



Hundred days of cognitive training enhance broad cognitive abilities in adulthood: findings from the COGITO study

Florian Schmiedek^{1,2*}, Martin Lövdén^{1,3} and Ulman Lindenberger¹

¹ Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany

² Center for Research on Education and Human Development, German Institute for International Educational Research, Frankfurt am Main, Germany

³ Department of Psychology, Lund University, Lund, Sweden

Edited by:

Lars Nyberg, Umeå University, Sweden

Reviewed by:

Bert Jonsson, Umeå University, Sweden

Cindy Lustig, University of Michigan, USA

*Correspondence:

Florian Schmiedek, Center for Lifespan Psychology, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany.
e-mail: schmiedek@mpib-berlin.mpg.de

We examined whether positive transfer of cognitive training, which so far has been observed for individual tests only, also generalizes to cognitive abilities, thereby carrying greater promise for improving everyday intellectual competence in adulthood and old age. In the COGITO Study, 101 younger and 103 older adults practiced six tests of perceptual speed (PS), three tests of working memory (WM), and three tests of episodic memory (EM) for over 100 daily 1-h sessions. Transfer assessment included multiple tests of PS, WM, EM, and reasoning. In both age groups, reliable positive transfer was found not only for individual tests but also for cognitive abilities, represented as latent factors. Furthermore, the pattern of correlations between latent change factors of practiced and latent change factors of transfer tasks indicates systematic relations at the level of broad abilities, making the interpretation of effects as resulting from unspecific increases in motivation or self-concept less likely.

Keywords: cognitive training, cognitive abilities, transfer, latent factors, working memory

INTRODUCTION

The goal of improving cognitive abilities through practice and training has been pursued in numerous intervention studies, particularly with older adults (for reviews, see Baltes and Lindenberger, 1988; Verhaeghen et al., 1992; Kramer and Willis, 2002; Rebok et al., 2007; Hertzog et al., 2009; Lustig et al., 2009; Noack et al., 2009). The main conclusions drawn from this research were that: (a) cognitive performance can be substantially improved through strategy training and practice up to very old age; (b) performance gains can be maintained up to several years; and (c) positive transfer of training to non-practiced tasks is generally non-existent or small (see also Owen et al., 2010).

The widespread absence of positive transfer in cognitive intervention studies is consistent with the longstanding distinction between the *acquisition of skills*, with limited applicability beyond trained tasks, and the *improvement of abilities*, denoting gains in general mechanisms and capacities that carry the potential for improved performance across a wide range of tasks (cf. Thorndike, 1906). If training does not just improve task-specific skills but also broad cognitive abilities (cf. Carroll, 1993), then even small effects could lead to important benefits for individuals' everyday intellectual competence, as these improvements would generalize to all sorts of cognitive activities. Furthermore, even small delays or reductions of age-associated declines in cognitive abilities could substantially prolong individuals' capacity for leading independent lives (Hertzog et al., 2009). However, if training and transfer effects are restricted to the tasks in question, such benefits would have little practical significance.

Going beyond earlier attempts to improve cognitive abilities through strategy training in the domains of episodic memory (EM) and fluid intelligence (Gf), recent approaches have targeted specific cognitive functions that are presumed to form the

basis of general cognitive abilities. Examples for such functions are working memory (WM) and executive functions (Klingberg et al., 2005; Diamond et al., 2007; Dahlin et al., 2008; Jaeggi et al., 2008; Li et al., 2008; Karbach and Kray, 2009), perceptual speed (PS; Ball et al., 2007), or sensory discrimination (Mahncke et al., 2006). Jaeggi et al. (2008) recently reported transfer of WM training to performance on a matrices test of Gf. However, as transfer of training had been demonstrated for only one task, the study does not warrant strong conclusions about the scope of transfer (Sternberg, 2008; Moody, 2009). Specifically, transfer may have just been restricted to a single test of Gf based on figural materials and may have been absent for other fluid tests with figural content, or for tests using numerical or verbal materials. Generally, demonstrating that transfer has occurred for a single test of a particular ability is insufficient for claiming that this ability has improved. Whenever transfer effects are reported for just one task, it is more parsimonious to assume that the observed gains, both on the trained and on the transfer task, result from commonalities in mechanisms and content that does not generalize to cognitive abilities.

To overcome these limitations, proposals have been made to base the evaluation of transfer effects on changes at the ability level (Lövdén et al., 2010; McArdle and Prindle, 2008; Noack et al., 2009). Using established hierarchical structures of cognitive abilities (e.g., Carroll, 1993) as a reference frame, arbitrary definitions of near vs. far transfer can be replaced with definitions of transfer in terms of narrow, broad, or general abilities. Such a shift of transfer assessment from tasks to abilities can be achieved in two steps. First, the targeted abilities have to be indexed by more than one task, and preferably with tasks that vary both in procedure and content to represent the given ability in a sufficiently broad way (Little et al., 1999).

Second, statistical analyses need to focus on the variance common to the different tasks. Factor analysis is a suitable statistical tool to this end, as it separates common variance from task-specific variance and measurement error. Specifically, confirmatory factor-analytic techniques (Bollen, 1989) can be used to test specific hypotheses about transfer effects by (a) specifying in advance which tasks represent which construct and (b) making sure that these relations are invariant across experimental groups and occasions. When applied to multiple measures administered before and after an intervention, these techniques allow differences in performance between pre-test and post-test to be extracted “at the latent level”, that is, at a level that represents the components of the variance common to the set of tasks indexing a given ability. If positive transfer effects, defined as reliably larger performance gains in the intervention group than in the control group, are found at the level of latent factors, then there is an empirical basis for interpreting the results as improvements in cognitive abilities. The definition of these abilities as narrow, broad, or general provides common ground for the comparison of transfer effects from different studies. Clearly, these definitions are contingent upon the structural model of cognitive abilities that is chosen as a point of reference. In our case, we opted for the taxonomy by Carroll (1993), which has been validated across a wide range of test batteries and samples. Because reliability of observed performance measures of single tasks is generally limited, it can even be the case that improvements on several individual tasks are not statistically significant, while the improvement at the level of a common factor of these tasks is. This finding would denote still stronger evidence for transfer at the ability level than significant improvements on an individual task that is not accompanied by improvement on the corresponding latent factor.

