Mem Cogn (2013) 41:611–624
DOI 10.3758/s13421-012-0280-7

Effects of working memory training in young and old adults

Claudia C. von Bastian · Nicolas Langer · Lutz Jäncke · Klaus Oberauer

Published online: 22 December 2012
© Psychonomic Society, Inc. 2012

Abstract Many cognitive abilities, including working memory and reasoning ability, decline with progressing age. In this study, we investigated whether four weeks of intensive working memory training would enhance working memory and reasoning performance in an age-comparative setting. Groups of 34 young (19–36 years) and 27 older (62–77 years) adults practiced tasks representing the three functional categories in the facet model of working memory capacity: storage and processing, relational integration, and supervision. The data were compared to those of a young and an old active control group who practiced tasks with low working memory demands. A cognitive test battery measuring near and far transfer was administered before and after training. Both age groups showed increased working memory performance in the trained tasks and in one structurally similar, but non-trained, task. Young adults also improved in a task measuring word-position binding in working memory. However, we found no far transfer to reasoning in either age group. The results provide evidence that working memory performance can be improved throughout the life span. However, in contrast to a previous study in which each facet of working memory capacity was trained separately, the present study showed that training multiple functional categories simultaneously induces less transfer.

Keywords Aging · Cognitive training · Reasoning · Transfer · Working memory capacity

Working memory (WM) is a cognitive system that provides temporary access to representations, and thereby builds the basis for complex cognition. Given that WM is an excellent predictor for a wide range of cognitive abilities, especially for reasoning (Conway, Kane & Engle, 2003; Engle, Kane, & Tuholski, 1999; Kyllonen & Christal, 1990; Oberauer, Süß, Wilhelm & Wittmann, 2008; Süß, Oberauer, Wittmann, Wilhelm & Schulze, 2002), a growing number of studies have investigated the effectiveness of process-based WM training (i.e., the repetitive practice of tasks assumed to measure WM capacity) and their possible positive impacts on other cognitive abilities, such as reasoning. Aging research has shown that WM and reasoning decline with age (Craik & Bialystok, 2006; Kramer & Willis, 2002; Park et al., 2002), but process-based training interventions focusing particularly on healthy older adults are still scarce. Therefore, the purpose of the present work was to compare the modifiability of WM performance in young and old adults and to examine transfer to nonpracticed WM and reasoning tasks.

To date, growing evidence has shown that WM training can lead to performance increases in nonpracticed WM tasks (see reviews by Klingberg, 2010; Morrison & Chein, 2011). Several studies have demonstrated that such positive effects are also possible in older adults (Buschkuehl et al., 2008; Smith et al., 2009; Zinke, Zeintl, Eschen, Herzog, & Kliegel, 2012), although the observed improvements are often smaller in old than in young adults (Brehmer, Westerberg, & Bäckman, 2012; Dahlin, Stigsdotter Neely, Larsson, Bäckman, & Nyberg, 2008; Dorbath, Hasselhorn & Titz, 2011; Karbach & Kray, 2009; Schmiedek, Lövden, & Lindenberger, 2010; but see Bherer et al., 2005; Li et al., 2008).

Previous findings regarding transfer to reasoning are less consistent. Some studies have established significant effects of WM training on reasoning measures in young (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Jaeggi et al., 2010; Klingberg, Forssberg, & Westerberg, 2002; von Bastian & Oberauer, 2012), and even in old (Basak, Boot, Voss, & Kramer, 2008; Borella, Carretti, Riboldi, & De Beni, 2010; Karbach & Kray, 2009; van Muijden, Band, & Hommel, 2012), adults. However, the results of other studies have been either inconclusive (Schmiedek et al., 2010) or did not

C. C. von Bastian (✉) · N. Langer · L. Jäncke · K. Oberauer
Department of Psychology, University of Zurich,
Binzmühlestrasse 14/22,
CH-8050 Zurich, Switzerland
e-mail: c.vonbastian@psychologie.uzh.ch

support training-induced changes in reasoning (Chein & Morrison, 2010; Chooi & Thompson, 2012; Dahlin, Nyberg, Bäckman, & Stigsdotter Neely, 2008; Owen et al., 2010; Redick et al., in press; Richmond, Morrison, Chein, & Olson, 2011).

The factors contributing to the success of training interventions in terms of transfer are still unclear, and comparisons across studies have been complicated mainly by three methodological issues (Conway & Getz, 2010; Moody, 2009; Shipstead, Redick, & Engle, 2010; von Bastian & Oberauer, 2012). First, prior studies have varied greatly in terms of training conditions. For example, the numbers of training sessions have ranged from only three (Borella et al., 2010) to more than 100 (Schmiedek et al., 2010) across studies, and between two and 188 training sessions within studies (Owen et al., 2010). Second, still only few studies have included active control groups that have completed alternative tasks that were similarly challenging and motivating as those performed by the training group. Evaluating training and transfer effects in comparison to an active control group controls not only for retest effects (as a nonactive or noncontact control group would also do), but also for intervention effects (e.g., effects of keeping to a regular training schedule or of completing regular computer-based tasks that required a high level of concentration) and expectancy effects (Oken et al., 2008). Third, although there is evidence that training is more efficient if the level of task difficulty is adapted to individual performance (Holmes, Gathercole, & Dunning, 2009; Klingberg et al., 2005; Metzler-Baddeley & Baddeley, 2009; Tallal et al., 1996), many previous training regimens for older adults have not included adaptive procedures that adjust task difficulty according to individual performance (e.g., Li et al., 2008; Schmiedek et al., 2010). Therefore, in order to examine WM training and transfer effects across the life span, the present study builds on results recently obtained in young adults with an extensive, well-controlled, and adaptive training regimen (von Bastian & Oberauer, 2012). In this study, each of three groups of participants had training focused on one specific functional category of WM capacity from the facet model of WM capacity (Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000; Oberauer, Süß, Wilhelm, & Wittmann, 2003; Süß et al., 2002). According to this model, WM capacity can be classified into three functional categories: storage and processing, relational integration, and supervision. Storage and processing is the simultaneous maintenance and manipulation of information; relational integration comprises the coordination of information elements into new structures; and supervision[1] is the selective activation of relevant and inhibition of irrelevant information. After four weeks of extensive and adaptive training of one specific functional category, transfer to multiple nonpracticed tasks measuring the construct trained was established by training storage and processing and by training supervision. Both groups also improved in reasoning. Although the group trained in relational integration did not show such broad transfer, we found a strong effect of relational-integration training on a word-position binding task measuring WM.

According to the rationale that transfer of training is driven by overlapping cognitive and neural mechanisms between training and transfer tasks (Buschkuehl, Jaeggi, & Jonides, 2012; Lustig, Shah, Seidler, & Reuter-Lorenz, 2009), even broader transfer effects should emerge for training interventions that target more than one facet of working memory. Specifically, this means that training storage and processing, relational integration, and supervision simultaneously could lead to additive transfer effects (i.e., transfer to nonpracticed WM tasks, to supervision tasks, and to reasoning). Therefore, in the present study, younger and older adults completed an extensive training intervention comprising tasks from all three functional categories, instead of from only a single category. As in the previous study, we included an active control group who practiced tasks with low working memory demand.

