches 50 repetitions	-	ur
		Inhibition			
Number Stroop (n)	Indicate which of two digits has the higher value. In congruent trials, differences in value and physical size match (i.e., digits with higher value are displayed larger), in incongruent trials they mismatch, and in neutral trials, both digits have the same size (Tzelgov, Meyer, & Henik, 1992).	288	33 congruent 33 incongruent 33 neutral	-	ur
Color Stroop (ve)	Indicate the hue of color words. In congruent trials, the hue matches the color word, in incongruent trials it does not, and in neutral trials, colored "X"s are presented (Stroop, 1935).	288	33 congruent 33 incongruent 33 neutral	-	ur
Global-Local (vi)	Indicate whether the small (local) shapes making up a bigger (global) shape are circles or squares. In congruent trials, the local and global shape match (e.g., a circle made up of small circles), in incongruent trials, they mismatch, and in neutral trials, line drawings of circles or squares are presented (Navon, 1977).	288	33 congruent 33 incongruent 33 neutral	-	ur
Simon (s)	Indicate the color of a green or red circle presented on the left, right or centrally by pressing the left arrow key for green and the right arrow key for red. In congruent trials, the location of the circle and the location of the response key match (e.g., a green circle is presented on the left), in incongruent trials they mismatch, and in neutral trials, the circle is displayed centrally (Simon, Sly, & Vilapakkam, 1981).	288	33 congruent 33 incongruent 33 neutral	-	ur

Note. All tasks included two practice trials (12 in the shifting and inhibition tasks) to familiarize participants with the tasks. Set size reflects the number of memoranda. N = numerical, ve = verbal, vi = visual, s = spatial, ur = unrestricted (until response); RAPM = Raven's Advanced Progressive Matrices.

Updating training. In the four memory updating tasks (adapted from Lewandowsky, Oberauer, Yang, & Ecker, 2010), participants had to remember an initial set of simultaneously presented stimuli. During the updating phase, participants had to transform individual stimuli (e.g., mentally move previously memorized circles in a grid or applying a simple arithmetic operation to a digit), enter the result of the transformation, and remember the result. In half of the trials, a cue indicated which of the stimuli had to be updated next. Half of the updating steps were switching trials (i.e., the to-be-updated stimulus was different from the one in the preceding updating step) and the other half were repetition trials (i.e., the to-be-updated stimulus was the same as the one in the preceding updating step). After nine updating steps, participants had to recall the most recent result of each stimulus.

Binding training. In the four associative binding tasks (adapted from Oberauer, 2005; Wilhelm, Hildebrandt, & Oberauer, 2013), participants had to memorize sequentially presented associations of two elements (e.g., pairings of symbols and digits, or fractals and their location). In the subsequent recognition phase, each association was randomly probed with one of the elements given as cue. Of the probes, 50% were positive (i.e., exact matches), 25% were intrusions (i.e., probes that were presented in the current trial, but associated with a different element or location), and 25% were distractors (i.e., probes not presented in the current trial).

Active control training. In the four visual search tasks, participants had to search for a target (e.g., a single-headed arrow) among distractors (e.g., double-headed arrows) and, if the target was present, to indicate whether it faced up, down, left, or right by pressing the corresponding arrow key (cf. von Bastian et al., 2013). In target-absent trials, participants had to press "A" instead. The search display was a warped 7 x 8 grid, generating a scattered distribution of stimuli on the screen (cf. Guye & von Bastian, 2017).

Adaptive training algorithm. Each training task's difficulty was adjusted to individual performance. The updating and binding training tasks started with a set size (i.e., number of memoranda) of two and a maximal response time limit of 3500 ms, and the active control training tasks with a search display containing six stimuli. The first training session served to evaluate participants' individual cognitive performance limit (cf. Guye & von Bastian, 2017). In each task, participants' mean accuracy was measured after every 10% of trials (1 trial in the updating tasks, 2 trials in the binding tasks, and 10 trials in the visual search tasks). If participants scored at least 85% correct, difficulty was increased, otherwise it remained on the current level. For the remaining 19 training sessions, participants' performance was checked after every 40% of trials (5 trials in the updating tasks, 10 trials in the binding tasks, and 40 trials in the visual search tasks). In the updating and binding tasks, difficulty was adjusted by decreasing the response time limit by 500 ms for four subsequent level-ups (e.g., when reaching levels 2 through 5), and by increasing set size by one additional memorandum every fifth level-up (e.g., when reaching level 6). The response time limit was reset to 3500 ms whenever the set size was increased and again reduced by 500 ms for the subsequent four level-ups. In the active control group, level-ups corresponded to an increase in set size (i.e., number of stimuli shown in the search display) by one additional stimulus. We refrained from adjusting response time in the active control group, to minimize demands on processing speed, which is strongly correlated with WM and reasoning (cf. Schmiedek et al., 2007). Participants were informed when they reached a higher difficulty level (i.e., "Congratulations, you achieved the next level"), and they started each session on the highest level they achieved in the preceding session (cf. von Bastian et al., 2013).

Questionnaires. To examine whether the training groups were similarly motivated during training, we asked participants after each training session to rate the enjoyment experienced ("Today's training session was fun to do") and the effort spent during the training session ("I tried to do well in today's training session"; both items adapted from the Intrinsic Motivation Inventory, IMI, Deci & Ryan, 2015), and the perceived fit between difficulty and ability ("The difficulty of today's training session was just right"; cf. von Bastian & Eschen, 2016) on a scale ranging from 1 (does not apply at all) to 7 (does apply very well).

In addition, participants were asked to monitor their learning progress during training by answering two questions at the beginning of each training session (i.e., "how would you rate your performance in the last training session?", "how well will you perform in today's training session?") and one at the end of each session (i.e., "how would you rate your performance in today's training session?"). These data will be reported elsewhere. Furthermore, participants completed the German version of the Questionnaire on Current Motivation (QCM, Rheinberg, Vollmeyer, & Burns, 2001) after training sessions 1 and 10, and the IMI (Deci & Ryan, 2015) after training session 20. These data have been reported in Guye et al. (2017). Finally, after training session 20, participants completed a questionnaire on strategy-use and training-related expectations. Strategy-use was assessed with a dichotomous item on whether they used strategies to complete the tasks and, if so, to briefly describe the strategy (primarily) used, and to rate how useful they were on a scale from 1 (not at all) to 5 (very). To assess training-related expectations, participants reported on a scale ranging from 1 (not at all) to 5 (very) whether they believed that they improved in the trained tasks, in the untrained cognitive tasks, and in everyday life.

Pre- and posttest. Before and after the training intervention, we assessed practice, near and far transfer, and mechanism-specific effects with a test battery of 28 computer-based tasks. We used the same test battery for pre- and posttest to facilitate between-groups baseline comparisons, required for investigating the comparability across groups and occasions (cf. Guye & von Bastian, 2017; von Bastian & Oberauer, 2013; von Bastian & Eschen, 2016). Up to four participants were tested simultaneously in a single lab session that lasted up to 5 h including three short breaks. To control for fatigue and practice effects, half of the participants in each group completed the test battery in reversed order (i.e., they started with the forward's order last task of the last block and finished the test battery with the first task of the first block, cf. von Bastian & Oberauer, 2013). To familiarize participants with the tasks, several practice trials were presented before test blocks of pseudorandomized trials. Each cognitive ability was assessed with four tasks with varying material. Table 2 lists the details for each task regarding stimuli, conditions, number of trials, set sizes, and timing parameters.

Before the pretest, participants completed a series of questionnaires: NEO-FFI (Borkenau & Ostendorf, 2008); Need for Cognition (Bless, Wänke, Bohner, Fellhauer, & Schwarz, 1994), Theories of Intelligence Scale (Dweck, 2000), Grit Scale (Duckworth & Quinn, 2009), Self-Efficacy Scale (Schwarzer & Jerusalem, 1995), Self-Efficacy to Regulate Exercise (Bandura, 2006). Findings from these measures are reported in Guye et al. (2017).

Training. To compare practice effects between training conditions, we administered test versions of the training tasks at pre- and posttest that were identical in structure and material to the training tasks above. All participants completed the same set of trials. Different to the training tasks, trials varied regarding set sizes, but with fixed timing parameters. The proportion of correctly recalled items at the end of each trial served as dependent variable for the updating tasks. The discrimination parameter d' from signal detection theory served as outcome measure for the associative binding tasks. It was computed from subtracting the z-transformed false alarms to intrusion probes from the z-transformed hit rates (cf. Oberauer, 2005). Participants' ability to search through a visual display was measured by the individual residuals from a simple linear regression model predicting the RTs of target-absent trials from the RTs of target-present trials (for a similar approach in dual tasks, see Oberauer, Lange, & Engle, 2004). Hence, for the visual search tasks, better performance was reflected by lower values.

Near transfer. To assess near transfer, we examined whether participants of the updating group showed transfer to the binding tasks and vice versa. The mean recall accuracy of the updating tasks and the discrimination parameter d' computed from the binding tasks served as dependent variables.

Far transfer. The four reasoning tasks required participants to either detect a rule behind a pattern or to integrate information for drawing a conclusion. The proportion of correctly answered items relative to the total number of items served as outcome measure. In the four shifting tasks, participants had to categorize bivalent stimuli according to one of two classification rules as indicated by a cue, which was displayed 150 ms before stimulus onset (cf. von Bastian, Souza, & Gade, 2016). Each task consisted of five blocks: two single-rule blocks (only one rule had to be applied, e.g., animacy classification followed by size classification), a mixed-rules block (two rules switched randomly, e.g., switching between animacy and size classifications), and another two single-rule blocks in reversed order (e.g., size classification followed by animacy classification). Half of the trials in the mixed-rules block were switching trials (i.e., the rule was different than the one in the preceding trial) and the other half were repetition trials (i.e., the rule was the same as the one in the preceding trial; the first trial was excluded from analyses as it constitutes neither a switch nor a repetition). Switching costs were calculated by subtracting RTs to repetition trials from RTs to switch trials in the mixed-rules block. Processing speed was measured by the average RTs in single-rule blocks of the shifting tasks. In the four inhibition tasks, participants were required to inhibit prepotent responses. These tasks comprised three conditions: a congruent condition (correct and prepotent response correspond), an incongruent condition (correct and prepotent response do not correspond) and a neutral condition (no prepotent response present). Each condition appeared equally often (cf. von Bastian et al., 2016). Stimuli of each condition were randomly presented within four blocks of 72 trials. Interference costs were computed by subtracting the RTs to neutral trials from RTs to incongruent trials.

WM processing. To assess focus switching, we used the RTs in the updating steps of the updating training tasks. Specifically, we computed the RT cost of having to switch between memory objects in two subsequent updating steps relative to object repetitions by subtracting RTs to repetition trials from RTs to switch trials (Oberauer & Hein, 2012). Thus, lower values reflected lower costs of switching the focus, which we interpreted as better performance.

Removal speed was measured with four modified memory updating tasks modeled after Ecker et al. (2014). As in the updating tasks, participants had to memorize an initial set of stimuli. In the subsequent updating phase, however, individual stimuli were substituted by new ones and participants had to press the space bar (or click a button in the spatial version of the task, cf. Table 2) as soon as they had memorized the new stimulus. A cue presented for either 200 ms (i.e., short cue-target interval, CTI) or 1500 ms (i.e., long CTI) indicated which stimulus was updated next. In contrast to the updating task, this paradigm consisted of switching trials only (i.e., the to-be-updated stimulus was always different from the one in the preceding updating step). After 1 to 18 updating steps, participants had to recall the most recent stimuli. Individual residuals from a simple regression model predicting the RTs of trials with short CTIs from RTs of trials with long CTIs were used as dependent variables (Ecker et al., 2010). Thus, lower values reflected more efficient removal from WM.

Finally, interference resolution performance was assessed by the WM training tasks. From the updating tasks, we extracted the proportion of transposition errors (i.e., recalling an item from the current trial but at a wrong position). From the binding tasks, we used the proportion of correct responses to intrusion probes.

Alternative mechanisms. To analyze the patterns of errors committed in the trained tasks, we additionally extracted the proportion of extra-list errors (i.e., wrongly recalled items not in the memory array) in the updating tasks, and the proportion of correct responses to matching and distractor probes in the binding tasks. As changes in accuracy towards probes in the binding tasks could reflect systematic shifts in response bias (e.g., saying NO more often would result in more misses and, so, in lower accuracy in matching probes, but higher accuracy in intrusion and distractor probes), we computed the criterion C by multiplying the sum of the z-transformed hit rate and z-transformed false alarm rate by -0.5. Values in the positive range reflect biases towards saying NO, and values in the negative range reflect biases towards saying YES. The higher the value, the stronger the bias. Self-reported strategies were categorized by each author separately, with an agreement rate of 86.18%. Discrepancies were resolved through discussion.

Analyses

Analyses were undertaken in four steps. First, we evaluated the proposed theoretical transfer model by fitting a measurement model to the pre- and posttest data. Second, we examined change over the course of the 20 training sessions in terms of the training level achieved and motivation. Third, we investigated gains from pre- to posttest for each experimental group relative to the control group. Specifically, we analyzed (1) practice effects on the trained tasks, (2) near and far transfer, and (3) changes in the proposed WM mechanisms of transfer (i.e., focus switching, removal, and resolution of interference). Finally, to explore alternative mechanisms of change in the trained abilities, we analyzed the response patterns and participants' self-reported strategy use. All analyses were conducted in R (R Core Team, 2015).

Bayesian analyses. We used the "BayesFactor" package (Morey & Rouder, 2015) with the default prior settings (i.e., Cauchy distribution with a scaling factor $r = 0.707$) to compute BFs. BFs range on a continuous scale from 0 to ∞, with a BF of 1 reflecting perfect ambiguity (i.e., the data support both hypotheses equally). BFs below 1 represent evidence for the hypothesis in the denominator (typically H_0), and BFs above 1 indicate evidence in favor of the hypothesis in the numerator (typically H_1). For example, a BF of 10 in favor of the H_1 means that the data are ten times more likely under H_1 than H_0. To facilitate interpretation, Table 3 lists verbal labels adapted from Wetzels and Wagenmakers (2012).

Table 3
Verbal Labels for Interpreting Bayes Factors

Bayes Factor		
H_1	H_0	Interpretation
> 100	< 1/100	Decisive
30 to 100	1/100 to 1/30	Very strong
10 to 30	1/30 to 1/10	Strong
3 to 10	1/10 to 1/3	Substantial
1 to 3	1/3 to 1	Ambiguous
1	1	No evidence

Note. Adapted from Wetzels and Wagenmakers (2012).

Baseline comparisons, training performance and motivation, and gains from pretest to posttest were analyzed with Bayesian linear mixed-effects (LME) models across each of the four tasks measuring the same ability using the lmBF() function of the BayesFactor package. LME models have the advantage that they simultaneously account for multiple sources of variance in the data. Two types of effects are distinguished in LME models: fixed effects (e.g., variance from experimental conditions or predictors) and random effects (e.g., variance from individual differences). We included participant and task as crossed-random effects (Baayen, Davidson, & Bates, 2008) to account for the fact that both participants and tasks included in our study are random samples drawn from larger populations. Hence, although this procedure does not specifically model latent change, it does allow for analyzing effects on the level of the assessed ability by modeling the task-specific variance as nuisance (cf.

Guye & von Bastian, 2017; von Bastian & Eschen, 2016; von Bastian & Oberauer, 2013; Zimmermann et al., 2017).

Confirmatory factor analysis (CFA). To investigate the relationship between the measured cognitive abilities at pre- and posttest, we conducted a latent-variable CFA using the "lavaan" package (Rosseel, 2012). We examined model fit evaluating the chi-square statistic (χ^2), the comparative fit index (CFI), the root mean-squared error of approximation (RMSEA) and its 90% confidence interval (CI), and the standardized root mean-squared residual (SRMR). Good fit is indicated by values above .95 for CFI, values less than 0.06 for RMSEA, and values below .08 for SRMR (Hu & Bentler, 1999).