Based on this rationale, we investigate whether practice in cognitive tasks can lead to enhancements at the latent ability level in younger and older adults. We report data from the COGITO Study, which is both more intensive and more extensive than any other cognitive intervention study conducted so far. The intervention was intensive because participants, 101 younger and 103 older adults, practiced twelve computerized tasks representing the cognitive abilities of PS, EM, and WM, and extensive because the amount of training comprised over 100 daily sessions of about 1 h each.

Normal aging is associated with a loss of cognition-related brain resources (Craik, 1983; Park et al., 1996; Lindenberger and Baltes, 1997). Animal models (Kempermann, 2008) and age-comparative training studies with humans (Brehmer et al., 2007; Shing et al., 2008) suggest that cognitive plasticity is reduced but not completely lost in old age. We therefore hypothesized that cognitive intervention would be more effective in improving cognitive abilities in younger than in older adults.

Similar to other recently developed training programs (Klingberg et al., 2005; Dahlin et al., 2008; Jaeggi et al., 2008), our battery included three tests of WM. In addition, it also contained three tests of EM and six tests of PS. Furthermore, for each of the three abilities the chosen tasks varied in procedures and content, consisting of verbal, numerical, or figural-spatial information. Hence, the psychometric space of human cognitive abilities was represented more broadly than in earlier studies, which has at least two advantages regarding the likelihood of producing transfer effects.

First, the varied selection of tasks allows increasing the dosage of training, compared to most existing studies, by keeping up motivation of the participants to practice for many sessions. Second, from a person-centered perspective, training tasks of more than one ability increases the chances of improvements in at least one of the abilities. By administering a large battery of transfer tasks at pre- and post-test, we were able to comprehensively examine the amount and scope of transfer effects. Specifically, the present battery included three near-transfer WM tasks, three far-transfer WM tasks, a word-pair EM task, a set of Raven Matrices, and a total of 27 tests from a standard paper-and-pencil battery designed to measure Gf, EM, and PS. To separate training from retest effects, age-matched test-retest-only control groups were assessed as well. We describe these groups as no-training control groups, as pre- and post-test sessions each involved 10 sessions of 2.0–2.5 h filled with mostly cognitive testing, so that labeling these groups as “no-contact” controls would be misleading.

MATERIALS AND METHODS

PARTICIPANTS AND PROCEDURE

During the training phase of the study, 101 younger (51.5% women, age: 20–31 years) and 103 older adults (49.5% women, age: 65–80 years) completed an average of 101 practice sessions (younger adults: $M = 100.8$, $SD = 2.6$, range = 87–109; older adults: $M = 101.0$, $SD = 2.7$, range = 90–106). Participants practiced individually in lab rooms containing up to six computer testing places. Participants in the no-training control group were 44 younger (47.7% women, age: 21–29 years) and 39 older adults (48.7% women, age: 65–81 years). Practice and control groups were matched on age, initial cognitive status, and education (Table 1). Both younger and older samples were quite representative regarding general cognitive functioning, as indicated by comparisons of Digit-Symbol performance with data from a population-based study and a meta-analysis (see Figure 1). Attrition rate for those participants who had entered the longitudinal practice phase was low (i.e., 15 out of 219 participants; for details on rates and reasons of dropout in the different study phases, see Schmiedek et al., 2010).

Before and after the longitudinal phase, participants completed pre- and post-tests during 10 sessions that consisted of 2.0–2.5 h of comprehensive cognitive test batteries and self-report questionnaires. On average, the time that elapsed between pre- and post-test was 197 vs. 193 days for the younger and 188 vs. 189 days for the older intervention and control groups, respectively. Participants in the intervention groups were paid between 1450 and 1950 EUR, depending on the number of completed sessions and their pace of completing the longitudinal phase of the study. Participants in the control groups were paid 460 EUR.

In each practice session, participants practiced 12 different tasks. For PS, these were three two-choice reaction tasks (odd vs. even numbers; consonants vs. vowels; symmetric vs. asymmetric figures) and three comparison tasks (two strings of digits; two strings of consonants; two three-dimensional figures). For EM, participants had to memorize word lists, number-word pairs, or object positions in a grid. WM tasks were adapted versions of the alpha span (Craik, 1986), numerical memory updating (Salthouse et al., 1991), and spatial n-back (Cohen et al., 1997)