Method

Over four weeks, participants had to complete 20 sessions of extensive cognitive training. We randomly assigned participants within each age group (young and old) to one of two training groups: WM training or active control (AC) training. The study was conducted in a double-blinded manner, which means that neither the participants nor the experimenter was aware which groups the participants were assigned to. Training and transfer effects were assessed by administering a broad battery of computer-based tests before and after training. Furthermore, all participants underwent electroencephalographic (EEG) recordings during a subset of the tasks (the three test versions of the WM training tasks and the *n*-back task; see the task descriptions below). Half of the participants additionally participated in functional and structural magnetic resonance imaging (MRI), as well as diffusion-tensor imaging (DTI). These measurements were conducted on another day than the one on which the behavioral assessments and EEG recordings took place. This study focuses on the behavioral findings only; the neuronal correlates will be reported elsewhere (Langer, von Bastian, Oberauer, & Jäncke, in press).

Participants

The participants were recruited for a "cognitive training study" by means of the participant pool at the University

[1] Supervision has so far been operationalized by task-shifting performance, and as such the Supervision factor corresponds to Shifting, one of three lower-level factors in the model of executive functions from the work of Miyake et al. (2000).

of Zurich, flyers distributed at the university's campus, newspaper advertisements, and senior Internet communities. A group of 66 young (43 women, 23 men; M_{age} = 23.27, SD = 3.85, age range 18–35 years) and 57 old (23 women, 34 men; M_{age} = 68.42, SD = 3.28, age range 61–77 years) participants completed the study and received CHF 100 (about US $127) or course credits. Additionally, they had the chance to earn a bonus up to a maximum of CHF 50, depending on the level of difficulty that they achieved during training. All of the participants were German native speakers or were highly proficient in German. The respective age groups did not differ in terms of demographic variables (age, gender, and education; see Table 1). In addition, no group differences emerged for the older participants in a German version of the Geriatric Depression Scale (GDS; Sheikh & Yesavage, 1986). Previous experience with computers and the Internet, and cognitive activity in daily life were assessed via self-constructed questionnaires before the pretest and showed that all of the participants were experienced with using a computer. All older adults participating in the study scored 25 points or more in the Mini-Mental-State Examination (Folstein, Folstein, & McHugh, 1975). All participants gave written consent to participate in the study, which was ethically approved by the Institutional Review Board of the "Kantonale Ethikkommission" (EK: E-80/2008). Six of the participants did not complete the study due to lack of interest (five) or technical problems (one), and six other participants withdrew consent without comment. We excluded four participants who completed fewer than 17 training sessions. Two other participants were excluded due to medical issues (one was diagnosed with Parkinson's disease, and another reached a clinical score on the GDS). The basic demographics of the participants who completed the study are listed in Table 1.

Table 1 Participant demographics

	Group			
	Young		Old	
Demographics	WM	AC	WM	AC
Sample size (n)	34	32	27	30
Gender (f/m)	22/12	21/11	11/16	12/18
Age ($M \pm SD$)	23 ± 4	23 ± 4	68 ± 4	69 ± 3
Education (M rank ± SD)[a]	5 ± 1	5 ± 1	6 ± 2	5 ± 2
GDS score ($M \pm SD$)	–	–	1.35 ± 1.70	1.21 ± 1.24

The training groups did not differ significantly (within age groups) in terms of basic demographics, as determined by two-tailed t tests (or a Mann–Whitney test, in the case of education). WM = working memory training, AC = active control. [a] The scale for education ranged from 1 (*no formal education*) to 8 (*doctorate*)

Design and materials

Training

Each group trained three tasks, each for approximately 10 min during each session. The order of the three tasks was randomized in each session. All participants within the respective groups started the first session at the same level of difficulty. Within and across sessions, task difficulty was adapted stepwise in response to the participants' individual performance (measured as percentages of correctly solved trials; see the Procedure section for details on the adaptive training algorithm). Training effects on the trained tasks were measured via performance gains during training and via test versions of each WM training task presented as pre- and posttests.

WM training

The experimental training comprised one task for each functional category of WM capacity: numerical complex span (storage and processing), Tower of Fame (relational integration), and figural task switching (supervision). The tasks were similar to those used in von Bastian and Oberauer (2012), but were adjusted slightly for the purposes of the present study. First, due to the age-comparative setting, we used an easier-to-understand processing task for numerical complex span (even/odd judgments instead of judgments of the correctness of equations). Second, in response to the participants' feedback after the previous study, we developed a more engaging version of the relational integration task. To this end, we used the names of famous people and descriptions of their neighborhood relations instead of the names of unknown people and descriptions of their kinship relations. Third, in the present study, we used only four instead of five different stimulus sets for task switching (the fifth set from the previous study had been used for the test version of the task; see below).

Numerical complex span Each trial started with a memory item (two-digit numbers) that was displayed centrally in black font for 0.5 s. This was followed immediately by a distractor (number with one digit) that was presented centrally in blue. The participants had to judge the parity (odd or even) of the digit as quickly and accurately as possible. The duration of the distracting task was 3 s. The distractor disappeared after the participant's response, and the remaining time was filled by a blank screen. Afterward, the next memory item followed. After a few memory–decision sequences, participants had to recall the memoranda in the correct serial order. Unlimited time was provided for recall. In each session, the participants completed 12 trials. The number of memory items intermixed with the decision tasks increased with the level of difficulty.

Tower of Fame We developed a task that required the integration of information elements and of the relations between these elements. Participants had to imagine a tower consisting of six floors, each comprising four apartments (A, B, C, and D). Sentences describing the location of a famous person's apartment in this building were presented sequentially. Each sentence was based on the previous one (e.g., "Tom Cruise lives in the second floor in apartment A," "Bruce Willis lives three floors above Tom Cruise, in the apartment to the right"). The participants were then asked to recall the correct apartments of the famous people that had been mentioned in the sentences previously presented (e.g., "Tom Cruise lives in?"—"2A"; "Bruce Willis lives in?"—"5B"). Participants completed 15 trials per session, and the percentage of correct answers served as the score. The level of difficulty was increased by randomizing the order of recall (e.g., "Bruce Willis lives in?" followed by "Tom Cruise lives in?"), and by increasing the number of sentences presented. The randomized order of recall would force participants to memorize not only the apartment numbers (i.e., "2A"), but also the names (i.e., "Tom Cruise"), and thus increase the number of bindings between information elements that would have to be maintained in memory. In each session, the participants completed 15 trials.

Figural task switching Bivalent stimuli (simple geometrical shapes) had to be categorized as accurately and quickly as possible according to rules given in alternating runs of two. The relevant categorization rule and the stimuli were presented simultaneously until participants responded or the display duration was exceeded. To increase the task difficulty, the display duration (i.e., the time to respond to the stimulus) was set to the 99th percentile of the individual reaction times (RTs) in the trials completed since the last adjustment of difficulty (for a more detailed description of this procedure, see von Bastian & Oberauer, 2012). Because this adjustment of task difficulty did not introduce novel stimuli, as was the case for the two other training tasks, variability was enhanced by replacing the sets of stimuli (i.e., new bivalent stimulus and new categorization rules) in every fifth session. Participants completed 384 trials in each session.

Active control training

To hold the variability of the training tasks constant, the active control groups completed three different tasks as well. These tasks were chosen because they required only little WM capacity. In our previous study (von Bastian & Oberauer, 2012), the active control group had practiced visual matching tasks (e.g., face matching). After training, the active control group showed large effects on processing speed, which is an important component of many WM and executive-function tasks (Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007). It is possible that the active control group also improved in their performance on these tasks and, hence, WM training effects were underestimated. For the present study, we therefore chose tasks in which the speed component was minimized.