Data preprocessing. Only RTs of correct responses were analyzed. RTs being 3 median absolute deviations away from the overall median (Leys, Ley, Klein, Bernard, & Licata, 2013) or shorter than 200 ms were defined as outliers and excluded from analyses. To reduce positive skew of speed-based outcome measures, we log-transformed RTs. All dependent variables were z-transformed across the three groups. To eliminate variance due to the two different orders of test administration in pre- and posttest, we arbitrarily selected one order as the reference condition and corrected the data of the other order for the mean difference between the two orders for each variable (cf. von Bastian & Druey, 2017; von Bastian & Oberauer, 2013).

Missing data. For each task, we excluded participants who showed signs of non-compliance (i.e., mean accuracy below chance level and proportion of RTs below 200 ms > M proportion + 3 SD). This concerned individuals in 13 tasks at pretest and 17 tasks at posttest (see Table A1 in the Appendix). As these data were not missing at random, we refrained from imputing those data and instead excluded the affected participants listwise from analyses including those measures. Seven participants (1 in the active control and 6 in the binding group) had difficulties pursuing their training schedule and, hence, completed only 19 (6 participants) or 18 (1 participant) sessions. Only participants with complete training data sets were included in the analyses of training progress and motivation. Three participants (one of each group) did not complete the questionnaire at the end of training; hence, these participants were excluded from the analyses of training-related expectations strategy usage.

Results

Data and scripts for running the analyses are available on the Open Science Framework (https://osf.io/fy5ku). Descriptive statistics and reliabilities for each outcome measure as a function of group and time are listed in Table A2 in the Appendix; between-task correlations at pretest and posttest are listed in Table A3 in the Appendix.

Baseline Comparability

We compared pretest performance of each experimental group with the active control group for each ability (Table 4). The evidence consistently supported the absence of baseline differences but was ambiguous in some instances. Specifically, at pretest, the active control group tended to show larger shifting switch costs than both experimental groups, longer RTs in the speed tasks than the binding group, and to slightly larger interference costs in the inhibition tasks than the updating group. Moreover, evidence consistently supported the absence of baseline difference in all updating response types, binding probe types, and binding bias ($BF_{H1} \geq 1/3.41$) but was ambiguous for updating extra-list errors, $BF_{H1} \geq 1/2.02 \pm 1.96\%$, with extra-list errors tending to be higher in the active control group than in the updating group. Therefore, we cannot exclude that some level of regression to the mean may have occurred for these outcomes. To reduce the impact of those baseline differences, we used standardized gain scores (i.e., mean of posttests scores minus mean of pretest scores divided by the pretest standard deviation; cf. Guye & von Bastian, 2017; von Bastian & Oberauer, 2013; von Bastian & Eschen, 2016) for each participant and each task.

Table 4
Baseline Group Comparability in Cognitive Abilities and Working Memory Processing

Ability	Updating vs. Active Control	Binding vs. Active Control
	Training and Near Transfer	
Updating	1/3.22 ± 2.17	1/3.35 ± 1.76
Binding	1/6.26 ± 1.24	1/6.94 ± 2.65
Visual Search	1/4.70 ± 4.42	1/4.68 ± 4.96
	Far Transfer	
Reasoning	1/6.94 ± 1.76	1/4.29 ± 1.37
Shifting	1/1.27 ± 1.32	1/2.02 ± 2.15
Speed	1/4.07 ± 1.38	1/1.90 ± 1.13
Inhibition	1/1.38 ± 1.41	1/8.15 ± 4.68
	Working Memory Processing	
Focus Switching	1/11.02 ± 2.16	1/11.36 ± 1.40
Removal	1/4.59 ± 2.30	1/10.64 ± 3.50
Interference Resolution		
Updating	1/8.69 ± 2.85	1/5.05 ± 2.98
Binding	1/8.50 ± 1.59	1/8.71 ± 1.55

Note. Values are Bayes factors in favor of the alternative hypothesis and their estimation error (%).

Evaluation of the Theoretical Transfer Model

To evaluate the theoretical transfer model, we conducted a latent-variable CFA. We first examined the proposed factor structure of seven correlated but separate cognitive abilities (i.e., updating, binding, visual search, reasoning, shifting, and general speed) by fitting the model to the data from all participants (excluding those with incomplete data) simultaneously at pretest (N = 185) and posttest (N = 180). While imposing the same factorial structure at both times of assessment (i.e., configural invariance), factor loadings, intercepts, and residual variances were allowed to vary freely between pretest and posttest. Likely due to their low zero-order correlations (see Table A3 in the Appendix), the four inhibition measures did not converge to a latent factor at neither time of assessment. We therefore excluded the inhibition factor from the model. The model including the remaining six factors for updating, binding, visual search, reasoning, shifting, and general speed fit the data reasonably well, $\chi^2(466)$ = 740.43, $p < .001$, CFI = .93, RMSEA = .06 [.05; .06], SRMR = .06. As depicted in Figure 2, all tasks loaded significantly on their respective factor. Moreover, all latent factors exhibited significant variance and correlations between the latent factors were significant, except for visual search, which showed no significant relation to any other factor except general speed at pretest. More specifically, at pretest, the two working memory factors, updating and binding, were strongly related (coefficient estimate = .91), and both factors showed moderate correlations with reasoning (coefficient estimates = .66 and .57, for updating and binding, respectively). At posttest, the pattern was the same, but the correlation between updating and binding was lower (coefficient estimate = .75), and the correlations to reasoning were stronger (coefficient estimate = .83 and .59, for updating and binding, respectively). As expected, the updating, binding, and reasoning factors were moderately related to general speed, with coefficient estimates between -.36 and -.48 at pretest, and between -.27 and -.37 at posttest. Furthermore, updating showed a weak, but significant correlation with shifting at pretest (coefficient estimate = -.18) which was non-significant at posttest (coefficient estimate = -.15).

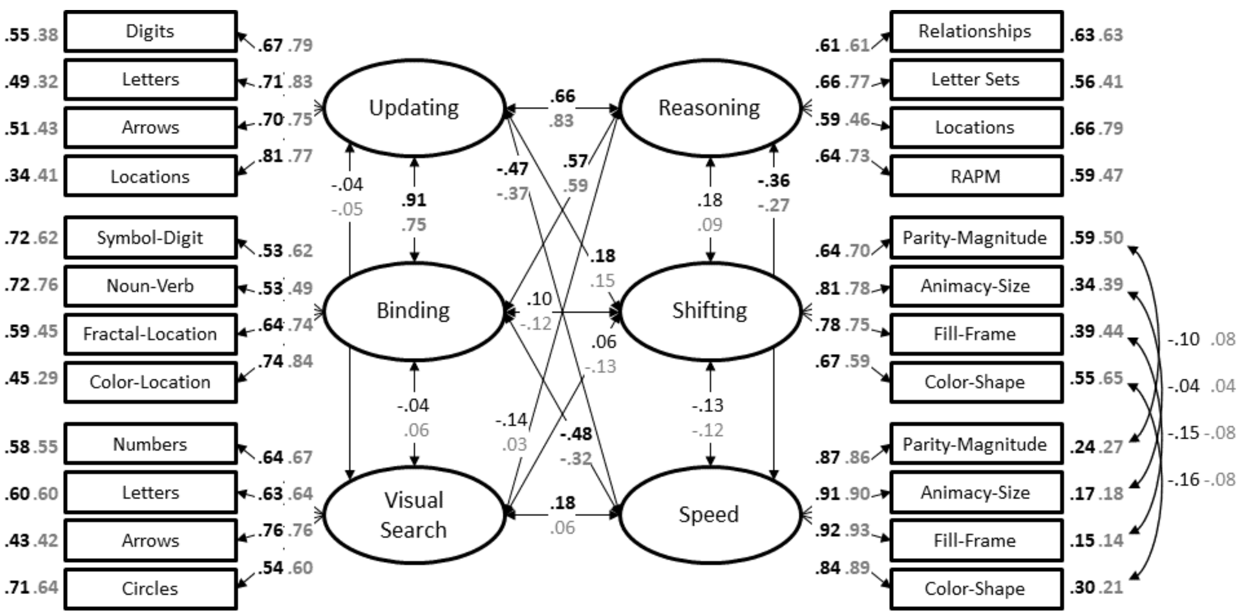

Figure 3. Measurement model for the theoretical transfer model for pretest data (printed in black) and posttest data (printed in gray). Rectangles represent manifest variables and ellipses latent factors. Single-headed arrows represent linear regressions and double-headed arrows correlations. Bold numbers indicate significance ($p < .05$). All latent factor variances were significant ($p < .05$). Residual errors of the shifting and speed variables were correlated,

because they originated from the same tasks. Note that lower values reflect better performance for visual search (i.e., lower residual search costs), shifting (i.e., lower task switching costs), and speed (i.e., shorter reaction times).

To examine whether the differences between the pretest and posttest model were statistically significant, we tested for measurement invariance across pretest and posttest. We first tested for weak invariance by constraining factor loadings to be equal across time, and compared the model fit to the baseline model with configural invariance by inspecting the differences in CFI and RMSEA in addition to running χ^2 difference tests (cf. Cheung & Rensvold, 2002). The loss in model fit caused by constraining factor loadings to be equal was negligible, $\Delta\chi^2 = 10.05$, $\Delta df = 18$, $p = .930$, $\Delta CFI < .01$, $\Delta RMSEA < .01$. Next, to test for strong invariance, we additionally constrained the intercepts to be equal across time, which again did not result in a significant loss of fit, $\Delta\chi^2 = 0.78$, $\Delta df = 18$, $p > .999$, $\Delta CFI < .01$, $\Delta RMSEA < .01$. Finally, we tested for strict invariance by additionally constraining the residuals to be equal across time. The loss in fit was again non-significant, $\Delta\chi^2 = 28.32$, $\Delta df = 24$, $p = .247$, $\Delta CFI < .01$, $\Delta RMSEA < .01$.

As we were able to establish strict measurement invariance, we next tested whether covariances and variances of the latent variables were also invariant across pretest and posttest. The loss of fit was non-significant, $\Delta\chi^2 = 14.56$, $\Delta df = 15$, $p = .483$, $\Delta CFI < .01$, $\Delta RMSEA < .01$. Finally, the loss in fit from constraining the latent factor variances to be equal was also negligible, $\Delta\chi^2 = 4.13$, $\Delta df = 6$, $p = .659$, $\Delta CFI < .01$, $\Delta RMSEA < .01$. Overall, this final model produced a slightly better fit than the configural baseline model, $\chi^2(547) = 798.27$, $p < .001$, CFI = .94, RMSEA = .05 [.04; .06], SRMR = .06. Hence, the most parsimonious model with strict measurement invariance with additional invariance of latent factor covariances and variances was retained, rendering the pretest to posttest differences in coefficient estimates negligible.

Summary. The CFA supported the theoretical assumption that updating and binding are highly correlated abilities unrelated to visual search. Following the rationale that transfer is driven by functional overlap between abilities, the factor correlations observed in our pretest and posttest data suggest that transfer is most likely to occur from one WM training intervention to the other (i.e., near transfer), followed by transfer to reasoning and speed (i.e., far transfer). In contrast, it should be unlikely that transfer from WM training to shifting and visual search would occur. Similarly, as we found only little shared variance between the visual search latent factor shares and all other latent factors, visual search training should have little impact on performance in tasks measuring any of the other abilities.

Training Performance and Motivation

To investigate performance gains over the course of the 20 training sessions, we ran Bayesian LMEs for each group using the set size (coded as linear contrast) achieved by the end of each training session as dependent variable, session as fixed effect, and participant and task as random effects. The reported estimates are means of the sampling from the posterior distribution with 10,000 iterations and reflect the increase in set size from one session to the next one alongside their 95% credible interval. As Figure 4 shows, all three groups improved substantially over the course of the 20 sessions. Evidence for monotonic increases in performance was decisive for the updating group, $M_{Diff} = 0.09$ [0.09, 0.09], $BF_{H1} > 100 \pm 1.22\%$, for the binding group, $M_{Diff} = 0.10$ [0.10, 0.11], $BF_{H1} > 100 \pm 1.80\%$, and for the visual search group, $M_{Diff} = 2.06$ [2.05, 2.08], $BF_{H1} > 100 \pm 2.70\%$.

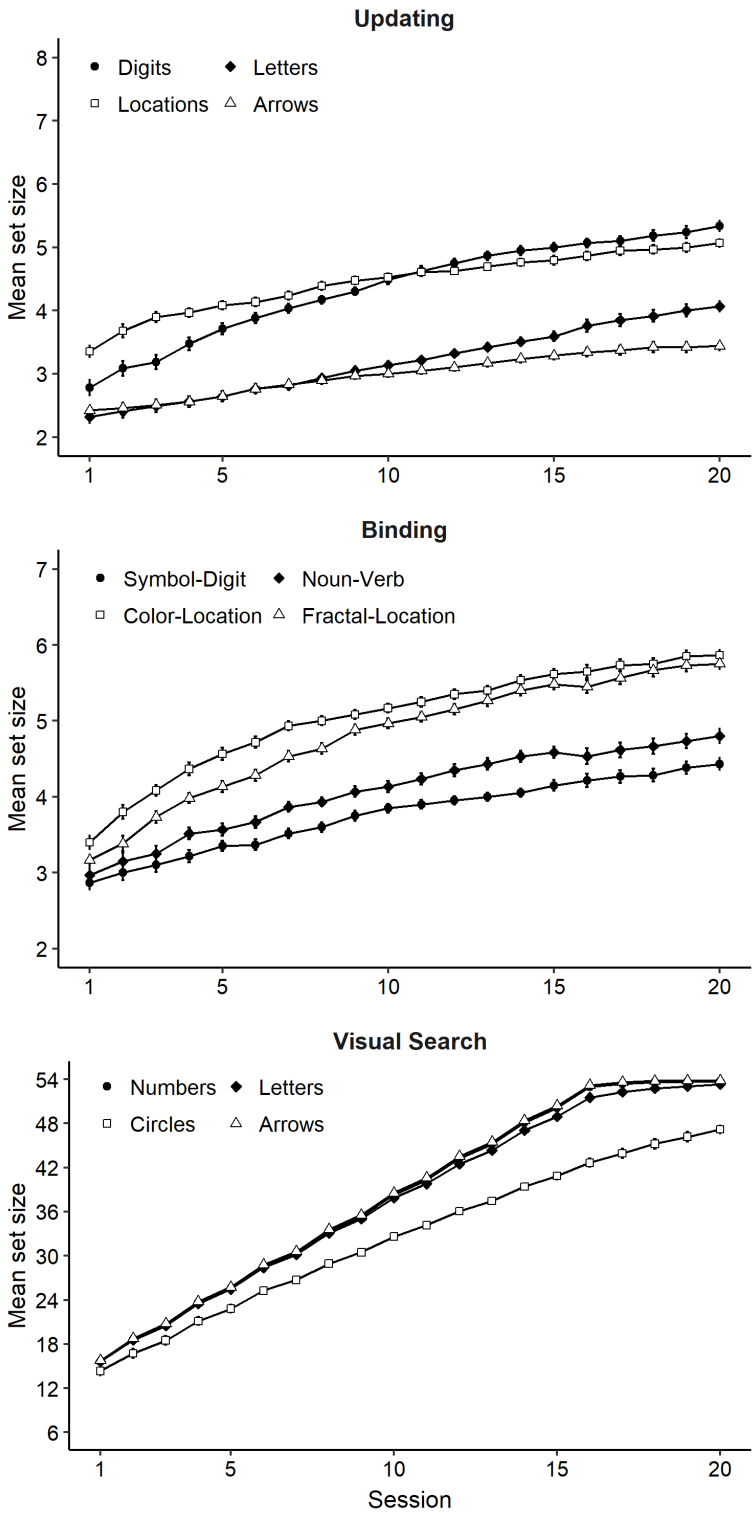

Figure 4. Performance gains during 20 sessions of training. Error bars denote the 95% confidence intervals for within-subjects comparisons calculated according to Cousineau (2005) and Morey (2008). Limits of the y-axes reflect the minimum and maximum of set sizes that could be achieved during training.