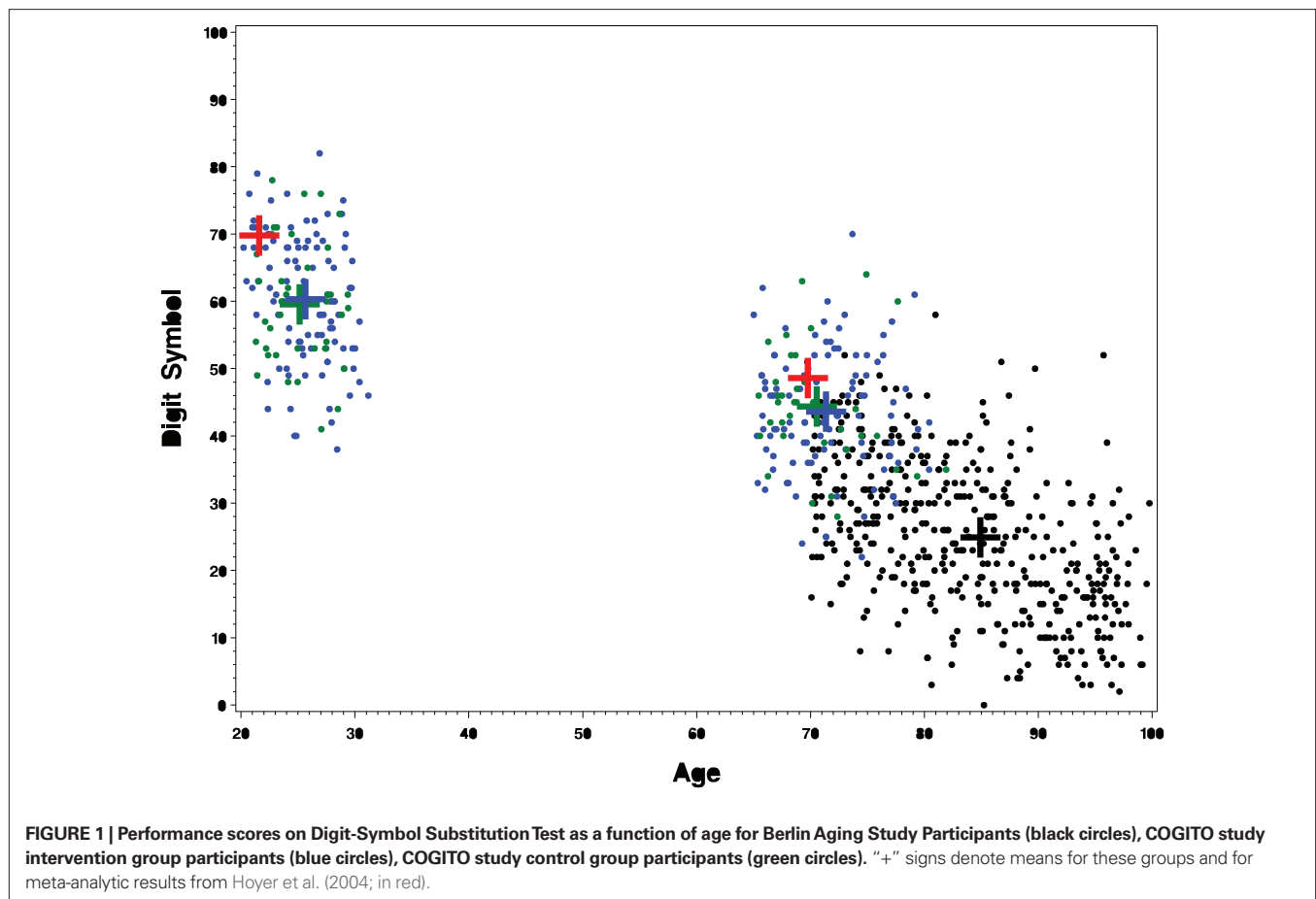


Table 1 | Sample means (SDs) of practice and control groups.

	Younger adults practice group	Control group	Older adults practice group	Control group
Age	25.6 (2.7)	25.2 (2.5)	71.3 (4.1)	70.6 (4.0)
Years of education	16.1 (3.2)	15.7 (2.7)	13.6 (3.6)	13.0 (3.9)
WAIS Digit-Symbol	60.3 (9.5)	59.6 (8.7)	43.6 (9.0)	44.4 (8.8)
Accuracy Spot-a-Word Test (Lehrl, 1991)	0.66 (0.10)	0.64 (0.11)	0.81 (0.10)	0.79 (0.09)

tasks. Tasks were carried out in small groups of 2–5 participants at a PC, using the keyboard, mouse, as well as special button boxes for the different tasks. To maximize and even out the cognitive challenge of these tasks across individuals, while also maintaining motivation, difficulty levels for the EM and WM tasks were individualized using different presentation times (PT) based on pre-test performance. For each task and each individual, mean accuracies for the different PT conditions at pre-test were fitted with exponential time-accuracy functions (including freely estimated parameters for onset, rate, and asymptote as well as a lower asymptote parameter fixed to different values for each task, e.g., 0.10 for memory updating). The fitted values from these functions were used to choose PTs that are clearly above random guessing but below some upper level. The upper level was defined by the midpoint between the lower asymptote level and perfect accuracy

[e.g., $(0.10 + 1.0)/2 = 0.55$ for memory updating], the minimum level was defined by the midpoint between lower asymptote level and the upper level [e.g., $(0.10 + 0.55)/2 = 0.325$ for memory updating]. The PT was then chosen so that the predicted performance level based on the TAF was above the minimum level and below the upper level. If performance was above the upper level for the second-but-fastest PT, then the fastest PT was chosen, even if predicted accuracy was below the minimum level for the fastest PT. Lower asymptote level was set to 0.10 for memory updating, to 0.50 for the 3-back and the choice reaction tasks, and to 0.00 for the EM tasks. For the Alpha span task, we deviated from the described procedure and chose 0.00 as lower asymptote, 0.40 as minimum level, and 0.60 as upper level, based on empirically observed TAFs. Regarding the choice reaction tasks, the fast masking time was chosen based on a medium level of 0.625 and an upper level of

0.75, while for the slow masking time, those levels were 0.875 and 0.95, respectively. PTs and masking times were kept constant over the intervention period.

PRACTICED TASKS

Perceptual speed: choice reaction tasks (CRTs)

All three CRTs were based on the same stimulus layout, the seven lines of the number “8” as displayed on hand calculators. Stimuli were masked with a stimulus that combined this “calculator 8” with extending lines in all 10 possible directions. Possible masking times were 2, 4, or 8 screen cycles (24, 47, or 94 ms). Depending on pre-test performance, two of these masking times (one fast and one slow condition) were chosen for each participant. Each CRT trial consisted of 40 stimuli, 20 for the fast and 20 for the slow condition, with randomly chosen stimuli from the two response categories. Two trials of each CRT were included in each daily session. The stimuli were odd and even numbers for the numerical CRT, consonants and vowels for the verbal CRT, and symmetric or asymmetric combinations of lines for the figural CRT.

Perceptual speed: comparison tasks

For the numerical version of the comparison task, two strings of five numbers each appeared on the left and right of the screen, with participants having to decide as quickly as possible whether both strings were exactly the same or different. If different, the strings differed by just one number. Number strings were randomly assembled using digits 1–9. The verbal version of this task was equivalent to the numerical one, using strings of five consonants. In the figural version, two “fribbles”, that is, three-dimensional colored objects consisting of several connected parts, were shown to the left and right of the screen, with participants having to decide as quickly as possible whether the two objects were exactly the same or different. If different, the objects differed with respect to one part. The fribble images in this task were courtesy of Michael J. Tarr, Brown University, <http://www.tarrlab.org/>). In each session, two trials of 40 items were included for each of the verbal, numerical, and figural tasks.