Quiz General knowledge quiz questions were presented, and participants had to choose one of four alternative answers. The response time was limited to 60 s for each question, and trials without responses were counted as incorrect. The training comprised 3,507 quiz questions provided by the Quiz-Fabrik GmbH (www.quiz-fabrik.de). Participants completed 100 trials in each session, and performance was measured by their percentages of correct answers. The level of training difficulty was increased by presenting more difficult questions; the difficulty of the questions ranged from very easy to very difficult and was rated by the providers of the questions.

Visual search Previous research has shown that prototypical visual search demands only little WM (Kane, Poole, Tuholski, & Engle, 2006; Poole & Kane, 2009; Sobel, Gerrie, Poole, & Kane, 2007; cf. Redick et al., in press). In the visual search task used in the active control group training, several circles with two gaps were displayed simultaneously. The participants had to search the display for the target item, a circle with only one gap, and to indicate the position of this gap by pressing the respective arrow key on the keyboard. Trials could also contain no target item, in which case the participants had to press "A." The display duration was 60 s or until the participant's response. Trials without responses were counted as incorrect; the percentage of correct answers served as the score. Participants completed 70 trials of this task in each session. Higher levels of difficulty corresponded to a greater number of circles displayed simultaneously.

Counting Blocks of identical digits between 1 and 6 were shown on the screen. These blocks comprised as many identical digits in a row as the digit indicated (e.g., five 5 s or three 3 s in a row). If this rule was broken for a digit, the participants were to press the respective number's key on the keyboard (e.g., in "5555," one 5 is missing; therefore, the correct response would be to press the "5" key). In the case that none of the blocks broke the rule, participants had to press the "0" key. Trials were displayed for 60 s or until the participant's response; trials without responses were counted as incorrect. One session comprised 70 trials. The level of difficulty was increased on the basis of the percentage of correct answers, by presenting more blocks of numbers simultaneously.

Pre- and postassessments

Overall, the test battery consisted of ten tasks that were designed to measure training on the three tasks trained, as

well as near transfer to three structurally similar tasks with different materials, intermediate transfer to two structurally dissimilar tasks that still measured the construct trained (i.e., WM), and far transfer to two tasks measuring a different but related construct (i.e., reasoning). Furthermore, we administered a control test to which we did not expect any transfer.

Trained tasks and near-transfer tasks

Each functional category of WM capacity was measured by the three tasks used for training, as well as by three structurally similar tasks that served to assess near transfer.

Storage and processing The complex span tasks consisted of 15 trials with varying list lengths (three to seven memoranda). The numerical version was identical to the training task; the verbal version used words as the memoranda. Memoranda were presented for 1 s, and in between memorization and recall, the participants had to decide whether a letter presented was a consonant or a vowel and to indicate their decision via a keypress. Each decision trial lasted 3 s, showing a blank screen after a participant's response for the remaining time in order to keep the retention time constant. The proportion of items recalled at the correct position was used as the dependent variable (partial-credit unit score; cf. Conway et al., 2005).

Relational integration The test version of the Tower of Fame task comprised 18 trials with the number of sentences (i.e., information elements to be integrated) ranging from two to four. Each sentence was presented for 5 s, and the order of recall was pseudorandomized. Unlimited time was provided to respond. The second task used to measure relational integration was the kinship integration task used in our previous study (von Bastian & Oberauer, 2012). Here, verbal descriptions of the relations between two people (e.g., "Anne is Barney's sister," "Barney is Carol's father") were presented sequentially for 5 s each. After two or three consecutive sentences, participants were asked to indicate the (implied, but not explicitly described) relationship between two people mentioned in the sentences previously presented (e.g., "Anne is Carol's?", with the correct answer being "aunt"). The test comprised 16 trials, and the proportion of correct answers was the outcome measure.

Supervision The task-switching tests comprised 80 bivalent stimuli each. The test version of figural task switching included stimuli similar to those in the training version (i.e., geometrical shapes), but the task set (i.e., the categorization rules) differed from those used during training. Participants had to decide either whether the stimulus shown was green or blue, or whether it was round or angular. In the verbal version, we presented words that had to be categorized as being either cities or rivers, or as being written in either green or blue. As in the training, the categorization rules switched after every second stimulus. A cue for the relevant task was shown simultaneously with the stimulus. The dependent variable measured was proportional switch costs, which were calculated by subtracting RTs in task switch trials from RTs in task repetition trials, and dividing the difference by the average RT (including both switch and repetition trials) per individual.

Intermediate transfer (WM)

A word-position binding task and an *n*-back task were used to assess transfer to structurally different WM tasks.

Binding In this task, two to five words were presented sequentially for 2 s each in different positions on the screen (cf. Oberauer, 2005). Participants had to memorize which word was shown at which position. Immediately afterward, probe words were displayed at the different positions. Positive probes were words from the previous list shown at the correct position, whereas negative probes were words shown at a different position than during learning. Across all 32 trials, the probes were 50 % positive and 50 % negative. The positive probes were distributed equally (± 1) across the serial positions, defined by the temporal order of presentation, and across the possible positions on the screen. Performance was measured by the discrimination parameter d' from signal detection theory, which takes hits and false alarms into account. It is calculated as $d' = z(\mathrm{FA}) - z(\mathrm{H})$, where H is the hit rate, FA the false alarm rate, and z refers to the z value corresponding to the probability of the given argument.

N-back Letters were presented sequentially, and participants had to decide whether the letter currently shown was the same as the one at n positions back, independent of whether or not the letter was displayed in capitals (e.g., as "A" or "a"). To increase recall based on recollection rather than familiarity (cf. Szmalec, Verbruggen, Vandierendonck, & Kemps, 2011), high-interference distractors were implemented (i.e., target letters that were shown at the wrong positions $n + 1$ and $n - 1$). The stimuli were presented for 500 ms each, followed by a 2,500-ms interstimulus interval. Participants had to respond to every item and could indicate their responses by keypresses during the whole trial (i.e., for 3,000 ms). Participants completed each level of n (2 to 4) for three consecutive blocks of trials, with each block consisting of $20 + n$ trials. Each block contained six matching letters and three high-interference distractors, with the remaining trials being mismatches. The proportion of correct answers was used as the dependent measure.

Far transfer (reasoning)

Far transfer to a different construct was measured by Raven's Advanced Progressive Matrices (RAPM; Raven, 1990). In this task, participants have to select the one of eight figures that completes a pattern presented. The 36 items of the RAPM were divided into odd and even items in order to create two test versions for the pre- and posttest assessments. The RAPM task was administered without a time limit. Previous studies examining transfer effects in young adults had occasionally reported trends toward ceiling effects (e.g., Jaeggi et al., 2008), and therefore we administered the Bochumer Matrizentest (BOMAT; Hossiep, Turck, & Hasella, 2001) to the young sample. The BOMAT is a matrix reasoning test similar to the RAPM, but more difficult. In the BOMAT, participants have to select one of six alternative figures to complete the patterns presented, and the test comprises 29 trials. We used the published parallel test versions A and B for the pre- and posttest assessments. The BOMAT was administered with a fixed time limit of 45 min, as determined by the manual.

Control test

A quiz on general knowledge served as a control test to which we did not expect any transfer of WM training. In addition, the quiz being part of pre- and postassessments increased the believability of the control training, because participants in the control group (like those in the experimental group) experienced a test similar to their training tasks. The questions in this test version differed from those used during the control group training, and therefore we did not expect any improvements from the control group in this task, either. The test comprised 16 open text questions.