We next evaluated motivation during training by running a Bayesian LME for each of the motivation measures assessed at the end of each session (i.e., enjoyment, effort and perceived fit between task difficulty and ability) as dependent variable, group (updating, binding, and active control) and the linear contrast of the 20 sessions as fixed effects, and participant as random effect. As Figure 5 illustrates, the three groups rated their average enjoyment similarly, with strong evidence supporting the absence of an effect of group, $BF_{H1} = 1/30.11 \pm 2.14\%$ as well as the absence of an effect of session, $BF_{H1} = 1/25.99 \pm 1.78\%$. Evidence was ambiguous regarding the interaction between group and the linear contrast of session, $BF_{H1} = 1.98 \pm 2.18\%$, with the updating group showing a stronger drop in enjoyment ratings than the active control towards the end of training, $BF_{H1} = 32.29 \pm 4.38\%$.

For the self-reported effort spent on training, evidence was decisive in favor of a linear effect of session, $BF_{H1} > 100 \pm 1.32\%$, indicating that effort ratings decreased over time. In addition, there was strong evidence in favor of a Group x Session interaction, $BF_{H1} = 16.55 \pm 1.55\%$. Following up on this interaction, we found that the active control group's ratings remained more stable over the course of training relative to the binding group, $BF_{H1} > 100 \pm 4.17\%$, but were similar to those of the updating group, $BF_{H1} = 1/7.74 \pm 2.40\%$.

Finally, there was ambiguous evidence tending to favor an effect of group on the ratings of perceived fit between task difficulty and individual ability, $BF_{H1} = 2.02 \pm 5.16\%$, caused by the slightly lower average ratings in the updating group relative to the active control group, $BF_{H1} = 5.15 \pm 14.10\%$. There was no effect of session, $BF_{H1} = 1/6.36 \pm 5.12\%$, and no Group x Session interaction, $BF_{H1} =1/14.32 \pm 5.24\%$.

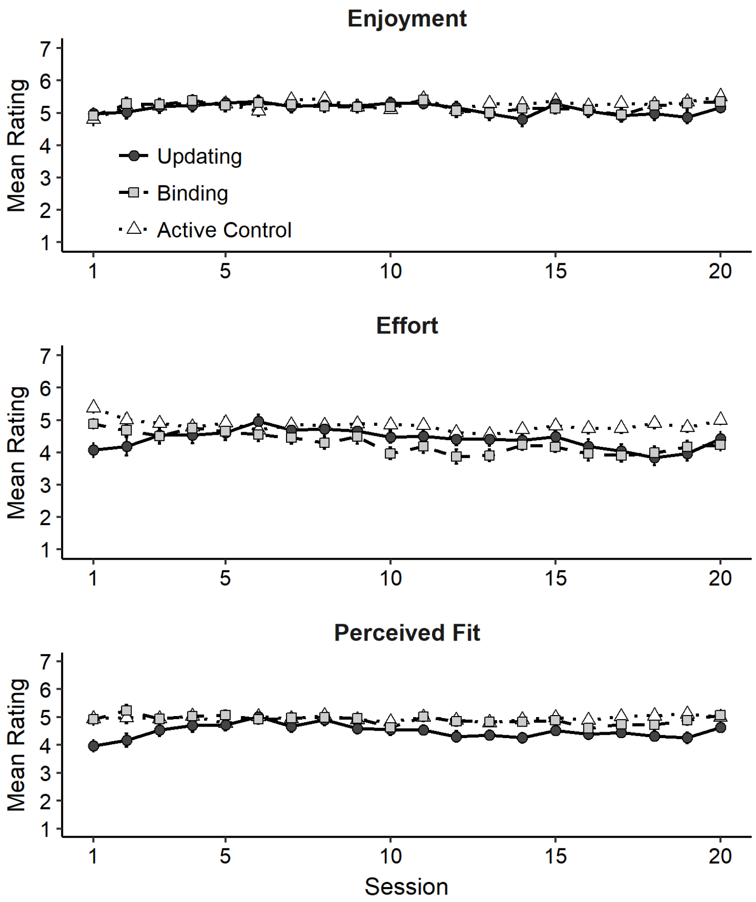

Figure 5. Motivation measures across the 20 training sessions as a function of group. Error bars denote 95% confidence intervals for within-subjects comparisons calculated according to Cousineau (2005) and Morey (2008).

Training-related expectations. To check whether the three training paradigms elicited different expectations that could affect posttest performance, we analyzed participants' ratings of subjective gains after the training intervention (see Table 5). Overall, evidence consistently favored the absence of group differences in ratings of subjective gains.

Table 5
Training-Related Expectations as a Function of Transfer Range

	Group M (SD)			Pairwise Comparisons (BF_{H1} ± Error)	
Range	Updating	Binding	AC	Updating vs. AC	Binding vs. AC
Training	3.81 (0.78)	3.88 (0.89)	3.69 (0.90)	1/3.96 ± 0.03	1/2.79 ± 0.00
Transfer	2.31 (1.16)	2.63 (1.22)	2.42 (1.09)	1/4.58 ± 0.05	1/3.29 ± 0.00
Everyday	1.81 (0.40)	1.89 (0.31)	1.89 (0.32)	1/2.68 ± 0.01	1/5.42 ± 0.00

Note. Participants rated their subjective gains on a scale from 1 (not at all) to 5 (very). AC = active control.

Summary. All three training groups showed substantial performance improvements across the 20 sessions. Motivation ratings were overall relatively similar, except that the updating group rated their intervention as somewhat less enjoyable towards the end of training, and the perceived fit between task difficulty on average as less optimal than the active control group. In addition, the reduction in self-reported effort spent over time was more pronounced in the binding group than in the active control group. Nevertheless, training-related expectations were relatively similar across groups.

Training and Transfer Gains

To evaluate whether training induced changes in cognitive performance, we compared the performance gains from pretest to posttest in each WM training group to those observed for the active control group using Bayesian LMEs across each of the four tasks measuring the same ability.[1] Results are summarized in Table 6 and illustrated in Figure 6.

[1] It would have been desirable to analyze those gains on the latent level within a structural-equation framework (e.g., Schmiedek et al., 2010). However, probably due the moderate group sizes for the purposes of structural-equation modeling, we could not establish measurement invariance at pretest for the multiple groups and, so, had to refrain from using latent difference-score modeling techniques (McArdle, 2009).

Table 6

Parameter Estimates for Fixed Effects of the Bayesian Linear Mixed-Effects Models

	Updating vs. Active Control		Binding vs. Active Control	
Ability	M_{Diff} [95% HDI]	BF ± error %	M_{Diff} [95% HDI]	BF ± error %
Training and Near Transfer Gains				
Updating	0.35 [0.17, 0.53]	> 100 ± 1.93	0.08 [-0.07, 0.23]	1/6.02 ± 1.79
Binding	0.08 [-0.11, 0.28]	1/8.24 ± 1.35	0.55 [0.30, 0.78]	> 100 ± 1.75
Visual Search[a]	0.52 [0.28, 0.78]	> 100 ± 3.14	0.57 [0.31, 0.82]	> 100 ± 1.45
Far Transfer				
Reasoning	-0.04 [-0.19, 0.11]	1/10.37 ± 1.34	-0.01 [-0.16, 0.14]	1/11.36 ± 5.38
Shifting[a]	-0.03 [-0.24, 0.18]	1/9.21 ± 3.30	0.02 [-0.18, 0.21]	1/9.77 ± 2.34
Speed[a]	-0.15 [-0.29, -0.01]	1.16 ± 1.24	-0.11 [-0.26, 0.04]	1/2.48 ± 4.83
Inhibition[a]	0.07 [-0.12, 0.26]	1/9.40 ± 1.67	-0.12 [-0.30, 0.08]	1/5.76 ± 4.91
Working Memory Processing				
Focus Switching[a]	-0.23 [-0.43, -0.03]	1.25 ± 3.03	0.03 [-0.15, 0.23]	1/11.50 ± 2.17
Removal[a]	0.09 [-0.13, 0.31]	1/8.46 ± 1.50	-0.04 [-0.28, 0.15]	1/11.30 ± 1.86
Interference Resolution				
Updating[a]	-0.26 [-0.47, -0.05]	1.40 ± 2.15	-0.09 [-0.27, 0.11]	1/7.58 ± 6.34
Binding	-0.11 [-0.29, 0.08]	1/5.57 ± 1.46	-0.07 [-0.29, 0.15]	1/8.84 ± 1.96

Note. Estimates are the mean group differences from 10,000 samples of the posterior distribution. HDI = highest density interval of the posterior distribution.

[a]Negative mean group differences reflect greater performance gains in the experimental group.

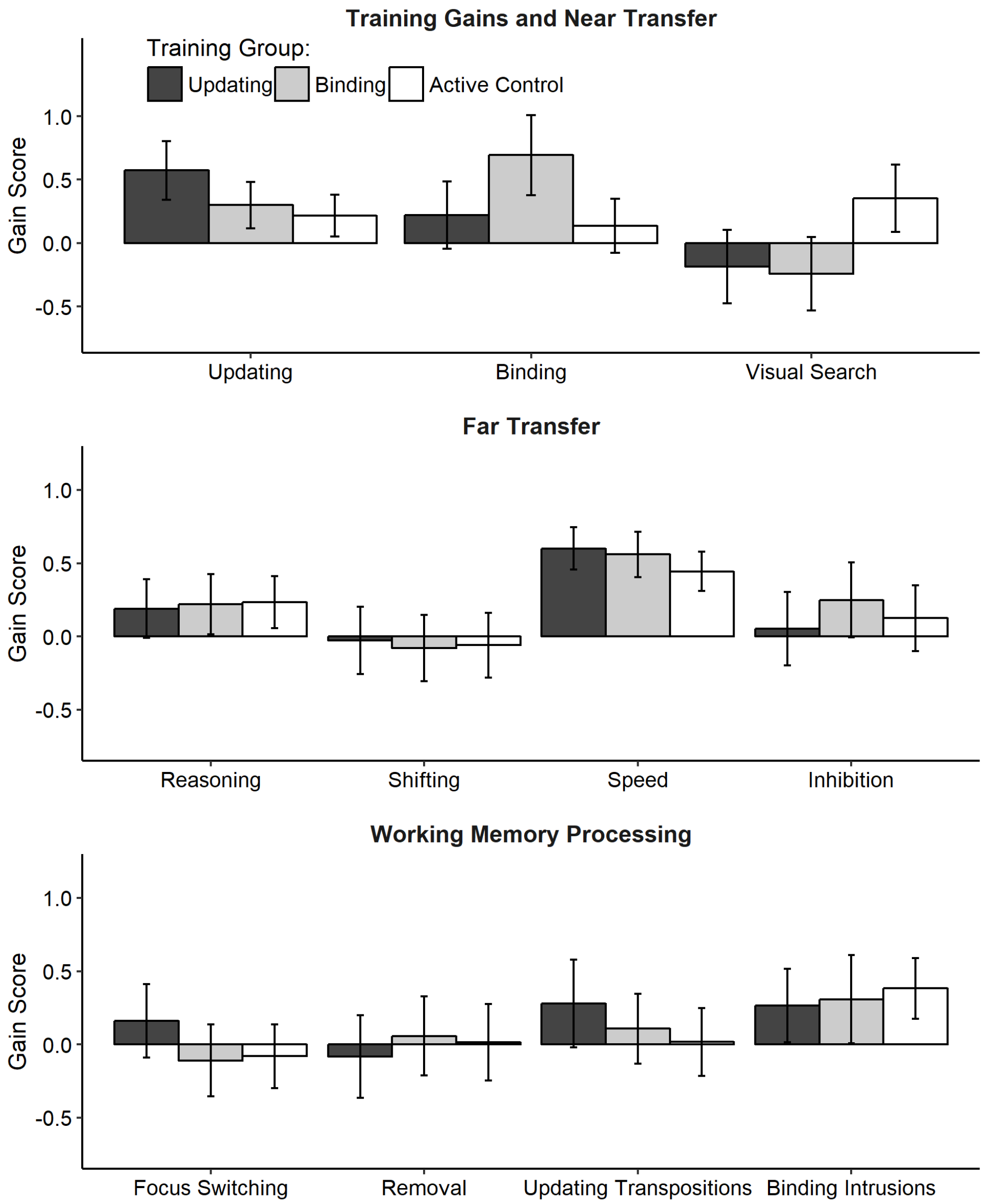

Figure 6. Gains from pretest to posttest in the trained tasks, structurally different working memory tasks (near transfer; both in top panel), working memory mechanisms (middle panel), and other but related abilities (far transfer; bottom panel). For illustration purposes, scores were averaged across the four tasks administered per ability and reverse-coded so that higher values correspond to greater performance gains. Error bars denote 95% confidence intervals.

Training gains. Evidence in favor of training effects from pretest to posttest for all three training groups was substantial to decisive (see top panel of Figure 6). Compared to the active control group, the updating group showed larger gains in updating, $BF_{H1} > 100 \pm 1.93\%$, and the binding group improved more strongly in binding, $BF_{H1} > 100 \pm 1.75\%$.

Similarly, the active control group showed larger gains in visual search than the updating training group, $BF_{H1} > 100 \pm 3.14\%$, and the binding group, $BF_{H1} > 100 \pm 1.45\%$.

Near transfer. Despite the large correlation between updating and binding on the latent level, and despite the large training effects, the evidence favored the absence of near transfer effects from updating training to binding, $BF_{H1} = 1/8.24 \pm 1.35\%$, as well as from binding training to updating, $BF_{H1} = 1/6.02 \pm 1.79\%$.

Far transfer. In line with the absence of near transfer, we found no far transfer effects on reasoning, shifting, speed, and inhibition (see middle panel of Figure 6), although some of these abilities correlated at least moderately with updating and binding at pretest. The absence of far transfer was largely supported by substantial to strong evidence (ranging between $1/5.76 \pm 4.91\%$ and $1/11.36 \pm 5.38$) except for speed, for which the evidence was ambiguous (updating: $BF_{H1} = 1.16 \pm 1.24\%$; binding: $BF_{H1} = 1/2.48 \pm 4.83\%$).

WM processing. Next, we investigated group differences in gain scores for focus switching, removal, and interference resolution (see bottom panel of Figure 6). To examine effects on focus switching, we analyzed object-switching costs derived from the updating tasks. The updating group tended towards improved focus switching, but the evidence was ambiguous, $BF_{H1} = 1.25 \pm 3.03\%$. The evidence strongly supported the absence of effects in the binding group, $BF_{H1} = 1/11.50 \pm 2.17$. The evidence also favored the absence of any training-specific improvements in the removal of information from WM (updating training: $BF_{H1} = 1/8.46 \pm 1.50\%$; binding training: $BF_{H1} = 1/11.30 \pm 1.86\%$).

To examine whether training improved interference resolution, we analyzed the proportion of transposition errors made at recall in the updating tasks, and the accuracy to intrusion probes in the binding tasks. There was a trend that updating training reduced the proportion of transposition errors, but the evidence was ambiguous only, $BF_{H1} = 1.40 \pm 2.15\%$. Updating training did not improve performance in binding intrusion probes, $BF_{H1} =$

1/5.57 ± 1.46%. Moreover, binding training affected neither indicator of interference resolution over and above the changes observed in the active control group (transposition errors: BF_{H1} = 1/7.58 ± 6.34%, intrusion probes: BF_{H1} = 1/8.84 ± 1.96%).

Summary. Although we observed large training gains for both experimental groups, the evidence consistently supported the absence of any transfer effects even to very strongly correlated abilities. Moreover, there was little evidence that training affected WM processing in terms of focus switching, removal, or interference resolution.

Alternative Mechanisms of Change

To further investigate the pattern of results, we explored two often discussed possible reasons for observed lack of transfer despite large training gains. One possibility is that training induced a stronger reliance on familiarity-based processing, which would become evident by higher rates of correct responses accompanied by higher rates of false-positive responses. For this purpose, we analyzed the proportion of errors committed in the WM training tasks. A second possibility is that participants acquired highly task-specific strategies that boosted performance in the trained tasks, but potentially hindered transfer. Therefore, we also analyzed the retrospective reports of strategy-use.

Error patterns. Updating. Besides recalling the correct item at the correct position (see updating training gains) and recalling a correct item at the wrong position (transposition errors, see WM processing), participants could recall an item that was not part of the current memory list at all (extra-list errors, see left column in Figure 7). We found strong evidence that the updating group improved more than the active control group in extra-list errors, M_{Diff} = -0.33 95% HDI [-0.51, -0.16], BF_{H1} = 58.27 ± 1.55%. The binding group did not improve, M_{Diff} = -0.07 95% HDI [-0.23, 0.09], BF_{H1} = 1/8.75 ± 1.43%.