Episodic memory

Word lists. Lists of 36 nouns were presented sequentially with PT individually adjusted based on pre-test performance. PT were 1000, 2000, or 4000 ms. ISI was 1000 ms. Word lists were assembled in such a way that word frequencies, word lengths, emotional valence, and imaginability were balanced across lists. After presentation, the first three letters of each word had to be entered in the correct order using the keyboard. The performance scores were based on the number of words correctly recalled multiplied by the accuracy of their order (ranging from 0 for reverse order to 1 for perfect order). Two trials were included in each daily session.

Number-noun pairs. Lists of 12 two-digit numbers and nouns in plural case pairs (e.g., 22 dogs) were presented sequentially with PT individually adjusted based on pre-test performance. PT were 1000, 2000, or 4000 ms. ISI was 1000 ms. After presentation, the nouns appeared in random order and the corresponding numbers had to be entered. Two trials were included in each daily session.

Object position memory. Sequences of 12 colored photographs of real-world objects were displayed at different locations in a 6 by 6 grid with PT individually adjusted based on pre-test performance. PT were 1000, 2000, or 4000 ms. ISI was 1000 ms. After presentation, objects appeared at the bottom of the screen and had to be moved in the correct order to the correct locations by clicking on the objects and the locations with the computer mouse. Two trials were included in each daily session.

Working memory

Alpha span. Ten upper-case consonants were presented sequentially together with a number below the letter. For each letter, participants had to decide as quickly as possible whether the number corresponded to the position of the current letter in the alphabet within the set of letters presented up to this step. Five of the ten items were targets. If the position numbers were incorrect (non-targets) they differed from the correct position by ± 1 . The presentation time for the letters was individually adjusted based on pre-test performance. The possible PT were 750, 1500, and 3000 ms. ISI was 500 ms. Eight trials were included in each daily session.

Memory updating numerical. Four single digits (ranging from 0 to 9) were presented simultaneously in four cells situated horizontally for 4000 ms. After an ISI of 500 ms, a sequence of eight updating operations were presented in a second row of four cells below the first one. These updating operations were additions and subtractions within a range of -8 to $+8$. Those updating operations had to be applied to the digits memorized from the corresponding cells above and the updated results had to be memorized. Each updating operation was applied to a different cell from the one a step earlier in the sequence, so that no two updating operations had to be applied to one cell in a sequence. The presentation time for each updating operation was individually adjusted based on pre-test performance. Possible PT were 500, 1250, and 2750 ms. ISI was 250 ms. At the end of each trial, the four end results had to be entered in the four cells in the upper row. All intermediate and end results ranged between 0 and 9. Eight trials were included in each daily session.

3-Back spatial. A sequence of 39 black dots appeared at varying locations in a 4 by 4 grid. Participants were supposed to recognize whether each dot was in the same position as the dot three steps earlier in the sequence or not. Dots appeared at random locations with the constraints that (a) 12 items were targets, (b) dots did not appear in the same location in consecutive steps, (c) exactly three items each were 2-, 4-, 5-, or 6-back lures, that is, items that appeared in the same position as the items 2-, 4-, 5-, or 6 steps earlier. No lures of lags longer than 6 were included. The presentation rate for the dots was individually adjusted based on pre-test performance by varying the ISI. The presentation time for the dots was always 500 ms. ISI was 500, 1500, or 2500 ms. Four trials were included in each daily session.

TRANSFER TASKS

Transfer tasks included computerized tasks as well as 27 tasks from the paper-and-pencil *Berlin Intelligence Structure Test* (BIS; Jäger et al., 1997). The three near-transfer WM tasks were based on the same three paradigms as the practiced WM tasks, but used different

content material. The far-transfer WM tasks were established complex span tasks. For EM, one computerized word paired-associates task and nine tasks from the BIS (three for each content domain) were used to assess transfer. Transfer in Gf was measured with 15 items from the *Raven's Advanced Progressive Matrices* as well as with nine tasks from the BIS, three for each content domain. Transfer on PS was measured with three BIS tasks for each content domain.

WM near transfer: updating tasks

Animal span. As in the alpha span task, a list of consecutively shown stimuli had to be ordered continuously. Instead of letters, six names of animals were shown one after the other, which had to be ordered by size and two-choice decisions on whether a given number corresponds to the current rank order of the present animal had to be made. PT was 3000 ms with an ISI of 1000 ms. Eight blocks were conducted in total.

3-Back numerical. As in the spatial version of the 3-back, two-choice decisions on whether the current stimulus matches the stimulus shown 3 steps earlier in the sequence had to be made. Instead of spatial positions, the 39 stimuli were one-digit numbers (1–9). PT was 3000 ms with an ISI of 1000 ms. Six blocks were conducted in total.

Memory updating spatial. In each block of this task, first a display of four 3×3 grids was shown for 4000 ms in each of which one black dot was present in one of the nine locations. Those four locations had to be memorized and updated according to shifting operations, which were indicated by arrows appearing below the corresponding field. PT of the arrows was 2750 ms with an ISI of 250 ms. After six updating operations, the four grids reappeared and the resulting end positions had to be clicked on. After 18 practice blocks with memory load two and six blocks with load three, twelve test blocks with load four were conducted and used for scoring.

WM far transfer: complex span tasks

Reading span. We used a version that differed from the original version in that participants did not have to memorize words but single letters (cf. Kane et al., 2004). Several sentences were presented successively. Below each sentence, a letter was displayed. Participants had to decide whether the sentences were semantically correct, to memorize the letter, and, after a sequence of sentence-letter combinations, recall the letters in their order of presentation. Twelve blocks of trials, three for each load-level (of 2–5) were included.

Counting span. Our version of Counting Span (CS) was similar to the one used by Kane et al. (2004). Several displays of blue circles (4–9), green circles (1–5), and blue squares (1–9) were presented. Participants had to count the blue circles and make decisions as to whether the number was odd or even. The numbers of blue circles had to be memorized for later recall in the order of their presentation. The number of displays ranged from 2–6 per block of trials. A total of 15 blocks was completed, three per load-level.