Procedure

Training

All of the participants had to complete 20 sessions of intensive training (approximately 25–30 min per session). Training was self-administered at home via the open-source software Tatool (von Bastian, Locher, & Ruflin, in press). After each training session, participants automatically uploaded their data to a Web server running Tatool Online, which permitted us to constantly control the participants' compliance. To enhance experimental control as much as possible, we took several steps, such as maximizing individual commitment by signing a participant agreement, alerting participants that their training data would be monitored, and automated online analysis of the training data in order to detect irregularities (e.g., accuracies below chance level). Furthermore, we stayed in regular contact with the participants via e-mail and phone. After half of the training sessions had been completed, each participant received an e-mail asking how the training had gone so far. In addition, participants could always contact the experimenters in case of any technical difficulties.

To adapt the level of task difficulty to individual performance, we used the adaptive score and level handler included in Tatool (see Fig. 1). This algorithm measured individual performance at intervals that represented 40 % of the trials of one session in each task (counted across sessions). For example, in the complex span task, 40 % of the trials corresponded to five trials. If the participant scored at least 80 % correct, the algorithm set the performance as the individual benchmark. If the participant's performance improved after another 40 % of the trials (e.g., the performance in the next five trials was greater than the individual benchmark), task difficulty was increased, and the algorithm recalculated the individual benchmark after the next 40 % of the trials. However, if performance was lower than the benchmark, the algorithm repeatedly checked the performance after every 40 % of the trials. If performance did not improve after three such unsuccessful retries, the level of task difficulty was decreased. Participants were informed about changes in the level of difficulty (e.g., "Congratulations, you achieved the next level"), and they started each session on the level that they had achieved in the previous session.

Pre- and postassessments

Participants were tested in groups of no more than five. To control for the effects of fatigue, half of the participants of

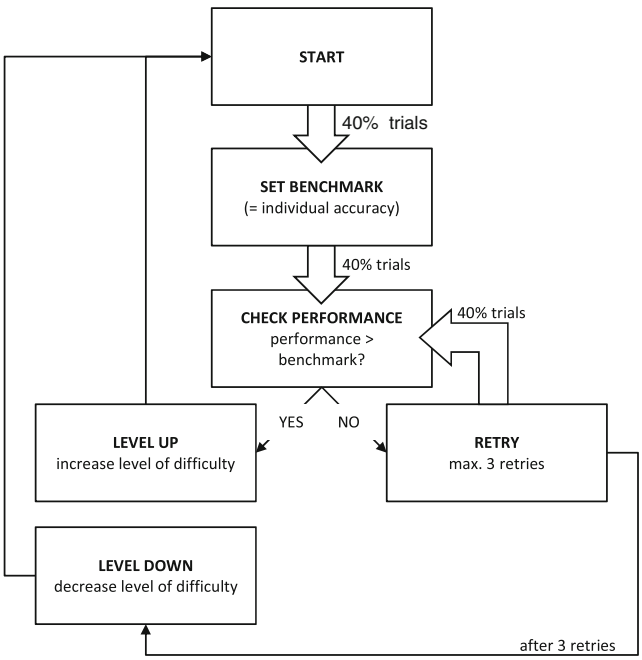

Fig. 1 Algorithm that adapts the level of task difficulty to individual changes in performance

each group completed the transfer tests in reverse order, relative to the other half. To minimize retest effects, different sets of stimuli (A and B) were used for the two occasions and were balanced with respect to groups and the order of test administration. For the computerized tests, we used Dell Optiplex GX620 PCs running Windows XP. The tasks were written in Tatool (von Bastian et al., in press). Stimuli were presented on a 17-in. TFT monitor, and manual responses were registered by a standard computer keyboard and a standard mouse.

Results

Missing data

Due to technical difficulties during the pretest assessment, we lost the data of one participant in the binding task. This participant was excluded from analyses that included this task. Two of the participants completed only 17 training sessions, one only 18 sessions, and six only 19 sessions, due to scheduling problems. Another four participants completed 21 training sessions. The results were the same, independent of whether or not the participants who completed more or less than 20 sessions were excluded; therefore, we included all of the participants in our analyses to maximize power.

Treatment of RT data

Task-switching scores (proportional switch costs) were based on the RTs of correct responses only. RTs of the responses immediately after wrong responses and RT outliers were excluded from the analysis. Outliers were defined as RTs exceeding a participant's mean by more than 3 *SD*s. On average, this led to 11 % of RTs being eliminated.

Analysis

First, to ensure that the effects that we found could be interpreted as being induced by training rather than baseline differences, we conducted two-tailed *t* tests for each transfer task in the pretest separately for both age groups. There were no significant baseline differences for any measurement (all *p*s > .184). However, there was a tendency for participants in the old control group to score worse in the RAPM than did participants in the old experimental group [$t(55) = 1.81$, $p = .076$]. Table 2 lists the means and standard deviations for each group in each task.

Training effects

Individual data inspection showed no signs of low engagement for any of the participants included (e.g., responding

Table 2 Mean performance on the test battery tasks as a function of training group and time of assessment

	Young				Old			
	WM		AC		WM		AC	
Task	T1	T2	T1	T2	T1	T2	T1	T2
Training Tasks								
Numerical complex span	.49 (.11)	.70 (.16)	.45 (.11)	.53 (.11)	.32 (.13)	.53 (.17)	.26 (.12)	.38 (.15)
Tower of Fame	.22 (.08)	.32 (.09)	.20 (.07)	.24 (.08)	.10 (.05)	.18 (.05)	.11 (.06)	.15 (.06)
Figural task switching	.10 (.12)	.11 (.11)	.12 (.11)	.10 (.08)	.03 (.22)	.06 (.07)	.08 (.11)	.04 (.10)
Near Transfer								
Verbal complex span	.74 (.16)	.84 (.14)	.77 (.09)	.79 (.10)	.51 (.11)	.64 (.09)	.53 (.11)	.57 (.16)
Kinship integration	.73 (.21)	.79 (.19)	.68 (.18)	.75 (.19)	.35 (.16)	.39 (.18)	.36 (.21)	.40 (.21)
Verbal task switching	.07 (.13)	.09 (.10)	.06 (.12)	.07 (.10)	.04 (.12)	.08 (.12)	.04 (.12)	.09 (.10)
Intermediate Transfer								
Binding	2.45 (0.77)	2.82 (0.73)	2.52 (0.65)	2.64 (0.62)	1.71 (0.58)	1.89 (0.59)	1.49 (0.60)	1.94 (0.64)
N-back	.81 (.06)	.86 (.07)	.78 (.12)	.84 (.13)	.56 (.25)	.57 (.26)	.54 (.23)	.60 (.23)
Far Transfer								
RAPM	13.44 (2.80)	13.85 (2.43)	13.69 (3.12)	14.59 (2.50)	9.04 (3.46)	8.74 (3.32)	7.30 (3.76)	8.40 (3.61)
BOMAT	16.91 (4.64)	19.18 (4.78)	15.53 (4.64)	18.59 (4.68)	–	–	–	–
Control Test								
Quiz	.46 (.18)	.45 (.17)	.44 (.14)	.41 (13)	.35 (.12)	.45 (.17)	.34 (.14)	.36 (.14)