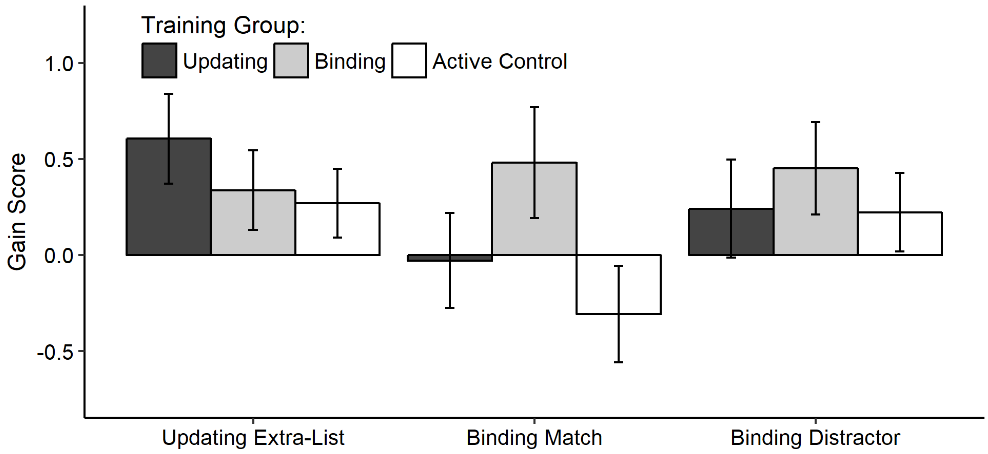

Figure 7. Gains from pretest to posttest in the proportion of errors committed in the trained tasks in addition to those reported above. For illustration purposes, scores were averaged across the four tasks administered per ability and reverse-coded so that higher values correspond to greater performance gains. Error bars denote 95% confidence intervals.

Binding. Recognition probes could be positive (matches) or negative, with the probe being either a correct item at a wrong position (intrusions, see WM processing) or an item that was not part of the current memory list at all (distractors). We observed decisive evidence that the binding group exhibited greater accuracy gains in matching probes than the active control group, M_{Diff} = 0.77 95% HDI [0.51, 1.05], BF_{H1} > 100 ± 1.78%. For the updating relative to the active control group, evidence was ambiguous, M_{Diff} = 0.27 95% HDI [0.03, 0.49], BF_{H1} = 1.32 ± 1.29%. For distractor probes, the evidence was ambiguous for the binding group, M_{Diff} = 0.22 95% HDI [0.04, 0.42], BF_{H1} = 1.22 ± 2.37%, and strongly favored the absence of a difference between the updating group and the active control group, M_{Diff} = 0.02 95% HDI [-0.17, 0.22], BF_{H1} = 1/10.85 ± 1.34%.

In recognition tasks, changes in response patterns can be the result of systematic changes in response bias. For example, a tendency to say YES more often at pretest, but a tendency to say NO more often at posttest would yield accuracy losses in matching probes accompanied by accuracy gains in intrusion and distractor probes. The pattern of changes seen in the active control group suggests such a change in response bias, whereas the gains in the binding group seem less bias-dependent. We evaluated the strength of evidence for systematic shifts in bias by testing for within- and between-group changes in the response criterion C. As illustrated in Figure 8, all groups started with a relatively relaxed criterion ($C < 0$) at pretest, but the active control group shifted towards a more conservative response criterion ($C > 0$) after training. This within-group bias shift was supported by decisive evidence, $M_{Diff} = 0.14$ 95% HDI [0.10, 0.18], $BF_{H1} > 100 \pm 2.36\%$. The updating training group showed a similar bias shift, albeit to a lesser degree, the $M_{Diff} = 0.06$ 95% HDI [0.03, 0.10], $BF_{H1} = 17.17 \pm 1.49\%$. In contrast, the binding training group's response criterion remained relatively stable from pretest to posttest, $M_{Diff} = -0.03$ 95% HDI [-0.07, 0.02], $BF_{H1} = 1/7.23 \pm 4.67\%$. Whereas the evidence was ambiguous regarding the difference in bias shift between the updating training and the active control group, $M_{Diff} = -0.28$ 95% HDI [-0.53, -0.05], $BF_{H1} = 1.39 \pm 2.01\%$, it was decisive in favor of a difference between the binding and active control group, $M_{Diff} = -0.59$ 95% HDI [-0.89, -0.32], $BF_{H1} > 100 \pm 1.09\%$.

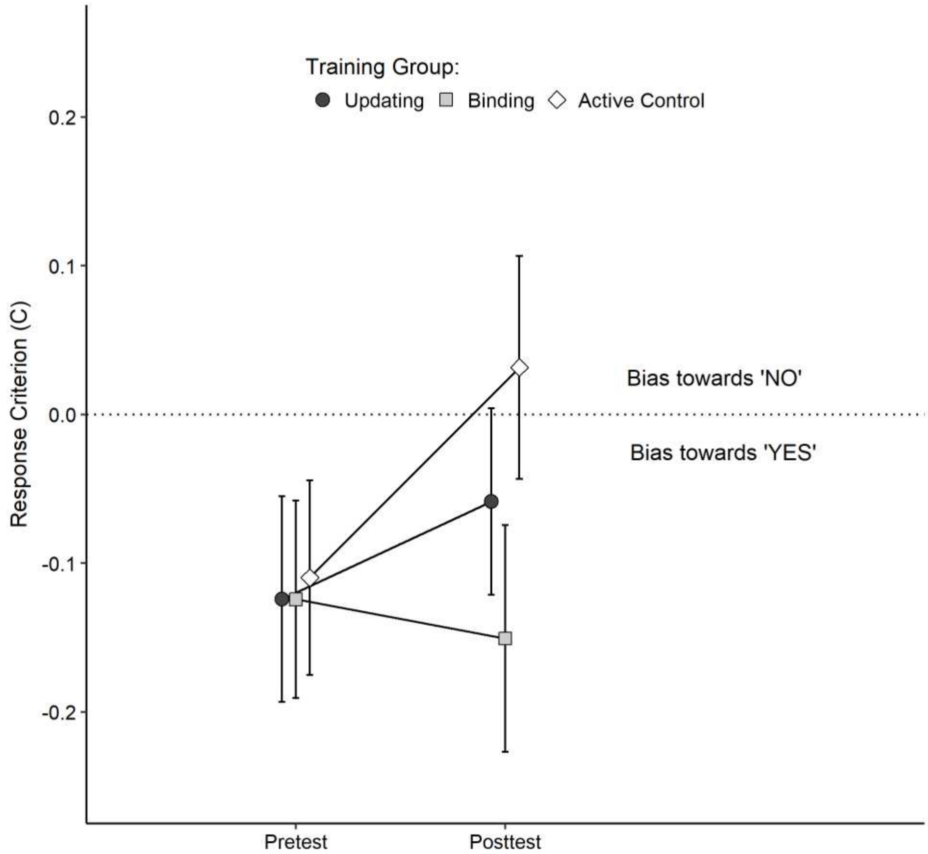

Figure 8. Change from pretest to posttest in response bias in the binding tasks. The dotted line represents a neutral response bias; $C > 0$ reflects a conservative response criterion (i.e., a

tendency to say ‘no’), $C < 0$ reflects a more relaxed response criterion (i.e., a tendency to say ‘yes’). Error bars denote 95% confidence intervals.

Strategy use. The reported strategies and the corresponding average ratings of perceived usefulness are reported in Table 7. Most participants reported to have used at least one strategy during training (updating: 83.04%; binding: 87.88%; active control: 62.50%). Of those, the majority of participants in the WM training groups mentioned that they primarily rehearsed or read aloud the memoranda, followed by strategies that involve some form of reducing memory load such as mapping the memoranda to the keyboard (updating training) or focusing efforts on remembering only a subset of stimuli (binding training). Participants in the active control most frequently reported having relied primarily on either holistic search strategies (i.e., attempting to attend to the full search array in parallel) or serial search strategies (i.e., scanning the search array item by item). Other strategies mentioned included chunking, visualizing, and recoding of elements, or a mixture of multiple strategies. Given that only few participants reported to have not used any strategies, we refrained from testing how the use of strategies affected transfer gains.

On average, usefulness ratings were higher in the updating group ($M = 4.18$, $SD = 0.93$) and in the binding group ($M = 4.28$, $SD = 0.77$) than in the active control ($M = 3.65$, $SD = 0.97$). Bayesian t-tests yielded substantial evidence in favor of higher ratings in the updating than in the control group, $M_{Diff} = -0.48$ 95% HDI [-0.85, -0.11], $BF_{H1} = 5.39 \pm 0.00\%$, and very strong evidence in favor of higher ratings in the binding than in the control group, $M_{Diff} = -0.58$ 95% HDI [-0.92, -0.25], $BF_{H1} = 64.67 \pm 0.00\%$. The two WM training groups rated the usefulness of the strategies they used similarly high, $M_{Diff} = -0.08$ 95% HDI [-0.38, 0.22], $BF_{H1} = 1/4.23 \pm 0.02\%$.

Table 7
Retrospectively Reported Strategy Use as a Function of Group

Strategy	Frequency (%)	Usefulness M (SD)
	Updating	
Rehearsal[a]	47.46	4.39 (0.79)
Mapping	11.86	3.00 (1.15)
Other[b]	15.25	4.56 (0.53)
Multiple	8.47	4.00 (0.71)
	Binding	
Rehearsal[a]	28.79	4.26 (0.73)
Subsetting	22.73	4.20 (0.86)
Other[c]	21.21	4.29 (0.83)
Multiple	15.15	4.40 (0.70)
	Active Control	
Holistic	25.00	4.11 (0.68)
Serial	20.83	3.27 (0.88)
Other[d]	16.67	3.58 (1.16)

Note. The perceived usefulness scale ranged from 1 (not at all) to 5 (very). Data from one participant per group was missing. One additional participant in the active control group indicated to have used a strategy but did not describe it.
[a]Also includes reading stimuli aloud.
[b]For example, chunking (n = 3), imagery (n = 3), and recoding (n = 2).
[c]For example, associating (n = 4), chunking (n = 3), imagery (n = 2), and recoding (n = 4).
[d]For example, attending to a specific part of the screen only (n = 2).

Summary. We found no evidence for training-induced changes in WM processing. Instead, updating training solely increased the proportion of correct responses and reduced the proportion of extra-list errors in the trained tasks. These improvements did not generalize to the other WM paradigm: the updating training group’s performance gains in binding matching and distractor probes were comparable to those observed in the active control group. Similarly, the binding training group showed marked improvements in matching probes and, to a lesser extent, in distractor probes. In contrast to the other two groups, these performance changes were not accompanied by systematic shifts in response bias. However, despite these bias-independent gains, the binding training group did not improve in correctly recalling memoranda or reducing extra-list errors in the updating tasks. Moreover, most participants reported to have used strategies which they rated as relatively useful for completing the tasks. Although both WM paradigms afforded similar strategies (e.g., rehearsal, reduction of memory load, chunking, or imagery), performance gains in one paradigm did not transfer to the other paradigm.

Discussion

In this study, we evaluated the evidence for and against the effectiveness of WM training in enhancing cognition using Bayesian inference. We found substantial and consistent evidence supporting the absence of both near and far transfer, suggesting that training did not enhance WM capacity. Furthermore, there was little evidence that WM training affected any of the processes that had been proposed to underlie transfer effects (i.e., focus switching, removal of no longer relevant information, or the resolution of interference). Instead, additional analyses of error patterns and self-reported use of strategies in the trained tasks suggest that the repetitive practice encouraged the development of stimuli-specific expertise and the use of paradigm-specific strategies.

Evaluation of the Theoretical Transfer Model

One prerequisite for training effects to generalize to other abilities is that the training and transfer abilities share a considerable portion of variance. To establish the magnitude of that shared variance, we subjected the assessment data to a latent-variable confirmatory factor analysis. Consistent with previous studies, updating and binding were strongly correlated (cf. Wilhelm et al., 2013). Both WM factors were moderately related to reasoning (e.g., Engle et al., 1999; Friedman et al., 2006; Kyllonen & Christal, 1990; Oberauer et al., 2008; Süß et al., 2002) and speed (e.g., Conway, Cowan, Bunting, & Minkoff, 2002; McAuley & White, 2011; Schmiedek et al., 2007). The correlations with shifting were only weak, a finding that has been observed occasionally in previous studies (e.g., Benedek, Jauk, Sommer, Arendasy, & Neubauer, 2014; Huizinga, Dolan, & van der Molen; Hull, Martin, Beier, Lane, & Hamilton, 2008).

The tasks administered to assess inhibition, however, were largely uncorrelated, and so we failed to establish a latent inhibition factor. Although this appears to be in conflict with a large body of literature (e.g., Friedman & Miyake, 2004b; Friedman et al., 2006; McVay & Kane, 2012; Miyake et al., 2000; Shipstead, Lindsey, Marshall, & Engle, 2014; Unsworth, Fukuda, Awh, & Vogel, 2014; Unsworth & McMillan, 2014), it is an increasingly common finding (e.g., Brydges, Reid, Fox, & Anderson, 2012; Guye & von Bastian, 2017; Hull et al., 2008; Krumm et al., 2009; Paap & Greenberg, 2013; Rey-Mermet, Gade, & Oberauer, in press; Rey-Mermet, Gade, Souza, von Bastian & Oberauer, 2018; von Bastian & Oberauer, 2013; von Bastian et al., 2016). Low zero-order correlations could be due to low psychometric properties of the tasks administered (Cooper, Gonthier, Barch, & Braver, 2017; Hedge, Powell, & Sumner, in press). Notably, however, reliability estimates for the inhibition measures in the present study ranged between .47 and .74, with 3 out 4 of the tasks exhibiting reliabilities ≥ .65. Alternatively, it is possible that individual differences in inhibition are highly task-specific, questioning inhibition as a psychometric construct (Rey-Mermet et al., 2017; Rey-Mermet et al., 2018). The present data add to this notion.

Paradigm-Specific Performance Gains

Another prerequisite for transfer effects is improved performance in the trained tasks. We found decisive evidence for training gains after both WM training interventions. Despite the considerable portion of shared variance between updating and binding, the training-induced gains were highly paradigm-specific: we observed substantial evidence favoring the absence of near transfer effects to the respective other WM factor. The lack of near transfer may seem surprising and at odds with meta-analytic near transfer effects (e.g., Melby-Lervåg et al., 2016). However, as recently demonstrated by Soveri et al. (2017), the degree to which training and near transfer tasks overlap in materials and surface structure needs to be taken into account. Specifically, Soveri et al.'s (2017) reanalysis of Melby-Lervåg et al.'s (2016) data from comparisons of healthy adults showed that the near transfer effect size was significantly reduced when considering only structurally different near transfer tasks. Indeed, the lack of near transfer in the present study is in line with findings both from our own previous studies (Guye & von Bastian, 2017; von Bastian & Eschen, 2016; von Bastian et al., 2013; von Bastian & Oberauer, 2013) as with those of other studies that assessed transfer to structurally different WM tasks (e.g., Harrison et al., 2013; Jaeggi et al., 2008; Jaeggi, Studer-Luethi, et al., 2010; Schmiedek et al., 2010; Sprenger et al., 2013).

The specificity of improvements after intensive practice is also in line with decades of research in skill acquisition showing that practice effects are often highly paradigm- and stimulus material-specific, yielding only limited transfer to new tasks (cf. Ericsson, Chase, & Faloon, 1980; Healy, Wohldmann, Sutton, & Bourne, 2006; Lewandowsky & Thomas, 2009). For example, Healy et al. (2006) showed that participants who learned to use a computer mouse that reverses either vertical, horizontal, or a combination of both movements, could not transfer this ability to one of the other reversal conditions (e.g., transfer from vertical to horizontal movements or vice versa).