Rotation span. This task combines recall of a sequence of short and long arrows, radiating from the center of the display, with a letter-rotation task (Kane et al., 2004; Wilhelm and Oberauer, 2006). First, a regular or mirror-reversed letter (rotated by 0–315°) was

displayed. The processing requirement was to decide whether letters were displayed regularly or mirror-reversed. After each processing step (ranging from 2–5 per block), short or long arrows were shown, pointing in one of the eight directions. At the end of one sequence, participants had to recall the direction and length of the arrows in the order of their presentation and indicate them by clicking on a layout with the 16 possible positions of the arrow head. There were 12 blocks of trials to complete, three per load level.

Episodic memory: word pairs

In this task, a set of 30 randomly combined pairs of nouns was shown for 5000 ms each with an ISI of 1000 ms. Immediately afterwards, the first nouns of the pairs appeared as cues in random order and the first three letters of the paired noun had to be entered with the keyboard.

Paper-and-Pencil Tasks

For each ability (Gf, EM, and PS), nine tasks from the BIS test were included in the transfer task battery, three from each of the three content domains. Descriptions of these tasks are available from the authors.

DATA ANALYSIS

Effect sizes for single tasks (d) were calculated separately for younger and older adults as mean pre-post differences in accuracy divided by the SD at pre-test. Net effects were obtained by subtracting the effect sizes for the control from those of the experimental groups. Whether these net effects were statistically significant was investigated separately for each age group by testing the interaction of occasion (pre vs. post) and group (practice vs. control) with mixed models that allowed for different variances at pre- and post-test. Effects at the latent level were analyzed with *latent difference score models* (McArdle and Nesselrode, 1994; McArdle and Prindle, 2008; **Figure 2**). In these models, latent factors are defined by a set of observed variables (i.e., transfer tasks). Factor loadings represent the strength of the relationship between the latent factor and the observed task, that is, to what degree relations to the latent factor account for the means and variance of the observed tasks. Latent difference score factors captured improvements at the latent level. To render the metric of latent factors interpretable, factor loadings and intercepts were constrained to be equal across occasions and experimental groups (strong measurement invariance). Nested model comparisons were used to test whether imposing these constraints was tenable. The evaluation of model fit was based on recommendations by Hair et al. (1998). Latent effect sizes were calculated by dividing the latent mean differences by the latent SDs at pre-test.

For analyses of the BIS test, tasks were parceled for each ability construct by calculating composites of standardized scores for the three tasks of each content domain. As these scores were thus already standardized based on pre-test SDs, mean differences are in effect-size metric and do not need to be divided by SDs.

For WM near transfer, model fit for strong measurement invariance was reasonable, $\chi^2[44] = 75.9$, CFI = 0.95, RMSEA = 0.10. For Gf, model fit for strong measurement invariance was also reasonable, $\chi^2[44] = 61.9$, CFI = 0.98, RMSEA = 0.08, and for EM, it was good, $\chi^2[44] = 47.6$, CFI = 1.00, RMSEA = 0.03.

RESULTS

For both age groups, training gains were reliable and of medium to strong size for practiced tasks, with the exception of the word list EM task in the older adult group (see **Table 2**). Pre-test differences between training and control groups were not significant for both age groups (all $p>0.05$) with the exception of animal span. For this

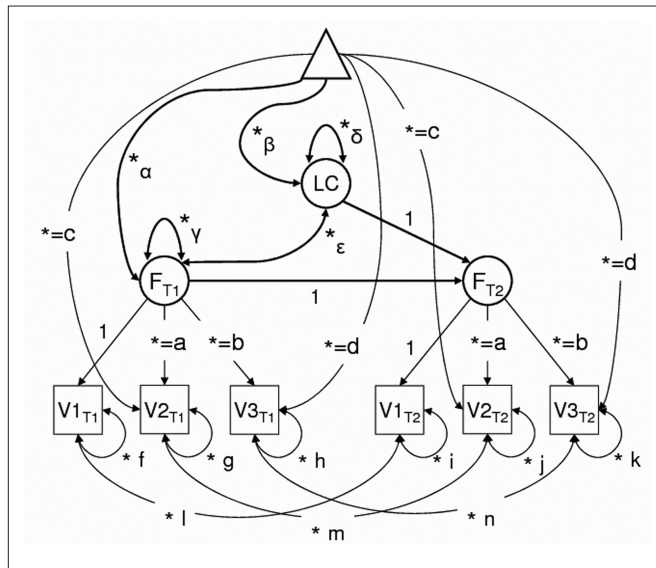


FIGURE 2 | Latent difference score model for modeling training-induced changes at the latent factor level. Squares represent observed variables, circles represent latent factors and the triangle serves to represent information regarding means and intercepts. Free parameters are indicated by asterisks. Parameters with equal sign and the same subscript are constrained to be equal to each other (i.e., strong measurement invariance with equal factor loadings and intercepts across occasions and across experimental and control groups). T1: Pre-test occasion; T2: Post-test occasion; V1–V3: observed variables (i.e., tasks of one ability); F: latent factor of ability; LC: Latent change factor; α : latent mean of ability factor at pre-test; β : mean difference between latent ability factors at pre- and post-test (=latent change, or latent difference score); γ : variance (individual differences) in latent ability at pre-test; δ : variance (individual differences) in latent ability changes between pre- and post-test; ϵ : covariance between individual differences in latent ability at pre-test and latent changes. For further information on two-occasion latent difference modeling in general, see McArdle and Nesselroade (1994).

task, the older training group had significantly lower accuracy at pre-test than the older control group (0.57 vs. 0.63; $t[140] = -2.75$, $P = 0.007$).

Regarding transfer of training to the WM near-transfer tasks (**Figure 3** and **Table 3**), significant experimental group \times occasion interactions were observed for animal span in the older group, $F(1,140) = 4.97$, $P = 0.027$, $d = 0.42$, and for 3-back numerical in the younger group, $F(1,143) = 8.15$, $P = 0.005$, $d = 0.42$. At the latent factor level, significant training gains were obtained for younger adults, $\chi^2[1] = 5.11$, $P = 0.024$, $d = 0.36$, and for older adults, $\chi^2[1] = 6.44$, $P = 0.011$, $d = 0.31$. The fit of the model with strong measurement invariance to the data was reasonable, $\chi^2[44] = 75.9$, CFI = 0.95, RMSEA = 0.10. The group \times occasion interaction for transfer of training to the WM far-transfer tasks (**Figure 3**) was significant only for older adults and the Rotation span (RoS) task, $F(1,140) = 14.95$, $P < 0.001$, $d = 0.60$. This might be due to the RoS task being more difficult and thereby leaving more room for improvement than CS and Reading Span (RS), for which performance was close to ceiling for many younger and older adults already at pre-test. Latent factors could not be modeled for the WM far-transfer tasks because the necessary requirements for measurement invariance were not met.