Standard deviations are given in parentheses. All values are given as accuracy proportions, except for task switching (proportional switch costs), binding (d'), and RAPM and BOMAT (numbers of correctly solved matrices). Only young participants completed the BOMAT. T1 = pretest, T2 = posttest. WM = working memory training, AC = active control

repeatedly with the same key or irregular RTs). Training effects were analyzed for each group and training task with analyses of variance (ANOVAs) for repeated measures, using training performance as the dependent variable, and age group and training session as independent variables. Training session was coded by a linear contrast to reflect monotonic trends rather than erratic fluctuations across sessions. As is illustrated in Fig. 2, all groups showed large training effects for each training task, indicated by significant linear effects of session (all *p*s < .001; see Table 3), except for figural task switching, for which the linear contrast was not significant in either age group. The main effect of age was significant for numerical complex span [$F(1, 52) = 20.24$, $p < .001$, $\eta_p^2 = .28$], reflecting that younger participants performed better than older participants. Furthermore, we found a significant interaction of age with the linear contrast of session [$F(1, 52) = 19.13$, $p < .001$, $\eta_p^2 = .27$], indicating larger improvements in young than in old participants. The same pattern was observed for the Tower of Fame task [age: $F(1, 52) = 31.96$, $p < .001$, $\eta_p^2 = .38$; Session × Age: $F(1, 52) = 17.44$, $p < .001$, $\eta_p^2 = .25$]. For task switching, an effect of age also emerged [$F(1, 52) = 4.33$, $p = .042$, $\eta_p^2 = .08$], but in this case, older participants performed better than younger participants (i.e., they showed smaller proportional switch costs). The linear contrast of the Session × Age interaction was not significant [$F(1, 52) < 0.01$, $p = .996$, $\eta_p^2 < .01$]. In the active control group, older participants performed better than younger participants in the quiz, $F(1, 58) = 23.98$, $p < .001$, $\eta_p^2 = .97$, and also showed larger gains during training, as reflected by a significant interaction of age with the linear contrast of session [$F(1, 58) = 10.05$, $p = .002$, $\eta_p^2 = .15$]. We found neither a main effect of age nor a Session × Age interaction for either visual search [$F(1, 58) = 0.32$, $p = .574$, $\eta_p^2 = .28$, and $F(1, 58) = 0.01$, $p = .931$, $\eta_p^2 < .01$, respectively] or counting [$F(1, 58) = 0.05$, $p = .823$, $\eta_p^2 < .01$, and $F(1, 58) = 0.04$, $p = .839$, $\eta_p^2 < .01$, respectively].

One general problem occurs when analyzing training gains on the basis of performance during training: All participants start the training phase on the same level of difficulty, independent of individual initial ability. Thus, people with higher initial ability will reach higher levels faster, even in the absence of training gains. As a consequence, performance gain during training is a measure that confounds initial ability and improvements in ability above this initial level. Therefore, we measured training gain also with test versions of the WM training tasks from the pre- and post-assessments. These tasks were structurally identical to the training versions, except for the absence of feedback during testing. A mixed-design ANOVA with age group, training group, and assessment (pre- vs. posttest) as independent variables showed that WM training induced greater performance gains from pretest to posttest, as compared to active control training, in the numerical complex span task, $F(1, 119) = 22.38$, $p < .001$, $\eta_p^2 = .16$, and the Tower of Fame task, $F(1, 119) = 23.44$, $p < .001$, $\eta_p^2 = .17$, but not for task switching, $F(1, 119) = 2.85$, $p = .094$, $\eta_p^2 = .02$ (cf. Table A1 in the Appendix). This confirms the effects found during training. Unlike the scores during training, however,

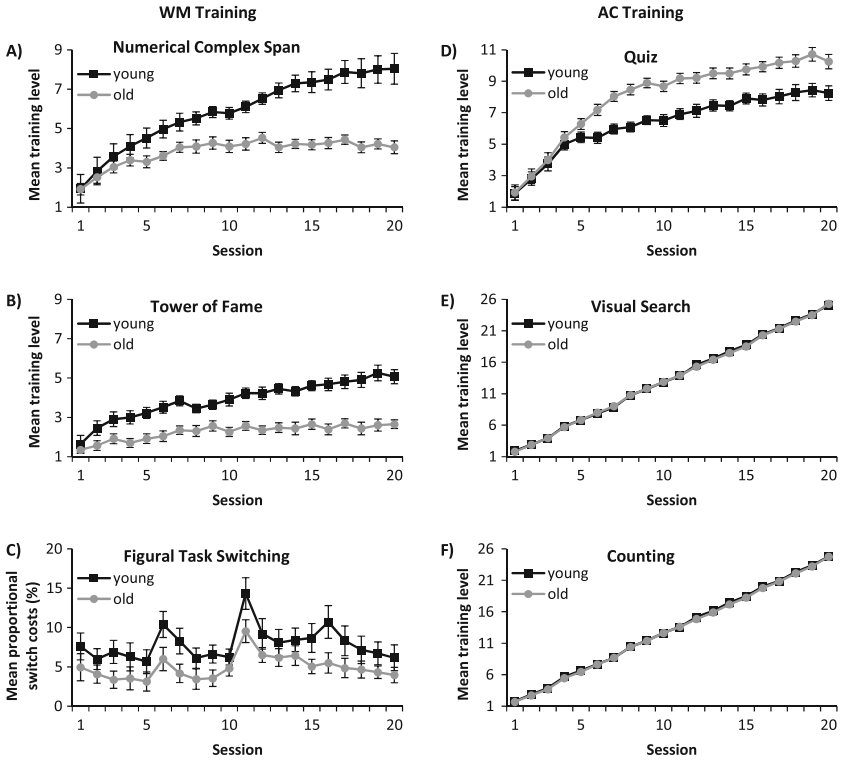

Fig. 2 Training gains during working memory (WM; panels **a–c**) and active control (**ac**; panels **d–f**) training. Error bars represent confidence intervals (95 %) for the within-subjects comparisons, calculated according to Cousineau (2005) and Morey (2008)

Table 3 Linear contrasts of training effects on performance in the trained tasks during training

Group/Age	Task	*M*	*SD*	*F*	*p*	η_p^2
WM						
Young	Numerical complex span	8.03	4.05	52.74	**<.001**	.64
	Tower of Fame	5.06	2.13	65.28	**<.001**	.69
	Figural task switching	0.01	0.05	0.78	.384	.03
Old	Numerical complex span	4.04	1.02	151.37	**<.001**	.87
	Tower of Fame	2.65	0.57	53.65	**<.001**	.71
	Figural task switching	0.004	0.03	3.26	.085	.13
AC						
Young	Quiz	8.25	2.26	182.12	**<.001**	.86
	Visual search	25.09	1.49	10,767.54	**<.001**	.99
	Counting	24.78	1.60	8,564.46	**<.001**	.99
Old	Quiz	10.25	1.99	295.65	**<.001**	.92
	Visual search	25.29	2.19	1,917.06	**<.001**	.99
	Counting	24.75	2.03	1,967.92	**<.001**	.99

Bold *p* values indicate significant effects. The means and standard deviations are given for the last training session. For figural task switching, mean improvements in switch costs from Sessions 1 to 20 and the respective standard deviations are given. Only participants with complete training data sets were included in the analyses. The dependent variable was the level of difficulty achieved in the last training session, except for figural task switching, for which proportional improvement in switch costs from the first to the last training session was used. The *df*s for the *F* statistics are as follows: WM training (young), 1, 30; WM training (old), 1, 22; AC training (young), 1, 31; AC training (old), 1, 27. WM = working memory training, AC = active control

performance gains in the test versions were not significantly modulated by age, as reflected in the Assessment × Age × Training Group interactions (*F*s < 1). Therefore, the age modulation during training was probably due to the lower initial performance of older than of younger participants.

Transfer effects

Transfer effects were assessed with mixed-design ANOVAs for each task, with Age (young vs. old) and Training Group (WM vs. active control) as between-subjects factors and Assessment (pre- vs. posttest) as a within-subjects factor. The complete results are listed in Table A1 in the Appendix. Significant Training Group × Assessment interactions, in combination with larger means for the WM training group (Table 2), provide evidence for positive effects of WM training on the respective measures; significant Age × Group × Assessment interactions indicate that these effects were modulated by age. Effect sizes (see Fig. 3 and Table A2 in the Appendix) were standardized by the standard deviation at pretest within each age group.