Given the lack of near transfer, it is unsurprising that we found substantial to strong evidence favoring the absence of far transfer effects to reasoning and shifting with $BF_{H1} \geq 1/9.21$. Evidence was ambiguous in respect to processing speed, but also tended to support the null hypothesis. The absence of far transfer in the present study corroborates meta-analytic findings for comparisons with active controls (e.g., Melby-Lervåg, 2016; Dougherty et al., 2015) and most of our own previous findings, with two exceptions. In two of our studies, we found some evidence for far transfer to reasoning (von Bastian & Oberauer, 2013; Zimmermann et al., 2016). Whereas the two studies varied in the training tasks administered and the age group examined, they have in common that we assessed long-term effects four (Zimmermann et al., 2016) and six months (von Bastian & Oberauer, 2013) after the posttest.

In both instances, the difference between the experimental group and the active control increased from posttest to follow-up assessment, which contributed to the overall transfer effect on reasoning being significant. Although such ‘sleeper effects’ have occasionally been reported elsewhere (e.g., Blakey & Carroll, 2015), they are theoretically not predicted and difficult to explain. We can only speculate about potential reasons for why the group differences were stronger at follow-up than at the immediate posttest. For example, it is possible that changes induced from taking part in any training (e.g., changes in motivation) mask WM training-specific effects immediately after the intervention but wear off over the course of the weeks leading up to follow-up assessment. At the same time, the effects may simply present chance findings and, thus, future research is needed to establish their replicability. Notably, though, current meta-analyses do not support the notion of reliable effects of WM training at follow-up assessments (e.g., Melby-Lervåg et al., 2016).

Mechanisms of Training-Related Change

In line with the lack of near and far transfer effects, we found no evidence for training-related changes in the WM mechanisms (i.e., focus switching, removal of no longer relevant information, or the resolution of interference) that were hypothesized to possibly underlie transfer of WM training effects. Hence, the question remains, what did change during training that may have led to the large gains observed in the practiced tasks? We suggest that participants acquired stimuli-specific expertise boosting their memory of item contents in addition to using paradigm-specific strategies.

Stimuli-specific expertise. Our additional analyses of error patterns revealed that training gains in the updating training tasks were primarily driven by increased rates of correctly recalled items and a reduction in extra-list errors. Similarly, in the binding training tasks, improvements were due to increased rates of correctly identifying matching probes and, although to a lesser degree, correctly rejecting distractor probes. These improvements were, again, highly paradigm-specific, with neither WM group showing analogous effects for the respective other set of tasks.

The changes specific to the correct items suggest that training may have increased familiarity-based processing. Following dual-process models of recognition, memory accuracy is based on two separable sources of information processing: familiarity and recollection (e.g., Atkinson, Herrmann, & Wescourt, 1974; Oberauer, 2005; 2008; Yonelinas, 2002). Whereas familiarity is based on the memory of an item regardless of its current context, recollection is based on the memory of the binding between an item and its current context (e.g., a word and its specific position within the memory array). If the training-related improvement in memory accuracy were due to enhanced familiarity-based (or context-independent) processing only, one would expect to observe a concurrent increase in responses involving the correct items but at wrong positions (i.e., more YES responses to intrusion probes in the binding tasks, and more transposition errors in the updating tasks). Conversely, if the improved memory accuracy were due to enhanced recollection-based (or context-dependent) processing only, one would expect to observe a concurrent decrease in those false-positive responses. In our study, the increase in correct responses was accompanied by relatively stable false-positive rates, suggesting that practice improved memory of items but without compromising recollection-based processing. Thus, it is unlikely that training solely led participants to simply rely more strongly on familiarity-based processing.

Alternatively, participants may have acquired stimuli-specific expertise through the repetitive practice with the same set of stimuli over an extended period of time, yielding enhanced resolution of the stimuli representations in memory. This tentative explanation borrows from a visual WM model recently introduced by Oberauer and Lin (2017). The authors suggest a memory system in which an item’s content (e.g., a color) is bound to its context (e.g., a location) in a two-dimensional, continuous binding space. At retrieval, contents are accessed through the context, with the probabilistic choice of a response being affected by the strength of the content-context binding, the precision of the representation on the content dimension, and the amount of background noise. Specifically, a context cue presented (e.g., a probed location) will activate the content representation bound to that context. In addition, although to a lesser degree, those representations that are bound to highly similar contexts (e.g., nearby locations) will also receive some activation, thereby creating interference.

During training, the repetitive encoding and retrieval of the same stimuli may have gradually enhanced the precision of their representations. Such a boost in precision of the content representations would benefit recall of correct items, but, as it would not affect the binding strength, would have little effect on the errors resulting from interference (i.e., transposition errors and rejection of intrusion probes). Consistent with this explanation and with the lack of transfer in our study, it has been shown that the precision of items in memory is task-dependent (Chow & Conway, 2015), and unrelated to fluid intelligence (Fukuda et al., 2010). However, as our training tasks involved discrete as opposed to continuous stimuli (e.g., colors or directions), our data does not allow for estimating the precision of content representations. To test our proposition more directly, future studies would need to employ tasks that allow for differentiating between precision and number of items in memory such as change-detection (e.g., Awh, Barton, & Vogel, 2007; Xu & Chun, 2006) or continuous-reproduction tasks (e.g., Bays, Catalao, & Husain, 2009; Zhang & Luck, 2008).

Paradigm-specific strategies. As strategy-based training interventions frequently yield only narrow transfer (for reviews, see Lustig et al., 2009; Morrison & Chein, 2011), lack of transfer after process-based WM training has been suggested to be caused by the development or use of highly paradigm-specific strategies (cf. von Bastian & Oberauer, 2014). For example, Laine, Fellman, Waris, and Nyman (2018) recently demonstrated that (internally developed or externally instructed) strategy usage during training can improve performance in trained and structurally similar untrained tasks while not affecting performance in structurally different tasks; a pattern in line with our results. Indeed, despite our efforts to reduce strategy use during training through gradually decreasing the response time limit (and, thus, the time to employ a strategy), more than 80% of WM trainees reported the use of WM-specific strategies. Hence, participants may have acquired strategies that aided performance in the trained tasks but were not applicable to the respective other set of tasks (Bailey, Dunlosky, & Kane, 2008). Self-reported strategies were quite similar across paradigms though. In both WM training groups, rehearsal of items was the most frequently mentioned strategy, which is in line with previous studies on strategy use in WM (e.g., Bailey et al., 2008; Dunlosky & Kane, 2007; Friedman & Miyake, 2004a; Logie, Della Sala, Laiacona, Chalmers, & Wynn, 1996; Morrison, Rosenbaum, Fair, & Chein, 2016; Unsworth & Spillers, 2010). On average, participants rated rehearsal as a useful strategy even though it has been shown to be normatively ineffective (Dunlosky & Kane, 2007), which is consistent with work showing that people often have only limited knowledge about strategy effectiveness (e.g., Hertzog, Price, & Dunlosky, 2008). Thus, it is possible that participants indeed attempted to use rehearsal also in the respective other paradigm, but that it simply had little effect on their performance.

Other frequently mentioned strategies included reducing WM load by either prioritizing a subset of items in the binding tasks (cf. Atkinson, Baddeley, & Allen, 2017) or by mapping the memoranda to mental shapes drawn on fingers, legs, or the keyboard (cf. Minear et al., 2016). While these strategies may be more effective in boosting performance, especially at the limits of WM capacity (e.g., Atkinson et al., 2017; Cusack, Lehmann, Veldsman, & Mitchell, 2009; Linke, Vicente-Grabovetsky, Mitchell, & Cusack, 2011), differences between the paradigms such as the response mode (i.e., recall in the updating tasks and recognition in the binding tasks, cf. Bailey, Dunlosky, & Hertzog, 2014) or the complexity of the memoranda (cf. Atkinson et al., 2017; see also Bengson & Luck, 2016) could have prevented participants from successfully applying these strategies to the respective other paradigm.

A third possibility is that, even if participants would have been able to apply the same strategy to both training and near transfer tasks, they may still not have chosen to do so, for example due to a lack of interest in making the cognitive effort (cf. Carretti, Borrella, Zavagnin, & De Beni, 2011). In a recent study, Morrison et al. (2016) asked participants to indicate the strategies they used across a range of six verbal WM paradigms. Although the distribution of strategies reported was relatively similar for the different paradigms, individuals varied in the consistency in which they employed a given strategy in any given paradigm. For example, rehearsal was the most frequently selected strategy for a complex span task across participants. However, for other tasks, distinct subgroups of participants emerged: whereas one subgroup primarily relied on rehearsal also in an item recognition but not at all in a running memory span task, another subgroup showed the opposite pattern. Hence, there are considerable individual differences in the consistency of strategy use even across paradigms that generally afford similar strategies. Although the ability to dynamically shift strategies appears to be unrelated to WM capacity (Unsworth, 2016), there is some tentative evidence that it may relate to better training outcomes (Dunning & Holmes, 2014). However, we assessed strategy use only retrospectively after the training phase, and only for the trained, but not for the non-trained tasks; thus, the present data does not allow for further investigating these issues. Future studies are needed to directly test the strategy affordance hypothesis in the context of WM training, and to better understand how consistency in strategy use is related to training and transfer effects.

Strengths

One of the major strengths of our study is the theory-driven selection of training and transfer tasks. We based our study on the three-embedded-components model of working memory (Oberauer, 2009; Oberauer & Hein, 2012), which describes WM as three different levels of information selection, each level narrowing down the information content of the preceding one. A theory-based approach of task selection allows for determining whether an improvement in a cognitive task is accomplished through an increase in WM capacity or enhanced WM efficiency (cf. von Bastian & Oberauer, 2014). Moreover, we systematically investigated the WM mechanisms that had been hypothesized to mediate near and far transfer effects, and we explored alternative mechanisms that may have led to the training-specific gains.

Another strength of our study was the relatively large sample size with 59 to 72 participants per group. In contrast to other cognitive training approaches (e.g., Hardy et al., 2015; Owen et al., 2010; Tennsted & Unverzagt, 2013), most previous WM training studies compared 30 participants per group or less. According to the meta-analysis by Melby-Lervåg et al. (2016) that included 87 WM training studies, there was only one study with group sizes larger (i.e., Estrada, Ferrer, Abad, Román, & Colom, 2015 with 114 to 193 participants per group) and one other study with group sizes comparable (i.e., Sprenger et al., 2013 with 57 and 70 participants per group in experiment 1) to ours. Indeed, the resulting Bayesian evidence was at least substantial in our study.

Moreover, we included an active control group. Although meta-analyses sometimes show no effect of the type of control group (i.e., active or passive) on the average transfer effect (e.g., Au et al., 2015; Karbach & Verhaeghen, 2015; but see Dougherty et al., 2015), including an active control group is still important from a methodological standpoint, as the absence of a statistical difference cannot exclude confounds from non-specific intervention effects in any given single study. An optimal control group should demand only little WM, and participants should perceive the control intervention as a believable and potentially effective intervention (cf. Morrison & Chein, 2011; Shipstead, Hicks, & Engle, 2012; von Bastian & Oberauer, 2014). Hence, the control treatment should be as similar as possible to the experimental treatment. We argue that visual search training met these criteria well. First, visual search did not correlate with the other cognitive abilities assessed as demonstrated by the transfer model (Figure 3). Second, training conditions were identical across interventions (e.g., number and duration of training sessions, adaptivity, and the amount of feedback and experimenter contact). Moreover, motivational measures assessed during training showed that participants of the control group perceived training at least as equally enjoyable as the two experimental groups; if anything, the updating group rated their training intervention as slightly less enjoyable, which is in line with that we observed the highest rate of compliance in the control group. We observed a substantial difference in the ratings of effort between the binding and the active control group, with the active control group, however, reporting having spent more effort during training. We can only speculate about reasons for this difference, but possibly, ratings of the effort spent are a consequence of the perceived performance in that session. More specifically, as the active control group showed steeper increases in training performance than the binding group, they on average experienced more level-ups per session than the binding group. Hence, participants might have used their progress in a session in terms of level-ups as reference to retrospectively rate their effort in this session. Importantly, these differences in rated enjoyment and effort did not affect training-related expectations.

Finally, we measured each cognitive ability with four tasks. The application of multiple indicators to assess a construct minimizes the problems arising from task impurity (Miyake & Friedman, 2012). More specifically, the score of a task reflects both cognitive ability as well as systematic and random influences (e.g., Shipstead, Redick, et al., 2012). For example, in our study, the latent updating factor explained between 45% and 66% of variance in the single updating tasks. Hence, there was still a considerable amount of variance from task-specific sources. Therefore, transfer effects found for single tasks may reflect changes in the task-specific portions of variance rather than an enhancement of the underlying ability (Schmiedek et al., 2010). To extract the ability-specific effects, multiple measures of an ability are needed. An often-voiced concern about measuring each ability with multiple indicators is that it increases testing time. For example, Green, Strobach, & Schubert (2014) argued that such increased testing time could lead to ego-depletion or fatigue effects and, hence, obscure training and transfer gains. Therefore, as in our previous training studies (e.g., von Bastian & Oberauer, 2013), we administered the test battery in counterbalanced order, thereby controlling for any potential linear effects of fatigue or ego-depletion. Moreover, in a recent reanalysis of data from four cognitive test batteries from previous studies of our lab (i.e., Herkert, 2012; von Bastian & Eschen, 2016; von Bastian et al., 2016; von Bastian & Oberauer, 2013) that comprised up to 20 tasks and lasted between 2.5 and 4.5 hours, we found no evidence for fatigue or ego-depletion effects (cf. De Simoni, Luethi, Oberauer, & von Bastian, 2018).

Limitations

A potential drawback of our study is that training was self-administered at home. Despite regularly keeping in touch with participants during training and closely monitoring their progress, experimental control was less tight than if participants had completed the training in the laboratory. For example, although we asked participants to complete their training sessions in a quiet area where they would not be disturbed, it is impossible to know whether all of them met this request throughout the intervention. Laboratory-administered interventions have been argued to strengthen the commitment and motivation of participants (cf. Lampit, Hallock, & Valenzuela, 2014). However, regular personal contact with the experimenter also yields an increased risk of experimenter effects (e.g., participants act as expected from the experimenter, cf. Rosenthal, et al. 2005), possibly confounding training and transfer data. Moreover, although laboratory-based training interventions increase control over the environment during training (e.g., minimized impact of distractions), home-administered training interventions increase ecological validity of the results. Empirically, it is yet unclear whether WM training effectiveness is affected by whether it training is administered at home or in the laboratory. Although some meta-analyses that include different age groups support such an effect (e.g., Lampit et al., 2014; Schwaighofer et al., 2015), others that focus solely on young adults do not (e.g., Au et al., 2015). Indeed, several laboratory-based cognitive training studies did not find any transfer effects either (e.g., Linares et al., 2017; Redick et al., 2013), whereas other self-administered interventions were successful in establishing even far transfer (e.g., Jaeggi, Buschkuehl, Shah, & Jonides, 2014; von Bastian & Oberauer, 2013; Zimmermann et al., 2016). Taken together, it seems unlikely that the self-administered training regime is responsible for the lack of transfer.

Another limitation is that the present sample consisted of young adults. One could argue that young adults are simply not an adequate target for training interventions, as they are at the peak of their cognitive functioning, and, hence, room for additional improvement is small at best (e.g., Bherer et al., 2008; Karbach & Kray, 2009, see also Titz & Karbach, 2014 for a review). However, in a recent training study with older adults that followed a similar design and yielded comparably strong Bayesian evidence as the present study, we also found no transfer effects (Guye & von Bastian, 2017). Moreover, fitting latent growth curve models to the data from the present study and that from Guye and von Bastian (2017), we estimated how baseline performance in the training tasks predicted the training trajectories (Guye, De Simoni, & von Bastian, 2017). For all three training groups, we found decisive evidence for individual differences in baseline performance (i.e., the intercept) as well as change therein (i.e., the slope). Contradictory to the hypothesis that lower initial cognitive performance yields stronger gains, we found a positive correlation between the intercept and the slope, suggesting magnification effects. Furthermore, it has been argued that individual differences may obscure transfer effects (cf. Jaeggi, Buschkuehl, Shah, & Jonides, 2014; see also Guye, et al., 2016; Katz, Jones, Shah, Buschkuehl, & Jaeggi, 2016). However, the individual differences we analyzed in Guye et al. (2017), ranging from indicators of motivation, cognition-related beliefs (e.g., grit, theories of intelligence, or need of cognition) to the personality big five (i.e., neuroticism, agreeableness, extraversion, openness, and conscientiousness), predicted the slope in neither younger nor older adults. Thus, it is rather unlikely that age or individual differences underlie the absence of transfer effects in the present study.