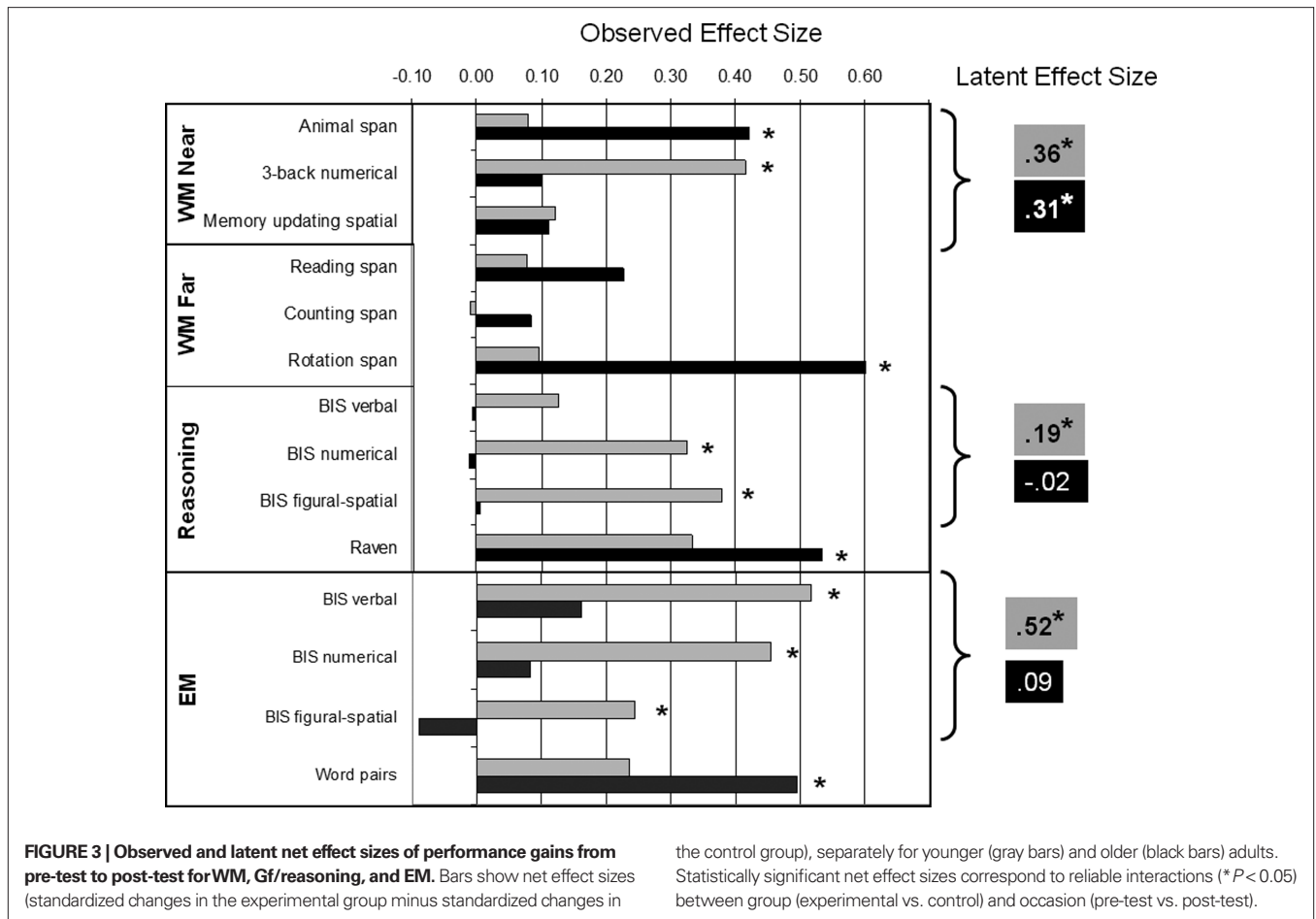
For transfer of training to the Gf tasks (**Figure 3**), results differed again by age group. For younger adults, the group \times occasion interaction was significant for the parcels of numerical tasks, $F(1,143) = 7.48$, $P = 0.007$, $d = 0.33$, and figural-spatial tasks, $F(1,143) = 7.84$, $P = 0.006$, $d = 0.38$, and marginally significant for the Raven, $F(1,142) = 3.73$, $P = 0.056$, $d = 0.33$. For older adults, performance increased more in the intervention than in the control group only for the Raven, $F(1,138) = 5.93$, $P = 0.022$, $d = 0.54$. At the latent factor level, significant training gains were obtained for younger adults, $\chi^2[1] = 11.62$, $P < 0.001$, $d = 0.19$, but not for older adults, $\chi^2[1] = 0.05$, $P = 0.830$, $d = -0.02$. The fit of the model with strong measurement invariance was reasonable, $\chi^2[44] = 61.9$, CFI = 0.98, RMSEA = 0.08.

Moving on to EM in younger adults, the group \times occasion interaction was significant for the verbal, $F(1,143) = 25.03$, $P < 0.001$, $d = 0.42$, the numerical, $F(1,143) = 11.99$, $P = 0.001$, $d = 0.46$, and the figural-spatial BIS parcels, $F(1,143) = 6.45$, $P = 0.012$, $d = 0.24$. In older adults, only the interaction for the word-pairs test was reliable, $F(1,140) = 10.42$, $P = 0.002$, $d = 0.50$. At the latent level, reliable

Table 2 | Practice effects for individual tasks.