Near transfer The only significant transfer was observed for the verbal complex span task, which was structurally similar to the numerical complex span task used for training, $F(1, 119) = 13.49$, $p < .001$, $\eta_p^2 = .10$. Again, this effect was not modulated by age ($F < 1$). Transfer effects emerged for neither kinship integration nor verbal task switching (*F*s < 1).

Intermediate transfer denotes transfer to structurally dissimilar tasks that measure the same theoretical construct that had been trained. We found no significant interactions indicating transfer (*F*s < 1), except for an Age × Group × Assessment interaction for performance on the binding task, $F(1, 118) = 8.06$, $p = .005$, $\eta_p^2 = .06$. We conducted post-hoc ANOVAs for each age group separately to identify the source of this interaction. The ANOVA for young adults showed a marginal Group × Assessment interaction, indicating that WM training might have led to superior performance in binding as compared to active control training, $F(1, 64) = 3.75$, $p = .057$, $\eta_p^2 = .06$. The Group × Assessment interaction was also significant for older adults, $F(1, 54) =$

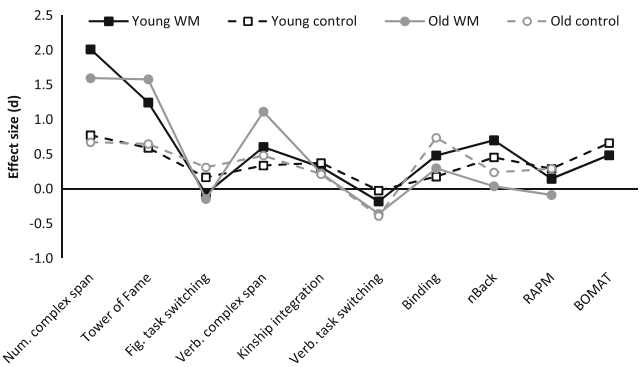

Fig. 3 Effects of working memory (WM) and active control training on the tasks included in the test battery. Only young participants completed the BOMAT

4.67, $p = .035$, $\eta_p^2 = .08$. In this case, however, the performance gain was larger in the active control group. Examining the means in Table 2 shows that the older active control group performed slightly (but not significantly) worse than the training group at pretest [$M_{diff} = 0.22$; $t(54) = 1.35$, $p = .184$], whereas the mean difference between the groups was rather small and also not significant at posttest [$M_{diff} = -0.05$; $t(54) = -0.27$, $p = .787$]. The older WM group's performance increased slightly, but not significantly [$t(25) = -1.93$, $p = .065$], whereas the older control group improved significantly from pre- to posttest [$t(29) = -6.16$, $p < .001$]. There is no obvious explanation for this effect.

Far transfer For the RAPM, a significant Group × Assessment interaction emerged, $F(1, 119) = 4.01$, $p = .047$, $\eta_p^2 = .03$. The means in Table 2, however, reveal that the source of this interaction was probably a larger performance gain in the active control group than in the WM training group. To investigate this supposition, we conducted post-hoc *t* tests between groups (WM training vs. control, with the two age groups conjoined) for each test assessment, and within groups (pre- vs. posttest). As with the binding task, the results showed that the groups differed neither at pretest, $t(121) = 1.17$, $p = .247$, nor at posttest, $t(121) = -0.01$, $p = .993$. The WM training group's change in RAPM performance from pre- to posttest was not significant, $t(60) = 0.30$, $p = .247$, but the change was significant for the control group, $t(61) = 3.03$, $p = .004$. We found no effect of WM training on performance in the BOMAT ($F < 1$).

Control test

There was no effect of training condition on changes in performance in the open-format quiz ($F < 1$).

Discussion

The present work had two goals. First, we examined whether transfer effects induced by WM training occur not only in younger, but also in older, adults. Second, on the basis of the rationale that transfer is driven by functional overlap between training and transfer tasks, we investigated the hypothesis that transfer should be broader if the training regimen targets multiple cognitive functions instead of focusing only on one specific process (Buschkuehl et al., 2012; Lustig et al., 2009). Therefore, our WM training regimen addressed the three functional categories in the facet model of WM capacity (Oberauer et al., 2000; Oberauer et al., 2003; Süß et al., 2002) simultaneously: storage and processing, relational integration, and supervision. In each age group, we compared a WM training group to an active control group that practiced tasks with only low WM demand.

Although we found large training effects for two of the training tasks (numerical complex span and Tower of Fame), no effects of training on proportional task-switching costs were apparent. The absence of a training effect on switch costs stands in contrast to previous findings (Karbach & Kray, 2009; von Bastian & Oberauer, 2012). An obvious difference between our previous study (von Bastian & Oberauer, 2012) and the present one is that participants in the supervision training group of the previous study completed more trials in each training session, simply because the four weeks of training focused on task switching only. In the study conducted by Karbach and Kray, however, a very short training intervention (only four sessions) led to improvements in switch costs. The difference between the switching paradigm used by Karbach and Kray and ours is that we used cues to indicate the relevant task, and they did not. It is possible that task-switching training is more effective when tasks are not cued, so that participants must keep track of the task sequence themselves. This speculation is in line with Minear and Shah (2008), who also used cued task-switching training and did not observe training-related decreases in switch costs, either. Future studies that compare training effects in cued and noncued task switching could shed further light on this matter.

A small transfer effect was observed in the verbal version of the complex span task for both age groups. In addition, we found a marginally significant effect for binding in young, but not in old, adults. The benefit of WM training on binding in young adults replicates an equivalent finding in our previous study (von Bastian & Oberauer, 2012). The facts that we found transfer to the verbal complex span task and obtained weak evidence for transfer to binding, but not to tasks more similar to the other training tasks (i.e., kinship integration, verbal task switching), suggests that probably only numerical complex span training was successful in terms of inducing transfer. Concerning transfer to reasoning, we did not find any evidence for WM-training-induced improvements. The absence of transfer to reasoning is not surprising, given the assumption that far transfer effects are generally smaller than near transfer effects (e.g., Klauer, 2001), and given that only small near transfer effects were found for the verbal complex span and binding tasks. The effects on the trained tasks and the verbal version of the complex span task were of the same magnitude for young and old adults, similar to the findings in some previous training studies with age-comparative settings (Bherer et al., 2005; Davidson, Zacks, & Williams, 2003; Li et al., 2008). The only effect that was modulated by age was the effect of binding, which was absent in old adults, and weak in young adults. The absence of transfer to binding exclusively in old adults matches the hypothesis of age-related impairments in associative memory (Oberauer, 2005; Old & Naveh-Benjamin, 2008). However, the effect was also only

small in young adults; hence, this age modulation should be interpreted cautiously.

Regarding our second research question, this study has provided evidence that transfer is not broader, but weaker when multiple functional categories are trained at once, as compared to a setting in which the training intervention focuses on only one specific functional category (von Bastian & Oberauer, 2012). Of course, because we kept the overall training intensity (20 sessions within four weeks) constant across both studies, each functional category was trained less intensively in the present training regimen than in the previous study. It is possible that in order to produce transfer, each functional category has to be trained with a certain minimum level of intensity, which lies beyond the amount of training that participants received in the present study. Therefore, we cannot exclude the possibility that training over a longer period (e.g., three times longer, to keep the training times for each functional category constant) would lead to broader transfer. For example, the higher training intensity in the study by Schmiedek et al. (2010) could be the reason why they found more transfer than the present study, although they used an equally broad training method. To investigate this matter, future studies will be required that comprise different intensity conditions and that directly contrast single-function with multifunction training regimens.