Conclusion

In the present study, WM updating and binding training yielded neither near nor far transfer effects, and did not affect WM processing. Thus, the findings contribute evidence that the repetitive practice of WM tasks is ineffective in increasing WM capacity and efficiency. Instead, the results suggest that training encouraged the development of stimulus-specific expertise alongside the use of paradigm-specific strategies. Taken together, the present findings add to the notion that interventions involving the mere repetitive practice of WM tasks are unlikely to elicit generalized improvements in cognition.

Context of the Research

The prospects of a relatively cheap, easy-to-administer intervention to boost cognition is highly attractive in terms of clinical and nonclinical applications, but also as a way to experimentally manipulate WM capacity. But does process-based WM training really fundamentally enhance cognition? Our previous work yielded some evidence in favor (von Bastian & Oberauer, 2013; Zimmermann et al., 2016), but also against this proposition (von

Bastian & Eschen, 2016; von Bastian et al., 2013) – a state of research that is closely mirrored by the mixed evidence from current meta-analyses (e.g., Au et al., 2015; Melby-Lervåg et al., 2016). However, as many studies featured only small sample sizes (including our own), the evidence was rather weak in either direction (cf. von Bastian et al., 2018). To shed further light on whether and under which circumstances WM training is effective, we set out to systematically examine the underlying mechanisms of WM transfer and the role of individual differences (cf. von Bastian & Oberauer, 2014) with larger-scale samples (see also Guye et al., 2017; Guye & von Bastian, 2017). Now, the results of that endeavor – that is, the consistent absence of the hypothesized effects across our studies, backed up by relatively strong Bayesian evidence – leave us to conclude that WM training interventions, as they are currently administered, are no quick-fix to enhance cognition. Nevertheless, the robust and large practice effects during WM training are still intriguing, as they go well over and above the expected range of performance. Thus, moving forward, we aim to more closely examine those practice effects to learn more about the role of expertise and task representation in WM. We hope this will not only advance our theoretical understanding of (individual differences in) WM capacity, but may also be utilized to develop more reliable interventions for improving cognitive performance.

References

Arthur, W., Jr., & Day, D. V. (1994). Development of a short form for the Raven advanced progressive matrices test. Educational and Psychological Measurement, 54(2), 394-403. doi: 10.1177/0013164494054002013

Atkinson, A., Baddeley, A. D., & Allen, R. J. (2017). Remember some or remember all? Ageing and strategy effects in visual working memory. The Quarterly Journal of Experimental Psychology. Advance online publication. doi: 10.1080/17470218.2017.1341537

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

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. Psychological Science, 18(7), 622-628. doi: 10.1111/j.1467-9280.2007.01949.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, 390–412. doi: 10.1016/j.jml.2007.12.005

Bailey, H. R., Dunlosky, J., & Hertzog, C. (2014). Does strategy-training reduce age-related deficits in working memory? Gerontology, 60(4), 346-356. doi: 10.1159/000356699

Bailey, H., Dunlosky, J., & Kane, M. J. (2008). Why does working memory span predict complex span? Testing the strategy affordance hypothesis. Memory & Cognition, 36(8), 1383-1390. doi:10.3758/MC.36.8.1383

Ball, K., Berch, D. B., Helmers, K. F., Leveck, M. D., Marsiske, M., Morris, J. N., . . . Willis, S. L. (2002). Effects of cognitive training interventions with older adults. JAMA, 288(18), 2271-2281. doi: 10.1001/jama.288.18.2271

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan, (Eds.), Adolescence and education: Vol. 5 (pp. 307–337). Greenwich, CT: Information Age.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time Constraints and Resource Sharing in Adults' Working Memory Spans. Journal of Experimental Psychology: General, 133, 83–100. doi: 10.1037/0096-3445.133.1.83

Bays, P. M. (2016). Evaluating and excluding swap errors in analogue tests of working memory. Scientific Reports, 6, 2016. doi: 10.1038/srep19203

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. Journal of Vision, 9(7), 1-11. doi: 10.1167/9.10.7

Bellander, M., Eschen, A., Lövdén, M., Martin, M., Bäckman, L., & Brehmer, Y. (2017). No evidence for improved associative memory performance following process-based associative memory training in older adults. Frontiers in Aging Neuroscience, 8, 1–11. doi: 10.3389/fnagi.2016.00326

Benedek, M., Jauk, E., Sommer, M., Arendasy, M., & Neubauer, A. C. (2014). Intelligence, creativity, and cognitive control: The common and differential involvement of executive functions in intelligence and creativity. Intelligence, 46, 73-83. doi: 10.1016/j.intell.2014.05.007

Bengson, J. J., & Luck, S. J. (2016). Effects of strategy on visual working memory capacity. Psychonomic Bulletin & Review, 23(1), 265-270. doi:10.3758/s13423-015-0891-7

Bherer, L., Kramer, A. F., Peterson, M. S., Colcombe, S., Erickson, K., & Becic, E. (2008). Transfer effects in task-set cost and dual-task cost after dual-task training in older and younger adults: further evidence for cognitive plasticity in attentional control in late adulthood. Experimental Aging Research, 34, 188–219. doi: 10.1080/03610730802070068

Blakey, E., & Carroll, D. J. (2015). A short executive function training program improves preschoolers' working memory. Frontiers in Psychology, 6, 1827. doi:10.3389/fpsyg.2015.01827

Bless, H., Wänke, M., Bohner, G., Fellhauer, R. F., & Schwarz, N. (1994). Need for cognition: eine Skala zur Erfassung von Engagement und Freude bei Denkaufgaben: need for cogntion: a scale measuring engagement and happiness in cognitive tasks. Zeitschrift für Sozialpsychologie, 25, 147–154.

Bogg, T., & Lasecki, L. (2015). Reliable gains? Evidence for substantially underpowered designs in studies of working memory training transfer to fluid intelligence. Frontiers in Psychology, 5, 1589. doi: 10.3389/fpsyg.2014.01589

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

Borkenau, P., & Ostendorf, F. (2008). NEO-Fünf-Faktoren-Inventar nach Costa und McCrea (NEO-FFI). Manual. Göttingen, Germany: Hogrefe.

Button, K. S., Ioannidis, J. P. a, Mokrysz, C., Nosek, B. a, Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. Nature Reviews. Neuroscience, 14, 365–76. doi: 10.1038/nrn3475

Brydges, C. R., Reid, C. L., Fox, A., M., & Anderson, M. (2012). A unitary executive function predicts intelligence in children. Intelligence, 40, 458-469. doi: 10.1016/j.intell.2012.05.006

Carretti, B., Borella, E., Zavagnin, M., & De Beni, R. (2011). Impact of metacognition and motivation on the efficacy of strategic memory training in older adults: Analysis of specific, transfer and maintenance effects. Archives of Gerontology and Geriatrics, 52, e192-e197. doi:10.1016/j.archger.2010.11.004

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling, 9(2), 233-255. doi: 10.1207/S15328007sem0902_5

Cooper, S. R., Gonthier, C., Barch, D. M., & Braver, T. S. (2017). The role of psychometrics in individual differences research in cognition: A case study of the AX-CPT. Frontiers in

Psychology, 8(1482), 1-16. doi: 10.3389/fpsyg.2017.01482

Cousineau, D. (2005). Confidence intervals in within-subjects designs: A simpler solution to Loftus and Masson’s method. Tutorials in Quantitative Methods for Psychology, 1(1), 42-45. doi: 10.20982/tqmp.01.1.p042

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

Cowan, N. (1995). Attention and memory: An integrated framework. New York, NY: Oxford University Press.

Cusack, R., Lehmann, M., Veldsman, M., & Mitchell, D. J. (2009). Encoding strategy and not visual working memory capacity correlates with intelligence. Psychonomic Bulletin & Review, 16(4), 641-647. doi: 10.3758/PBR.16.4.641

Dahlin, E., Neely, A. S., Larsson, A., Backman, L., & Nyberg, L. (2008). Transfer of Learning After Updating Training Mediated by the Striatum. Science, 320, 1510–1512. doi: 10.1126/science.1155466

Dahlin, E., Nyberg, L., Bäckman, L., & Neely, A. S. (2008). Plasticity of executive functioning in young and older adults: immediate training gains, transfer, and long-term maintenance. Psychology and Aging, 23, 720–30. doi: 10.1037/a0014296

De Simoni, C., Luethi, M. S., Oberauer, K., & von Bastian, C. C. (2018). No consistent evidence for ego-depletion effects across multiple cognitive tasks and domains. Manuscript in preparation.

Deci, E. L., & Ryan, R. M. (2015). Intrinsic motivation inventory. Retrieved from http://selfdeterminationtheory.org/questionnaires/10-questionnaires/50

De Simoni, C., & von Bastian, C. C. (2018, March31). Working Memory Updating and Binding Training: Bayesian Evidence Supporting the Absence of Transfer. Retrieved from osf.io/fy5ku

Dorbath, L., Hasselhorn, M., & Titz, C. (2011). Aging and executive functioning: A training study on focus-switching. Frontiers in Psychology, 2, 1–12. doi: 10.3389/fpsyg.2011.00257

Dougherty, M. R., Hamovitz, T., & Tidwell, J. W. (2015). Reevaluating the effectiveness of n-back training on transfer through the Bayesian lens: Support for the null. Psychonomic Bulletin & Review. doi: 10.3758/s13423-015-0865-9

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the short grit scale (grit-s). Journal of Personality Assessment, 91, 166–174. doi: 10.1080/00223890802634290

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

Dweck, C. S. (2000). Self-theories: their role in motivation, personality, and development.

Philadelphia, PA: Taylor & Francis.

Ecker, U. K. H., Lewandowsky, S., & Oberauer, K. (2014). Removal of information from working memory: A specific updating process. Journal of Memory and Language, 74, 77–90. doi: 10.1016/j.jml.2013.09.003

Ecker, U. K. H., Lewandowsky, S., Oberauer, K., & Chee, A. E. H. (2010). The Components of Working Memory Updating: An Experimental Decomposition and Individual Differences, 36, 170–189. doi: 10.1037/a0017891

Ekstrom, R., French, J., Harman, H., & Dermen, D. (1976). Manual for kit of factor referenced cognitive tests. Retrieved from http://www.ets.org/Media/Research/pdf/Manual_for_Kit_of_Factor-Referenced_Cognitive_Tests.pdf

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & 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., Chase, W. G., & Faloon, S. (1980). Acquisition of a Memory Skill. Science, 208, 1181–1182. doi: 10.1126/science.7375930

Estrada, E., Ferrer, E., Abad, F. J., Román, F. J., & Colom, R. (2015). A general factor of intelligence fails to account for changes in tests’ scores after cognitive practice: A longitudinal multi-group latent-variable study. Intelligence, 50, 93–99. doi: 10.1016/j.intell.2015.02.004

Foster, J. L., Harrison, T. L., Hicks, K. L., Draheim, C., Redick, T. S., & Engle, R. W. (2017). Do the effects of working memory training depend on baseline ability level? Journal of Experimental Psychology: Learning, Memory, and Cognition, 43(11), 1677-1689. doi: 10.1037/xlm0000426

Friedman, N. P., & Miyake, A. (2004a). The reading span test and its predictive power for reading comprehension ability. Journal of Memory and Language, 51(1), 136-158. doi: 10.1016/j.jml.2004.03.008

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

Friedman, N. P., Miyake, A., Schmeichel, B. J., & Tang, D. (2006). Not all executive functions are related to intelligence. Psychological Science, 17, 172–179. doi: 10.1111/j.1467-9280.2006.01681.x

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2009). Individual differences in executive functions are almost entirely genetic in origin. Journal of Experimental Psychology: General, 137(2), 201–225. doi: 10.1037/0096-3445.137.2.201

Gibson, B. S., Gondoli, D. M., Johnson, A. C., Steeger, C. M., & Morrissey, R. A. (2012). The future promise of Cogmed working memory training. Journal of Applied Research in Memory and Cognition, 1, 214–216. doi: 10.1016/j.jarmac.2012.07.003

Goghari, V. M., & Lawlor-Savage, L. (2017). Comparison of cognitive change after working memory training and logic and planning training in healthy older adults. Frontiers in Aging Neuroscience, 9(39), 1-12. doi: 10.3389/fnagi.2017.00039

Green, C. S., Strobach, T., & Schubert, T. (2014). On methodological standards in training and transfer experiments. Psychological Research, 78, 756–772. doi: 10.1007/s00426-013-0535-3

Guye, S., De Simoni, C., & von Bastian, C. C. (2017). Do individual differences predict change in cognitive training performance? A latent growth curve modeling approach. Journal of Cognitive Enhancement, 1(4), 374-393. doi: 10.1007/s41465-017-0049-9

Guye, S., Röcke, C., Mérillat, S., von Bastian, C. C., & Martin, M. (2016). Plasticity in different age groups: Adult lifespan. In T. Strobach & J. Karbach (Eds.), Cognitive Training: An Overview of Features and Applications. Berlin: Springer.

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

Halsey, L. G., Curran-Everett, D., Vowler, S. L., & Drummond, G. B. (2015). The fickle P value generates irreproducible results. Nature Methods, 12, 179–185. doi: 10.1038/nmeth.3288

Hardy, J. L., Nelson, R. A., Thomason, M. E., Sternberg, D. A., Katovich, K., Farzin, F., & Scanlon, M. (2015). Enhancing cognitive abilities with comprehensive training: A large, online, randomized, active-controlled trial. PLoS One, 10(9), e0134467. doi: 10.1371/journal.pone.0134467

Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. Psychological Science, 24, 2409–2419. doi: 10.1177/0956797613492984

Healy, A. F., Wohldmann, E. L., Sutton, E. M., & Bourne, L. E. (2006). Specificity effects in training and transfer of speeded responses. Journal of Experimental Psychology. Learning, Memory, and Cognition, 32, 534–546. doi: 10.1037/0278-7393.32.3.534

Hedge, C., Powell, G., & Sumner, P. (in press). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. Behavior Research Methods. doi: 10.3758/s13428-017-0935-1

Herkert, J. (2012). Trainings- und Transfereffekte des Arbeitsgedächtnisses. (Unpublished master's thesis). University of Zurich, Zurich, Switzerland.

Hertzog, C., Price, J., & Dunlosky, J. (2008). How is knowledge generated about memory encoding strategy effectiveness? Learning and Individual Differences, 18(4), 430-445. doi: 10.1016/j.lindif.2007.12.002

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6, 1–55. doi: 10.1080/10705519909540118

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. Neuropsychologia, 44, 2017-2036.

Hull, R., Martin, R. C., Beier, M. E., Lane, D., & Hamilton, A. C. (2008). Executive function in older adults: A structural equation modeling approach. Neuropsychology, 22(4), 508-522. doi: 10.1037/0894-4105.22.4.508

Hussey, E. K., Isaiah Harbison, J., Teubner-Rhodes, S. E., Mishler, A., Velnoskey, K., & Novick, J. M. (2016). Memory and Language Improvements Following Cognitive Control Training. Journal of Experimental Psychology: Learning, Memory, and Cognition, 42. doi: 10.1037/xlm0000283

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 of the United States of America, 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, 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

Jeffreys, H. (1961). Theory of probability (3rd ed.). New York: NY: Oxford University Press.

Judd, C. H. (1908). The relation of special training to general intelligence. Educational Review, 88(197-227).

Kane, M. J., Poole, B. J., Tuholski, S. W., & Engle, R. W. (2006). Working memory capacity and the top-down control of visual search: Exploring the boundaries of “executive attention”. Journal of Experimental Psychology. Learning, Memory, and Cognition, 32, 749–777. doi: 10.1037/0278-7393.32.4.749

Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. Developmental Science, 12, 978–990. doi: 10.1111/j.1467-7687.2009.00846.x

Katz, B., Jones, M. R., Shah, P., Buschkuehl, M., & Jaeggi, S. M. (2016). Individual differences and motivational effects in cognitive training research. In T. Strobach & J. Karbach (Eds.), Cognitive training: An overview of features and applications (pp. 157-166). Berlin: Springer.