	Net effect size younger adults	Interaction pre-post \times experimental group	Net effect size older adults	Interaction pre-post \times experimental group
WM				
Alpha span	1.08	$F(1,143) = 35.69$, $P < 0.001$	1.36	$F(1,140) = 69.38$, $P < 0.001$
Memory updating numerical	1.20	$F(1,143) = 80.74$, $P < 0.001$	0.98	$F(1,140) = 44.42$, $P < 0.001$
N-Back figural	0.90	$F(1,143) = 28.18$, $P < 0.001$	1.46	$F(1,140) = 56.67$, $P < 0.001$
EPISODIC MEMORY				
Word lists	0.40	$F(1,141) = 6.51$, $P = 0.012$	0.25	$F(1,138) = 1.00$, $P = 0.320$
Number-noun pairs	0.43	$F(1,143) = 13.82$, $P < 0.001$	0.81	$F(1,140) = 14.12$, $P < 0.001$
Object position memory	0.82	$F(1,143) = 30.68$, $P < 0.001$	1.00	$F(1,140) = 31.22$, $P < 0.001$

Practice effects were calculated for accuracy averaged across the four presentation time conditions.



training gains were again found for younger adults, $\chi^2[1] = 31.97$, $P < 0.001$, $d = 0.52$, but not for older adults, $\chi^2[1] = 1.22$, $P = 0.269$, $d = 0.09$. The fit of the model with strong measurement invariance was good, $\chi^2[44] = 47.6$, CFI = 1.00, RMSEA = 0.03.

Finally, for the BIS PS tasks, younger adults showed significant group \times occasion interactions for the numerical parcel, $F(1,143) = 6.93$, $P = 0.009$, $d = 0.20$, and the figural-spatial parcel, $F(1,143) = 4.58$, $P = 0.034$, $d = -0.26$. At the latent level, interactions were not reliable.

All effects reported so far are based on comparisons of the intervention to no-training control groups. This raises the question to what degree non-specific influences like general improvements in achievement motivation or self-concept contributed to these effects. To address this issue, we further examined the patterns of correlations among improvements in the experimental group. Specifically, we investigated how strongly individual differences in improvements on the latent factors of the *practiced* tasks were related to individual differences in improvement on the latent factors of the *transfer* tasks. Thus, we correlated the latent change factors of practiced EM and WM with the latent change factors of transfer EM and near-transfer WM (cf. McArdle and Prindle, 2008). Correlations with the Gf latent change factor could not be estimated because individual differences in improvements were not reliable for this factor. In a model combining the latent change models of the practiced EM tasks, the EM transfer tasks, the practiced WM tasks,

and the WM near-transfer tasks, the two WM latent change factors (practiced and transfer) turned out to be perfectly correlated. In a modified model, the correlation of these factors was therefore fixed to one, indicating the presence of one general factor of change in WM. Fit of this model was reasonable, $\chi^2[234] = 368.5$, CFI = 0.91, RMSEA = 0.08. The correlation of the EM transfer latent change factor with the EM practiced latent change factor was $r = 0.58$ and significant ($\Delta\chi^2[1] = 8.4$, $P = 0.004$), while its correlation with the WM latent change factors was only $r = 0.25$ and not significant ($\Delta\chi^2[1] = 1.2$, $P = 0.273$). Similarly, in contrast to their perfect correlation among each other, the correlation of the latent WM factors with the latent change factor of EM practiced tasks was much smaller, $r = 0.52$ ($\Delta\chi^2[1] = 6.3$; $P = 0.012$). Thus, latent correlations of practiced tasks correlated more strongly with transfer tasks of the corresponding than with transfer tasks from a different ability. This dissociation strengthens the interpretation of the demonstrated latent transfer effects as *ability-specific* – that is, broader than task-specific effects but more narrow than potential effects generated by general improvements in motivation or self-concept.

DISCUSSION

We found positive transfer to the latent factor of WM ability in both younger and older adults when the tasks constituting this factor had similar processing requirements but were based on different content material than the practiced tasks. Transfer to the latent factor of

Table 3 | Transfer effects for individual tasks.

	Net effect size younger adults	Interaction pre-post × experimental group	Net effect size older adults	Interaction pre-post × experimental group
WM – NEAR				
Animal span	0.08	$F(1, 143) = 0.26, P = 0.611$	0.42	$F(1, 140) = 4.97, P = 0.027$
N-Back numerical	0.42	$F(1, 143) = 8.15, P = 0.005$	0.10	$F(1, 140) = 0.44, P = 0.507$
Memory updating spatial	0.12	$F(1, 143) = 0.69, P = 0.409$	0.11	$F(1, 140) = 0.62, P = 0.434$
WM – FAR				
Reading span	0.08	$F(1, 143) = 0.20, P = 0.655$	0.23	$F(1, 140) = 1.94, P = 0.166$
Counting span	-0.01	$F(1, 143) = 0.00, P = 0.964$	0.08	$F(1, 140) = 0.21, P = 0.644$
Rotation span	0.10	$F(1, 143) = 0.50, P = 0.481$	0.60	$F(1, 140) = 14.95, P < 0.001$
REASONING				
Verbal	0.13	$F(1, 143) = 1.26, P = 0.263$	-0.01	$F(1, 140) = 0.00, P = 0.951$
Numerical	0.33	$F(1, 143) = 7.48, P = 0.007$	-0.01	$F(1, 140) = 0.01, P = 0.906$
Figural/spatial	0.38	$F(1, 143) = 7.84, P = 0.006$	0.01	$F(1, 140) = 0.00, P = 0.960$
Raven	0.33	$F(1, 142) = 3.73, P = 0.056$	0.54	$F(1, 138) = 5.39, P = 0.022$
EPISODIC MEMORY				
Verbal	0.52	$F(1, 143) = 25.03, P < 0.0001$	0.16	$F(1, 140) = 2.91, P = 0.090$
Numerical	0.46	$F(1, 143) = 11.99, P = 0.001$	0.08	$F(1, 140) = 0.35, P = 0.557$
Figural/spatial	0.25	$F(1, 143) = 6.45, P = 0.012$	-0.09	$F(1, 140) = 0.56, P = 0.457$
Word pairs	0.24	$F(1, 143) = 2.12, P = 0.147$	0.50	$F(1, 140) = 10.24, P = 0.002$
PERCEPTUAL SPEED				
Verbal	0.07	$F(1, 143) = 0.61, P = 0.435$	-0.14	$F(1, 140) = 1.92, P = 0.168$
Numerical	0.20	$F(1, 143) = 6.93, P = 0.009$	0.01	$F(1, 140) = 0.02, P = 0.887$
Figural/spatial	-0.26	$F(1, 143) = 4.58, P = 0.034$	0.04	$F(1, 140) = 0.19, P = 0.667$