For all tasks in our test battery, we found main effects of age, indicating better performance in young than in older adults (see Table A1 in the Appendix), except for task switching, where the effect was either in the opposite direction (figural version) or absent (verbal version). At first glance, this finding seems to indicate the rather counterintuitive conclusion that executive functioning was better in older than in young adults. Absent or reverse age effects on switch costs, however, are a common finding in the literature (Reimers & Maylor, 2005; Wasylyshyn, Verhaeghen, & Sliwinski, 2011; Whitson, Karayanidis, & Michie, 2012). One possible reason for this was suggested by Mayr (2001). Theoretically, it is assumed that the relevant task set has to be selected only after a task switch. Smaller switch costs are, therefore, often interpreted as revealing increased efficiency in task set selection in switch trials. The smaller switch costs in older adults could, however, also reflect that older adults have to rely on task selection not only in switch trials, but also in repetition trials. This assumption is supported by our finding that the age effect found in all other tasks in our test battery favored younger adults.

To conclude, our study provides evidence that WM training targeting multiple functional categories is less efficient than WM training focusing on single processes only. Given that transfer effects of WM training are generally rather small (if they are observed at all), our results suggest that future training interventions (at least those that extend only four weeks) should better focus on specific functional categories in order to enhance the probability of observing transfer. The magnitude of training effects was not modulated by age; transfer was, however, very narrow for both age groups.

Author note This research is part of the first author's doctoral thesis and was supported by grants to the first author from the Forschungskredit of the University of Zurich and the Suzanne and Hans Biäsch Foundation for Applied Psychology. We thank André Locher for programming the training tasks in Tatool, and Sabrina Guye, Veronica Heusser, Melanie Künzli, Eszter Montvai, Katharina Vogt, Helen Wirz, and Marc Züst for their assistance with collecting the data.

Appendix

Table 4 ANOVA effects of working memory training on the test battery measurements

Task	$F(1, 119)$	p	η_p^2
Training Effects			
Numerical complex span			
Age	64.01	**<.001**	.35
Group	23.27	**<.001**	.16
Age × Group	<0.01	.954	<.01
Assessment	178.06	**<.001**	.60
Age × Assessment	0.26	.612	<.01
Group × Assessment	22.38	**<.001**	.16
Age × Group × Assessment	0.74	.391	<.01
Tower of Fame			
Age	81.46	**<.001**	.90
Group	6.57	**.012**	.05
Age × Group	2.33	.129	.02
Assessment	153.86	**<.001**	.56
Age × Assessment	1.34	.249	.01
Group × Assessment	23.44	**<.001**	.17
Age × Group × Assessment	0.89	.347	.01
Task switching figural			
Age	10.74	**.001**	.08
Group	0.30	.587	<.01
Age × Group	0.20	.656	<.01
Assessment	0.06	.801	<.01
Age × Assessment	0.11	.918	<.01
Group × Assessment	2.85	.094	.02
Age × Group × Assessment	0.58	.448	.01
Near Transfer			
Complex span verbal			
Age	137.95	**<.001**	.54
Group	0.76	.385	.01
Age × Group	0.11	.742	<.01

Table 4 (continued)

Task	$F(1, 119)$	p	η_p^2
Assessment	63.94	**<.001**	.35
Age × Assessment	1.81	.182	.02
Group × Assessment	13.49	**<.001**	.10
Age × Group × Assessment	<0.01	.948	<.01
Kinship integration			
Age	124.96	**<.001**	.91
Group	0.27	.602	<.01
Age × Group	0.80	.372	.01
Assessment	18.49	**<.001**	.13
Age × Assessment	1.09	.299	.01
Group × Assessment	0.02	.889	<.01
Age × Group × Assessment	0.02	.896	<.01
Task switching verbal			
Age	0.31	.579	<.01
Group	0.13	.724	<.01
Age × Group	0.17	.685	<.01
Assessment	5.90	**.017**	.047
Age × Assessment	1.70	.194	.014
Group × Assessment	0.13	.717	<.01
Age × Group × Assessment	0.28	.596	<.01
Intermediate Transfer			
Binding[a]			
Age	58.54	**<.001**	.93
Group	0.39	.536	<.01
Age × Group	0.02	.890	<.01
Assessment	38.50	**<.001**	.25
Age × Assessment	0.63	.429	.01
Group × Assessment	<0.01	.993	<.01
Age × Group × Assessment	8.06	**.005**	.06
N-back			
Age	66.31	**<.001**	.36
Group	0.10	.759	<.01
Age × Group	0.17	.684	<.01
Assessment	17.37	**<.001**	.13
Age × Assessment	0.80	.373	.01
Group × Assessment	1.82	.179	.02
Age × Group × Assessment	0.87	.353	.01
Far Transfer			
RAPM			
Age	114.87	**<.001**	.49
Group	0.28	.598	.01
Age × Group	2.21	.140	<.01
Assessment	5.05	**.026**	.04
Age × Assessment	0.30	.587	<.01
Group × Assessment	4.01	**.047**	.03
Age × Group × Assessment	0.91	.341	.01
BOMAT[b]			
Group	0.86	.357	.01
Assessment	33.47	**<.001**	.34
Group × Assessment	0.75	.389	.01
Control Task			
Quiz			
Age	9.57	**.002**	.07
Group	1.48	.227	.01
Age × Group	0.01	.929	<.01
Assessment	0.46	.498	<.01
Age × Assessment	5.65	**.019**	.05
Group × Assessment	0.89	.348	.01
Age × Group × Assessment	0.19	.666	<.01

Bold p values indicate significant effects ($p < .05$, two-tailed). [a] $df = 1, 118$ (missing data of one old participant in the experimental group at pretest). [b] $df = 1, 64$ (only young participants completed the BOMAT)

Table 5 Effect sizes (Cohen's *d*) of training on the test battery measurements for the working memory (WM) and active control (AC) groups

Task	Young			Old		
	WM	AC	Net ES	WM	AC	Net ES
Training Effects						
Numerical complex span	2.01	0.77	1.20	1.60	0.60	0.58
Tower of Fame	1.25	0.59	0.80	1.58	0.65	0.74
Figural task switching	−0.07	0.17	−0.23	−0.14	0.31	−0.39
Near Transfer						
Verbal complex span	0.60	0.34	0.53	1.11	0.48	0.64
Kinship integration	0.31	0.38	0.00	0.23	0.21	−0.04
Verbal task switching	−0.18	−0.02	−0.17	−0.36	−0.39	0.03
Intermediate Transfer						
Binding	0.48	0.18	0.36	0.30	0.74	−0.46
N-back	0.70	0.45	−0.09	0.04	0.24	−0.19
Far Transfer						
RAPM	0.15	0.29	−0.17	−0.09	0.29	−0.38
BOMAT	0.49	0.66	−0.17	–	–	–
Control Task						
Quiz	−0.08	−0.19	0.08	0.46	0.15	0.27

Net effect sizes (Net ES) were calculated by subtracting the control group's gain scores (from pre- to posttest) from the WM training group's gain scores and standardizing the result by the standard deviation at pretest within each age group

References

Basak, C., Boot, W. R., Voss, M. W., & Kramer, A. F. (2008). Can training in a real-time strategy video game attenuate cognitive decline in older adults? *Psychology and Aging, 23,* 765–777. doi:10.1037/a0013494

Bherer, L., Kramer, A. F., Peterson, M. S., Colcombe, S., Erickson, K., & Becic, E. (2005). Training effects on dual-task performance: Are there age-related differences in plasticity of attentional control? *Psychology and Aging, 20,* 695–709.