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

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., Gillberg, C. S., Forssberg, H., & Westerberg, H. (2005). Computerized training of working memory in children with ADHD - a randomized, controlled trial. Journal of the American Academy of Child & Adolescent Psychiatry, 44(2), 177-186. doi: 10.1097/00004583-200502000-00010

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

Kovacs, K., & Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. Psychological Inquiry, 27(3), 151-177. doi: 10.1080/1047840X.2016.1153946

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! Intelligence, 14, 389–433. doi: 10.1016/S0160-2896(05)80012-1

Kramer, A. F., Larish, J. F., & Strayer, D. L. (1995). Training for attentional control in dual task settings: A comparison of young and old adults. Journal of Experimental Psychology: Applied, 1(1), 50-76. doi: 10.1037/1076-898X.1.1.50

Krumm, S., Schmidt-Atzert, L., Buehner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. Intelligence, 37, 347-364. doi: 10.1016/j.intell.2009.02.003

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

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. doi: 10.1371/journal.pmed.1001756

Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. H. (2010). A working memory test battery for MATLAB. Behavior Research Methods, 42, 571–585. doi: 10.3758/BRM.42.2.571

Lewandowsky, S., & Thomas, J. (2009). Expertise: Acquisition, limitations, and control. Reviews of Human Factors and Ergonomics, 5, 140–165. doi: 10.1518/155723409X448044.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology, 49, 764–766. doi: 10.1016/j.jesp.2013.03.013

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, 135–41. doi: 10.3758/s13423-012-0335-6

Linares, R., Borella, E., Lechuga, M. T., Carretti, B., & Pelegrina, S. (2017). Training working memory updating in young adults. Psychological Research. Advance online publication. doi: 10.1007/s00426-017-0843-0

Linke, A. C., Vicente-Grabovetsky, A., Mitchell, D. J., & Cusack, R. (2011). Encoding strategy accounts for individual differences in change detection measures of VSTM. Neuropsychologia, 49(6), 1476-1486. doi: 10.1016/j.neuropsychologia.2010.11.034

Logie, R. H., Baddeley, A. D., Mané, A., Donchin, E., & Sheptak, R. (1989). Working memory in the acquisition of complex cognitive skills. Acta Psychologica, 71, 179-196. doi: 10.1016/0001-6918(89)90005-X

Logie, R. H., Della Sala, S., Laiacona, M., Chalmers, P., & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. Memory & Cognition, 24(3), 305-321. doi: 10.3758/BF03213295

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

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. Annual Review of Psychology, 60, 577-605. doi: 10.1146/annurev.psych.60.110707.163612

McAuley, T., & White, D. A. (2011). A latent variables examination of processing speed, response inhibition, and working memory during typical development. Journal of Experimental Child Psychology, 108, 453-468. doi: 10.1016/j.jecp.2010.08.009

McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. Journal of Experimental Psychology: General, 141(2), 302-320. doi: 10.1037/a0025250

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

Melby-Lervåg, M., Redick, T., & 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, 512–534. doi: 10.1177/1745691616635612

Minear, M., Brasher, F., Guerrero, C. B., Brasher, M., Moore, A., & Sukeena, J. (2016). A simultaneous examination of two forms of working memory training: Evidence for near transfer only. Memory & Cognition, 44, 1014-1037. doi: 10.3758/s13421-016-0616-9

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

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

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). Tutorials in Quantitative Methods for Psychology, 4(2), 61-64. doi: 10.20982/tqmp.04.2.p061Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs. Retrieved from https://cran.r-project.org/package=BayesFactor

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, 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

Navon, D. (1977). Forest before trees: The precedence of global fetures in visual perception. Cognitive Psychology, 9, 353–383.

Noack, H., Lövdén, M., & Schmiedek, F. (2014). On the validity and generality of transfer effects in cognitive training research. Psychological Research, 78, 773–789. doi: 10.1007/s00426-014-0564-6

Oberauer, K. (2005). Binding and inhibition in working memory: individual and age differences in short-term recognition. Journal of Experimental Psychology: General, 134, 368–387. doi: 10.1037/0096-3445.134.3.368

Oberauer, K. (2006). Is the focus of attention in working memory expanded through practice? Journal of Experimental Psychology. Learning, Memory, and Cognition, 32, 197–214. doi: 10.1037/0278-7393.32.2.197

Oberauer, K. (2008). How to say no: Single- and dual-process theories of short-term recognition tested on negative probes. Journal of Experimental Psychology. Learning, Memory, and Cognition, 34, 439–459. doi: 10.1037/0278-7393.34.3.439

Oberauer, K. (2009). Design for a working memory. Psychology of Learning and Motivation, 51, 45-100. doi:10.1016/S0079-7421(09)51002-X

Oberauer, K., & Hein, L. (2012). Attention to Information in Working Memory. Current Directions in Psychological Science, 21, 164–169. doi: 10.1177/0963721412444727

Oberauer, K., Lange, E., & Engle, R. W. (2004). Working memory capacity and resistance to interference. Journal of Memory and Language, 51, 80–96. doi: 10.1016/j.jml.2004.03.003

Oberauer, K. (2008). How to say no: Single- and dual-process theories of short-term recognition tested on negative probes. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(3), 439–459. doi: 10.1037/0278-7393.34.3.439

Oberauer, K., & Lewandowsky, S. (2016). Control of information in working memory: Encoding and removal of distractors in the complex-span paradigm. Cognition, 156, 106–128. doi: 10.1016/j.cognition.2016.08.007

Oberauer, K., Süss, H.-M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), Variation in Working Memory (pp. 49–75). Oxford: Oxford University Press.

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

Oelhafen, S., Nikolaidis, A., Padovani, T., Blaser, D., Koenig, T., & Perrig, W. J. (2013). Increased parietal activity after training of interference control. Neuropsychologia, 51, 2781–2790. doi: 10.1016/j.neuropsychologia.2013.08.012

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., Howard, R. J., & Ballard, C. G. (2010). Putting brain training to the test. Nature, 465, 775-779. doi: 10.1038/nature09042

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. Cognitive Psychology, 66, 232-258. doi: 10.1016/j.cogpsych.2012.12.002

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

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, 359–379. doi: 10.1037/a0029082

Rey-Mermet, A., Gade, M., & Oberauer, K. (in press). Should we stop thinking about inhibition? Journal of Experimental Psychology: Learning, Memory, and Cognition. doi: 10.1037/xlm0000450

Rey-Mermet, A., Gade, M., Souza, A. S., von Bastian, C. C., & Oberauer, K. (2017). Executive control is not related to working memory capacity and fluid intelligence. Manuscript in preparation.

Rheinberg, F., Vollmeyer, R., & Burns, B. D. (2001). FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen. Diagnostica, 47, 57–66. doi: doi: 10.1026//0012-1924.47.2.57

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. Journal of Statistical Software, 48, 1–36. Retrieved from http://www.jstatsoft.org/v48/i02/

Shipstead, Z., Lindsey, D. R. B., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. Journal of Memory and Language, 72, 116-141. doi: 10.1016/j.jml.2014.01.004

Schmiedek, F., Lövdén, 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, 1–10. doi: 10.3389/fnagi.2010.00027

Schmiedek, F., Oberauer, K., Wilhelm, O., Süss, H. M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. Journal of Experimental Psychology: General, 136(3), 414–429. doi: 10.1037/0096-3445.136.3.414

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

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portf*olio. Causal and control beliefs (pp. 35–37). Windsor, UK: NFER-NELSON.

Shipstead, Z., Hicks, K. L., & Engle, R. W. (2012). Working memory training remains a work in progress. Journal of Applied Research in Memory and Cognition, 1, 217–219. doi: 10.1016/j.jarmac.2012.07.009

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? Psychological Bulletin, 138, 628–654. doi: 10.1037/a0027473\r10.1037/0033-

2909.131.1.30

Simon, J. R., Sly, P. E., & Vilapakkam, S. (1981). Toward the Stimulus Source. Acta Psychologica, 47, 63–81.

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, 103–186. doi: 10.1177/1529100616661983

Sobel, K. V, Gerrie, M. P., Poole, B. J., & Kane, M. J. (2007). Individual differences in working memory capacity and visual search: the roles of top-down and bottom-up processing. Psychonomic Bulletin & Review, 14, 840–845. doi: 10.3758/BF03194109

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 & Review, 24, 1077-1096. doi: 10.3758/s13423-016-1217-0

Sprenger, A. M., Atkins, S. M., Bolger, D. J., Harbison, J. I., Novick, J. M., Chrabaszcz, J. S., Weems, S. A., Smith, V., Bobb, S., Bunting, M. F., & Dougherty, M. R. (2013). Training working memory: Limits of transfer. Intelligence, 41, 638–663. doi: 10.1016/j.intell.2013.07.013

Sternberg, R. J. (2008). Increasing fluid intelligence is possible after all. Proceedings of the National Academy of Sciences of the United States of America, 105(19), 6791-6792. doi: 10.1073/pnas.0803396105

Stroop, R. J. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18(6), 643-662. doi: 10.1037/h0054651

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

Szmalec, A., Verbruggen, F., Vandierendonck, A., & Kemps, E. (2011). Control of interference during working memory updating. Journal of Experimental Psychology. Human Perception and Performance, 37, 137–151. doi: 10.1037/a0020365

Tennsted, S. L., & Unverzagt, F. W. (2013). The ACTIVE study: Study overview and major findings. Journal of Aging and Health, 25(8), 3S-20S. doi: 10.1177/0898264313518133

Titz, C., & Karbach, J. (2014). Working memory and executive functions: effects of training on academic achievement. Psychological Research, 78, 852–868. doi: 10.1007/s00426-013-0537-1

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 166–179. doi: 10.1037/0278-7393.18.1.166

Unsworth, N. (2016). Working memory capacity and recall from long-term memory: Examining the influences of encoding strategies, study time allocation, search efficiency, and monitoring abilities. Journal of Experimental Psychology: Learning, Memory, and Cognition, 42(1), 50-61. doi:10.1037/xlm0000148

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. Cognitive Psychology, 71, 1-26. doi: 10.1016/j.cogpsych.2014.01.003

Unsworth, N., & McMillan, B. D. (2014). Similarities and differences between mind-wandering and external distraction: A latent variable analysis of lapses of attention and their relation to cognitive abilities. Acta Psychologica, 150, 14-25. doi: 10.1016/j.actpsy.2014.04.001

Verhaeghen, P., Cerella, J., & Basak, C. (2004). A working memory workout: how to expand the focus of serial attention from one to four items in 10 hours or less. Journal of Experimental Psychology. Learning, Memory, and Cognition, 30, 1322–37. doi: 10.1037/0278-7393.30.6.1322

von Bastian, C. C., Guye, S., & De Simoni, C. (2018). How strong is the evidence for the effectiveness of working memory training? Manuscript accepted for publication in M. F. Bunting, J. M. Novick, M. R. Dougherty, & R. W. Engle (Eds.), Cognitive and Working Memory Training: Perspectives from Psychology, Neuroscience, and Human Development. New York, NY: Oxford University Press.

von Bastian, C. C., & Druey, M. D. (2017). Shifting between mental sets: An individual differences approach to commonalities and differences of task switching components. Journal of Experimental Psychology: General, 146(9), 1266-1285. doi:10.1037/xge0000333

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., Langer, N., Jäncke, L., & Oberauer, K. (2013). Effects of working memory training in young and old adults. Memory & Cognition, 41, 611–624. doi: 10.3758/s13421-012-0280-7

von Bastian, C. C., Locher, A., & Ruflin, M. (2012). Tatool: A Java-based open-source programming framework for psychological studies. Behavior Research Methods, 45, 108–115. doi: 10.3758/s13428-012-0224-y

von Bastian, C. C., & Oberauer, K. (2013). Distinct transfer effects of training different facets of working memory capacity. Journal of Memory and Language, 69, 36–58.

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

von Bastian, C. C., Souza, A. S., & Gade, M. (2016). No evidence for bilingual cognitive advantages: A test of four hypotheses. Journal of Experimental Psychology: General, 145(2), 246-258. doi: 10.1037/xge0000120

Waris, O., Soveri, A., & Laine, M. (2015). Transfer after working memory updating training. PLoS ONE, 10, 1–19. doi: 10.1371/journal.pone.0138734

Weicker, J., Villringer, A., & Thöne-Otto, A. (2015). Can impaired working memory functioning be improved by training? A meta-analysis with a special focus on brain injured patients. Neuropsychology, 30, 1–23. doi: 10.1037/neu0000227

Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. Psychonomic Bulletin & Review, 1057–1064. doi: 10.3758/s13423-012-0295-x

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? Frontiers in Psychology, 4, 1–22. doi: 10.3389/fpsyg.2013.00433

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. Nature, 440(2), 91-95. doi: 10.1038/nature04262

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. Nature, 453(8), 233-236. doi: 10.1038/nature06860

Zimmermann, K., von Bastian, C. C., Röcke, C., Martin, M., & Eschen, A. (2016). Transfer after process-based object-location memory training in healthy older adults. Psychology and Aging, 31(7), 798–814. doi: 10.1037/pag0000123

Appendix

Table A1
Excluded Data from Pretest and Posttest

Task	Pretest			Posttest		
	UPD	BIN	AC	UPD	BIN	AC
Updating						
Accuracies						
Arrows			1		1	
Locations		1				1
Focus Switching						
Digits	1	1			2	1
Letters	2	2	1		2	2
Arrows	1	9	5		7	5
Locations		1	3		1	2
Binding						
Symbol-Digit						1
Fractal-Location				1		
Color-Location		1				
Visual Search						
Letters	1	4	1	5	5	
Arrows			1	1		
Circles			2			
Removal						
Arrows		1	1			
Reasoning						
Diagramming Relationships						1
Letter Sets					1	
Locations Test					1	1
RAPM						1
Shifting						
Parity-Magnitude					1	
Color-Shape				1	1	
Fill-Frame		1	1			
Inhibition						
Global-Local	3	1	3	2		2

Note. UPD = updating group; BIN = binding group; AC = active control group.

Table A2
Task Performance as a Function of Training Group and Time of Assessment

	Group							
	Updating		Binding		Active Control		Reliability	
Tasks	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Updating								
Digits	0.75 (0.16)	0.80 (0.16)	0.71 (0.20)	0.75 (0.20)	0.72 (0.19)	0.74 (0.20)	0.89	0.91
Letters	0.59 (0.21)	0.73 (0.17)	0.49 (0.21)	0.53 (0.22)	0.56 (0.19)	0.57 (0.20)	0.86	0.89
Arrows	0.40 (0.16)	0.49 (0.18)	0.35 (0.17)	0.41 (0.19)	0.38 (0.15)	0.43 (0.16)	0.83	0.84
Locations	0.43 (0.17)	0.58 (0.19)	0.39 (0.21)	0.46 (0.22)	0.39 (0.19)	0.46 (0.20)	0.89	0.91
Binding								
Symbol-Digit	1.00 (0.52)	1.01 (0.64)	0.87 (0.54)	1.13 (0.84)	0.91 (0.51)	1.03 (0.61)	0.7	0.83
Noun-Verb	1.46 (0.72)	1.47 (0.81)	1.51 (0.71)	1.48 (0.76)	1.36 (0.72)	1.34 (0.80)	0.81	0.84
Fractal-Location	0.87 (0.51)	1.03 (0.70)	0.92 (0.53)	1.77 (0.99)	0.91 (0.51)	0.99 (0.63)	0.64	0.88
Color-Location	1.12 (0.72)	1.48 (0.87)	1.08 (0.63)	1.52 (0.80)	1.04 (0.60)	1.18 (0.67)	0.8	0.85
Visual Search[a]								
Numbers	-0.03 (0.22)	0.00 (0.18)	0.03 (0.17)	0.07 (0.20)	0.00 (0.19)	-0.06 (0.17)	0.73	0.75
Letters	-0.03 (0.18)	0.02 (0.20)	0.01 (0.16)	0.07 (0.19)	0.01 (0.15)	-0.07 (0.16)	0.72	0.73
Arrows	-0.02 (0.17)	0.03 (0.19)	0.02 (0.17)	0.03 (0.19)	-0.01 (0.18)	-0.05 (0.17)	0.68	0.76
Circles	-0.02 (0.14)	-0.01 (0.15)	0.02 (0.13)	0.06 (0.14)	0.00 (0.14)	-0.04 (0.14)	0.42	0.43
Reasoning								
Relationships	0.74 (0.14)	0.75 (0.13)	0.70 (0.14)	0.73 (0.17)	0.74 (0.15)	0.77 (0.13)	0.72	0.74
Letter Sets	0.73 (0.12)	0.74 (0.16)	0.72 (0.15)	0.74 (0.15)	0.71 (0.16)	0.73 (0.14)	0.66	0.77
Locations	0.53 (0.16)	0.59 (0.17)	0.46 (0.18)	0.53 (0.19)	0.49 (0.17)	0.58 (0.17)	0.63	0.7
RAPM	0.59 (0.23)	0.65 (0.23)	0.56 (0.25)	0.61 (0.20)	0.60 (0.19)	0.63 (0.18)	0.72	0.65
Shifting[a]								
Parity-Magnitude	0.37 (0.14)	0.35 (0.16)	0.29 (0.15)	0.30 (0.16)	0.31 (0.17)	0.32 (0.17)	0.87	0.9
Animacy-Size	0.40 (0.16)	0.42 (0.13)	0.33 (0.15)	0.32 (0.15)	0.36 (0.17)	0.35 (0.16)	0.82	0.88
Color-Shape	0.35 (0.18)	0.34 (0.17)	0.26 (0.16)	0.28 (0.14)	0.30 (0.17)	0.34 (0.17)	0.87	0.88
Fill-Frame	0.29 (0.13)	0.32 (0.13)	0.22 (0.12)	0.26 (0.15)	0.28 (0.15)	0.27 (0.13)	0.84	0.81

Speed[a]								
Parity-Magnitude	6.27 (0.08)	6.23 (0.08)	6.29 (0.13)	6.23 (0.13)	6.27 (0.12)	6.23 (0.13)	0.99	0.99
Animacy-Size	6.39 (0.09)	6.32 (0.08)	6.40 (0.13)	6.34 (0.14)	6.38 (0.11)	6.32 (0.12)	0.99	0.99
Fill-Frame	6.19 (0.09)	6.12 (0.09)	6.21 (0.12)	6.16 (0.13)	6.18 (0.13)	6.15 (0.15)	0.99	0.99
Color-Shape	6.23 (0.08)	6.14 (0.08)	6.25 (0.11)	6.17 (0.11)	6.23 (0.12)	6.15 (0.11)	0.99	0.99
Inhibition[a]								
Numerical Stroop	0.05 (0.04)	0.05 (0.04)	0.07 (0.06)	0.06 (0.04)	0.06 (0.05)	0.06 (0.04)	0.69	0.65
Color Stroop	0.04 (0.07)	0.02 (0.05)	0.06 (0.07)	0.04 (0.06)	0.06 (0.07)	0.04 (0.07)	0.79	0.74
Global-Local	0.02 (0.06)	0.03 (0.05)	0.05 (0.06)	0.03 (0.06)	0.04 (0.05)	0.04 (0.06)	0.74	0.74
Simon	0.03 (0.01)	0.02 (0.01)	0.03 (0.02)	0.03 (0.02)	0.03 (0.02)	0.02 (0.02)	0.39	0.47
Focus switching[a]								
Digits	0.38 (0.15)	0.32 (0.13)	0.35 (0.17)	0.35 (0.16)	0.38 (0.20)	0.37 (0.16)	0.79	0.82
Letters	0.42 (0.19)	0.31 (0.11)	0.38 (0.23)	0.43 (0.20)	0.38 (0.21)	0.40 (0.18)	0.62	0.7
Arrows	0.54 (0.28)	0.52 (0.17)	0.57 (0.31)	0.57 (0.30)	0.53 (0.26)	0.54 (0.25)	0.73	0.8
Locations	0.48 (0.17)	0.53 (0.19)	0.52 (0.21)	0.53 (0.22)	0.50 (0.18)	0.51 (0.19)	0.82	0.86
Removal[a]								
Digits	-0.01 (0.14)	0.00 (0.13)	-0.01 (0.14)	0.00 (0.14)	0.01 (0.13)	0.00 (0.14)	0.86	0.85
Letters	-0.01 (0.14)	0.01 (0.11)	0.01 (0.14)	0.00 (0.14)	0.00 (0.11)	0.00 (0.13)	0.76	0.82
Arrows	-0.02 (0.11)	0.00 (0.10)	0.00 (0.11)	-0.01 (0.12)	0.02 (0.12)	0.01 (0.13)	0.57	0.64
Locations	0.00 (0.09)	-0.01 (0.06)	0.01 (0.09)	0.00 (0.09)	-0.01 (0.09)	0.01 (0.08)	0.63	0.65
Interference Resolution								
Updating[a]								
Digits	0.10 (0.07)	0.08 (0.08)	0.11 (0.09)	0.11 (0.08)	0.10 (0.07)	0.11 (0.09)	0.73	0.78
Letters	0.13 (0.08)	0.09 (0.07)	0.15 (0.08)	0.16 (0.08)	0.13 (0.07)	0.13 (0.06)	0.59	0.59
Arrows	0.18 (0.07)	0.17 (0.08)	0.19 (0.08)	0.16 (0.06)	0.18 (0.07)	0.17 (0.07)	0.38	0.42
Locations	0.17 (0.05)	0.15 (0.07)	0.17 (0.07)	0.15 (0.07)	0.17 (0.06)	0.16 (0.07)	0.42	0.54
Binding								
Symbol-digit	0.65 (0.14)	0.67 (0.14)	0.61 (0.14)	0.65 (0.20)	0.61 (0.13)	0.69 (0.13)	0.6	0.75
Noun-verb	0.71 (0.15)	0.73 (0.14)	0.72 (0.15)	0.73 (0.14)	0.69 (0.17)	0.72 (0.16)	0.74	0.74
Fractal-location	0.59 (0.14)	0.64 (0.16)	0.62 (0.14)	0.71 (0.19)	0.63 (0.15)	0.68 (0.14)	0.6	0.76
Color-location	0.65 (0.17)	0.72 (0.18)	0.64 (0.16)	0.68 (0.18)	0.65 (0.16)	0.72 (0.15)	0.75	0.79

Other Updating Errors								
Extra-list errors								
Digits	0.16 (0.10)	0.12 (0.10)	0.19 (0.13)	0.15 (0.13)	0.18 (0.13)	0.15 (0.13)	0.84	0.87
Letters	0.29 (0.15)	0.18 (0.12)	0.36 (0.16)	0.32 (0.16)	0.31 (0.14)	0.30 (0.16)	0.81	0.86
Arrows	0.42 (0.14)	0.35 (0.14)	0.47 (0.14)	0.42 (0.16)	0.45 (0.13)	0.40 (0.13)	0.76	0.78
Locations	0.39 (0.14)	0.27 (0.14)	0.45 (0.16)	0.39 (0.18)	0.44 (0.15)	0.38 (0.15)	0.85	0.88
Other Binding Probes								
Matches								
Symbol-digit	0.72 (0.10)	0.69 (0.10)	0.71 (0.10)	0.72 (0.13)	0.72 (0.11)	0.67 (0.14)	0.69	0.82
Noun-verb	0.78 (0.11)	0.75 (0.12)	0.78 (0.10)	0.76 (0.13)	0.76 (0.10)	0.72 (0.13)	0.78	0.85
Fractal-location	0.72 (0.09)	0.73 (0.10)	0.71 (0.12)	0.84 (0.09)	0.70 (0.11)	0.67 (0.14)	0.71	0.85
Color-location	0.74 (0.10)	0.78 (0.10)	0.74 (0.12)	0.83 (0.09)	0.72 (0.11)	0.71 (0.13)	0.77	0.78
Distractors								
Symbol-digit	0.73 (0.12)	0.75 (0.13)	0.72 (0.14)	0.76 (0.15)	0.74 (0.13)	0.77 (0.13)	0.7	0.75
Noun-verb	0.85 (0.11)	0.85 (0.12)	0.84 (0.14)	0.86 (0.14)	0.83 (0.12)	0.84 (0.12)	0.71	0.78
Fractal-location	0.67 (0.12)	0.72 (0.14)	0.69 (0.13)	0.80 (0.13)	0.69 (0.15)	0.73 (0.13)	0.69	0.74
Color-location	0.78 (0.13)	0.83 (0.13)	0.76 (0.13)	0.82 (0.12)	0.79 (0.11)	0.82 (0.13)	0.77	0.75

Note. Values are Ms ± SDs. Scores are accuracies (proportion correct) or proportion of errors (extra-list errors), except for binding (d'), visual search and removal (residuals), inhibition, shifting and focus switching (log-transformed reaction time differences), and updating intrusions (proportion of errors). Reliabilities for accuracies and proportion of errors are Cronbach's alpha; all other reliabilities are split-half reliabilities corrected with the Spearman-Brown prophecy formula.

[a]Lower values reflect better performance.

Table A3

Correlations between Tasks at Pretest and Posttest

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Updating																											
1. Digits		.70	.59	.54	.40	.34	.38	.49	.03	-.03	.02	.00	.63	.33	.47	.45	.04	.15	.20	.01	-.29	-.28	-.30	-.27	.03	-.29	-.03
2. Letters	.58		.61	.66	.35	.29	.34	.51	.00	-.01	.02	-.10	.52	.36	.49	.42	.06	.17	.26	.09	-.35	-.31	-.37	-.31	-.02	-.27	-.03
3. Arrows	.50	.57		.62	.44	.40	.39	.55	.06	-.01	.02	-.06	.47	.37	.41	.37	.01	.06	.21	.00	-.34	-.30	-.35	-.32	-.01	-.24	-.01
4. Locations	.49	.58	.56		.46	.42	.51	.69	.01	.04	.03	-.01	.45	.38	.46	.32	-.07	.06	.11	-.06	-.26	-.28	-.28	-.26	-.12	-.31	-.01
Binding																											
5. Symbol-Digit	.29	.34	.33	.42		.53	.50	.50	-.06	-.06	.01	-.03	.37	.15	.35	.18	-.10	-.02	.10	.11	-.16	-.28	-.21	-.19	.07	-.22	.01
6. Noun-Verb	.42	.31	.39	.41	.42		.39	.39	-.01	-.01	.14	.11	.23	.28	.33	.17	-.10	.01	.03	.15	-.18	-.20	-.17	-.16	.05	-.19	.04
7. Fractal-Location	.35	.41	.32	.56	.38	.42		.67	.15	.13	.09	.07	.39	.27	.32	.25	-.07	-.04	-.01	-.03	-.30	-.32	-.23	-.27	-.03	-.25	-.05
8. Color-Location	.45	.44	.45	.66	.37	.29	.49		.04	.08	.09	.04	.42	.29	.37	.24	-.12	-.02	.02	-.13	-.27	-.27	-.24	-.28	-.09	-.30	-.03
Visual Search																											
9. Digits	-.01	-.03	.06	-.05	.01	-.05	-.04	-.06		.45	.50	.41	.14	.00	.13	.08	-.05	-.10	-.01	-.06	-.04	.00	.00	.01	.01	.02	-.04
1. Letters	-.12	-.04	.01	.00	-.11	-.15	-.01	.01	.35		.49	.36	.01	-.06	.09	-.07	-.07	-.07	.01	-.02	.02	.00	.04	.04	-.10	.10	-.03
11. Arrows	-.06	-.04	.03	-.07	-.04	-.02	-.10	.01	.55	.45		.46	.06	.03	.15	.11	.00	-.03	.03	-.01	.03	.03	.00	-.01	-.02	.01	-.14
12. Circles	-.05	-.05	.04	.05	-.04	-.01	-.04	.08	.31	.46	.35		-.01	.02	.09	.04	-.11	-.11	-.09	-.06	.04	.09	.13	.14	-.04	.11	-.12
Reasoning																											
13. Relationships	.44	.40	.31	.33	.22	.27	.31	.29	.01	-.04	.02	-.06		.30	.59	.54	.05	.13	.15	-.02	-.26	-.23	-.25	-.22	-.01	-.26	-.09
14. Letter Sets	.21	.20	.23	.39	.23	.29	.33	.25	-.15	-.09	-.14	-.06	.37		.44	.35	-.01	.06	.14	.05	-.23	-.22	-.22	-.32	-.01	-.21	-.13
15. Locations	.41	.36	.37	.38	.26	.23	.27	.30	-.08	-.17	-.05	-.10	.47	.29		.47	-.01	.01	.17	.00	-.14	-.13	-.10	-.12	-.01	-.24	-.07
16. RAPM	.32	.31	.28	.25	.11	.19	.17	.22	-.02	-.10	-.05	.03	.47	.42	.41		.07	.07	.15	.04	-.19	-.16	-.16	-.10	.05	-.15	-.15
Shifting																											
17. Parity-Magnitude	.12	.12	.23	.06	-.11	.03	.02	.03	-.08	.04	.00	-.13	.17	.13	.10	.16		.58	.54	.39	.02	-.02	-.05	-.09	-.02	-.04	-.21
18. Animacy-Size	.00	.06	.16	.06	.06	.11	.02	.07	-.08	.12	.14	-.06	.06	.09	.06	.05	.51		.54	.49	-.03	-.12	-.12	-.19	.00	-.18	-.13
19. Color-Shape	.10	.16	.22	.11	.00	.04	.09	.12	-.06	.15	.10	.04	.13	.10	.09	.07	.51	.61		.45	-.13	-.14	-.17	-.20	.02	-.12	-.19
20. Fill-Frame	.06	-.01	.08	-.01	.10	.05	-.01	-.07	-.09	.05	.09	-.12	.09	.10	.05	.06	.39	.57	.52		-.03	-.11	-.05	-.11	.13	-.07	-.07
Speed																											
21. Parity-Magnitude	-.37	-.34	-.38	-.34	-.26	-.25	-.25	-.39	.07	.08	.13	.05	-.25	-.35	-.26	-.26	-.10	-.12	-.13	-.03		.76	.80	.73	-.14	.44	-.15
22. Animacy-Size	-.29	-.32	-.31	-.33	-.31	-.21	-.29	-.32	.10	.06	.18	.05	-.18	-.25	-.17	-.18	.01	-.12	-.10	.00	.77		.84	.81	-.11	.36	-.13
23. Color-Shape	-.37	-.35	-.34	-.35	-.30	-.26	-.26	-.36	.12	.09	.17	.09	-.25	-.32	-.15	-.21	-.04	-.14	-.16	.02	.79	.85		.83	-.05	.40	-.09
24. Fill-Frame	-.29	-.27	-.36	-.37	-.27	-.24	-.17	-.36	.12	.10	.08	.07	-.26	-.30	-.26	-.22	-.09	-.18	-.18	-.12	.71	.77	.78		-.07	.41	.00
Inhibition																											
25. Numerical Stroop	-.04	.05	-.12	-.08	.06	-.07	-.01	.04	-.06	.01	.08	.02	-.09	-.15	.03	-.09	-.07	.02	-.08	.05	-.08	-.14	-.10	-.03		.04	.05
26. Color Stroop	-.20	-.23	-.26	-.30	-.14	-.18	-.25	-.23	.08	-.03	.15	-.02	-.02	-.25	-.09	-.10	.02	-.07	-.09	.02	.43	.42	.47	.37	.05		-.04
27. Global-Local	-.10	-.08	-.11	-.09	.06	-.08	-.03	-.15	.04	-.02	-.06	.05	-.10	-.09	-.16	-.14	-.05	-.08	-.12	-.03	-.02	-.09	-.06	.04	.13	.02	
28. Simon	-.06	-.02	-.10	-.27	.00	-.07	-.06	-.17	.09	.08	.21	-.11	.06	-.09	-.02	-.05	.01	.00	-.07	.00	.15	.20	.14	.21	-.03	.08	.00

Note. Correlations at pretest are illustrated below the diagonal and correlations at posttest above.