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

Brehmer, Y., Westerberg, H., & Bäckman, L. (2012). Working-memory training in younger and older audlts: Training gains, transfer, and maintenance. *Frontiers in Human Neuroscience, 6*(63), 1–7. doi:10.3389/fnhum.2012.00063

Buschkuehl, M., Jaeggi, S. M., Hutchison, S., Perrig-Chiello, P., Däpp, C., Müller, M., & Perrig, W. J. (2008). Impact of working memory training on old-old adults. *Psychology and Aging, 23,* 743–753.

Buschkuehl, M., Jaeggi, S. M., & Jonides, J. (2012). Neuronal effects following working memory training. *Developmental Cognitive Neuroscience, 25,* S167–S179. doi:10.1016/j.dcn.2011.10.001

Chein, J. M., & Morrison, A. B. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. *Psychonomic Bulletin & Review, 17,* 193–199. doi:10.3758/PBR.17.2.193

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

Conway, A. R. A., & Getz, S. J. (2010). Cognitive ability: Does working memory training enhance intelligence? *Current Biology, 20,* R362–R364.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12,* 739–786. doi:10.3758/BF03196772

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7,* 547–552. doi:10.1016/j.tics.2003.10.005

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1,* 42–45.

Craik, F. I. M., & Bialystok, E. (2006). Cognition through the lifespan: Mechanisms of change. *Trends in Cognitive Sciences, 10,* 131–138. doi:10.1016/j.tics.2006.01.007

Dahlin, E., Nyberg, L., Bäckman, L., & Stigsdotter Neely, A. (2008). Plasticity of executive functioning in young and older adults: Immediate training gains, transfer, and long-term maintenance. *Psychology and Aging, 23,* 720–730. doi:10.1037/a0014296

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

Davidson, D. J., Zacks, R. T., & Williams, C. C. (2003). Stroop interference, practice, and aging. *Aging Neuropsychology, and Cognition, 10,* 85–98.

Dorbath, L., Hasselhorn, M., & Titz, C. (2011). Aging and executive functioning: A training study on focus-switching. *Frontiers in Psychology, 2*(257), 1–12. doi:10.3389/fpsyg.2011.00257

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). Cambridge, U.K.: Cambridge University Press.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198. doi:10.1016/0022-3956(75)90026-6

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

Hossiep, R., Turck, D., & Hasella, M. (2001). *BOMAT, advanced short version*. Göttingen, Germany: Hogrefe.

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

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,* 625–635. doi:10.1016/j.intell.2010.09.001

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

Klauer, K. J. (Ed.). (2001). *Handbuch Kognitives Training [Handbook of cognitive training]* (2nd ed.). Göttingen, Germany: Hogrefe.

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., & Westerberg, H. (2005). Computerized training of working memory in children with ADHD—A randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry, 44,* 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

Kramer, A. F., & Willis, S. L. (2002). Enhancing the cognitive vitality of older adults. *Current Directions in Psychological Science, 11,* 173–177.

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

Langer, N., von Bastian, C. C., Oberauer, K., & Jäncke, L. (in press). *The effects of working memory training on functional brain network efficiency*. Manuscript in preparation. doi:10.1037/a0029082

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

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

Mayr, U. (2001). Age differences in the selection of mental sets: The role of inhibition, stimulus ambiguity, and response-set overlap. *Psychology and Aging, 16,* 96–109. doi:10.1037//0882-7974.16.1.96

Metzler-Baddeley, C., & Baddeley, R. J. (2009). Does adaptive training work? *Applied Cognitive Psychology, 23,* 254–266. doi:10.1002/acp. 1454

Minear, M., & Shah, P. (2008). Training and transfer effects in task switching. *Memory & Cognition, 36,* 1470–1483. doi:10.3758/MC.336.8.1470

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

Moody, D. E. (2009). Can intelligence be increased by training on a task of working memory? *Intelligence, 37,* 327–328.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4,* 61–64.

Morrison, A. B., & Chein, J. M. (2011). Does working memory traning 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

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., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—Facets of a cognitive ability construct. *Personality and Individual Differences, 29,* 1017–1045.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittmann, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence, 31,* 167–193.

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

Oken, B. S., Flegal, K., Zajdel, D., Kishiyama, S., Haas, M., & Peters, D. (2008). Expectancy effect: Impact of pill administration on cognitive performance in healthy seniors. *Journal of Clinical and Experimental Neuropsychology, 30,* 7–17. doi:10.1080/13803390701775428

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging, 23,* 104–118. doi:10.1037/0882-7974.23.1.104

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

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging, 17,* 299–320. doi:10.1037/0882-7974.17.2.299

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

Raven, J. C. (1990). *Advanced Progressive Matrices: Sets I, II*. Oxford, U.K.: Oxford Psychologists Press.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., . . . Engle, R. W. (in press). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General.* doi:10.1037/a0029082

Reimers, S., & Maylor, E. A. (2005). Task switching across the life span: Effects of age on general and specific switch costs. *Developmental Psychology, 41,* 661–671. doi:10.1037/0012-1649.41.4.661

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

Schmiedek, F., Lövden, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: Findings from the COGITO study. *Frontiers in Aging Neuroscience, 2*(27), 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,* 414–429. doi:10.1037/0096-3445.136.3.414

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist, 5,* 165–173. doi:10.1300/J018v05n01_09

Shipstead, Z., Redick, T. S., & Engle, R. W. (2010). Does working memory training generalize? *Psychologica Belgica, 50,* 245–276.

Smith, G. E., Housen, P., Yaffe, K., Ruff, R., Kennison, R. F., Mahncke, H. W., & Zelinski, E. M. (2009). A cognitive training program based on principles of brain plasticity: Results from the improvement in memory with plasticity-based adaptive cognitive training (IMPACT) study. *Journal of the American Geriatrics Society, 1–10*. doi:10.1111/j.1532-5415.2008.02167.x

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

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

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

Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., & Merzenich, M. M. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science, 271,* 81–84.

van Muijden, J., Band, G. P. H., & Hommel, B. (2012). Online games training aging brains: Limited transfer to cognitive control functions. *Frontiers in Human Neuroscience, 6*(221), 1–13. doi:10.3389/fnhum.2012.00221

von Bastian, C. C., Locher, A., & Ruflin, M. (in press). *Tatool: A Java-based open-source programming framework for psychological studies*. Manuscript submitted for publication. doi:10.3758/s13428-012-0224-y

von Bastian, C. C., & Oberauer, K. (2012). *Distinct transfer effects of training different facets of working memory capacity.* Manuscript submitted for publication.

Wasylyshyn, C., Verhaeghen, P., & Sliwinski, M. J. (2011). Aging and task switching: A meta-analysis. *Psychology and Aging, 26,* 15–20. doi:10.1037/a0020912

Whitson, L. R., Karayanidis, F., & Michie, P. T. (2012). Task practice differentially modulates task-switching performance across the adult lifespan. *Acta Psychologica, 139,* 124–136. doi:10.1016/j.actpsy.2011.09.004

Zinke, K., Zeintl, M., Eschen, A., Herzog, C., & Kliegel, M. (2012). Potentials and limits of plasticity induced by working memory training in old-old age. *Gerontology, 58,* 79–87. doi:10.1159/000324240