complex span WM tasks was not discernible; inspection of raw data revealed that this might be due to ceiling effects. In younger adults, reliable transfer of training was observed for the latent factor of Gf. While the size of this effect was small, its scope was impressive, given that the factor was based on nine very heterogeneous reasoning tasks that share no obvious overlap in task-specific characteristics with the practiced tasks. In older adults, transfer to Gf was not reliable. For EM, results at the latent level were even stronger, but again restricted to younger adults. Finally, the lack of transfer to PS may reflect the considerable lack of similarity between the sensory and psychomotor requirements of the computerized two-choice reaction tasks used for cognitive practice and the paper-and-pencil transfer tasks. Also, visual search and psychomotor skills are known to show little generalization across tasks (Clancy and Hoyer, 1994).

As expected, the results of this study demonstrate that cognitive intervention is more effective in earlier than in later adulthood, supporting the notion that cognitive plasticity declines throughout adulthood and old age (Brehmer et al. 2007; Shing et al., 2008). Surprisingly, and contrary to the predicted and observed pattern of age-related reductions in the scope and amount of transfer, the effect sizes for transfer to WM near were of similar size in the two age groups. Possibly, the individually adjusted but fixed difficulty levels of the WM tasks posed less of a challenge to younger adults than to older adults in later portions of practice so that older adults were operating at a more plasticity-inducing difficulty level than younger adults. Future age-comparative studies with adaptive training procedures (e.g., Klingberg et al., 2005) are needed to resolve this issue. Some

ideas how transfer effects in later adulthood might be strengthened and broadened come to mind. First, we did not adapt task difficulty across the training phase dynamically but kept it at fixed levels based on pre-test performance, which potentially led to less than optimal performance demands. Second, the recent training literature suggests that the most effective ingredients of the training program may have been the WM tasks (e.g., Klingberg et al., 2005; Jaeggi et al., 2008), so the overall effectiveness of the training may increase if more time was allotted to WM practice. Third, most of the paper-and-pencil tasks used for transfer assessment were paced, possibly restricting the opportunity for older adults to display improvements in Gf and EM. Older adults' reliable transfer to the Raven and the Word Pairs task supports this interpretation because both of these tasks were carried out under less time pressure and with less of a need to move on from one task to the next than was the case for the BIS test. To be able to better identify factors that influence the effectiveness of the training, future research should also include training groups that receive more specific training (e.g., only WM tasks) as well as active control groups who participate in interventions that can increase motivation and self-concept but, based on theoretical considerations, should not produce transfer at the ability level.

Whereas flexibility-based changes in cognitive strategies and plasticity-based acquisition of knowledge (e.g., stimuli-response mapping) are probably responsible for the lion's share of improvements in practiced tasks, the observed transfer effects are unlikely to be a mere result of such changes. Rather, the broad and far nature of the transfer effects observed in the younger adults, and their pres-

ence at the ability level, suggests that aspects of cognitive processing efficiency have been improved (Lövdén et al., 2010). If this is the case, then one may also expect that the neural underpinnings of transfer effects are changes that affect the efficiency of the neural system in relatively pronounced ways. Little is however currently known about the neural mechanisms of transfer effects. In one study (Dahlin et al., 2008), transfer of improvements from working-memory (updating) training to an unpracticed working-memory task (n-back) was mediated by changes in striatal activity measured by the BOLD response. Considering the link between release of dopamine and the BOLD response (Schott et al., 2008), the putative role of the dopaminergic system in memory updating (e.g., Hazy et al., 2006), and the sensitivity of the dopaminergic system to working-memory training (McNab et al., 2009), it seems likely that alterations in the dopaminergic system may mediate transfer effects that reflect improved working-memory efficiency. Other neural mechanisms may also play important roles as mediators of transfer effects. For example, white-matter integrity displays experience-dependent plasticity (Scholz et al., 2009) in humans and improvements in white matter may have pronounced effects on the synchronous operations of brain regions that higher-order cognition is highly dependent on (Fields, 2008). Investigating the neural mechanisms of transfer effects is an important avenue for further research.

REFERENCES

- Ball, K. K., Edwards, J. D., and Ross, L. A. (2007). The impact of speed of processing training on cognitive and everyday functions. *J. Gerontol. B* 62, 19–31.
- Baltes, M. M., Maas, I., Wilms, H.-U., Borchelt, M., and Little T. D. (1999). Everyday competence in old and very old age: theoretical considerations and empirical findings. In *The Berlin Aging Study: Aging from 70 to 100*, P. B. Baltes and K. U. Mayer, eds. (New York, NY: Cambridge University Press), pp. 384–402.
- Baltes, P. B., and Lindenberger, U. (1988). On the range of cognitive plasticity in old-age as a function of experience - 15 years of intervention research. *Behav. Ther.* 19, 283–300.
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York, NY: Wiley.
- Brehmer, Y., Li, S.-C., Müller, V., von Oertzen, T., and Lindenberger, U. (2007). Memory plasticity across the lifespan: uncovering children's latent potential. *Dev. Psychol.* 43, 465–478.
- Carroll, J. B. (1993). *Human Cognitive Abilities*. Cambridge: Cambridge University Press.
- Clancy, S. M., and Hoyer, W. J. (1994). Age and skill in visual search. *Dev. Psychol.* 30, 545–552.
- Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., and Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature* 386, 604–608.
- Craik, F. I. M. (1983). On the transfer of information from temporary to permanent memory. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 302, 341–359.
- Craik, F. I. M. (1986). "A functional account of age differences in memory," in *Human Memory and Cognitive Capabilities*. eds F. Klix and H. Hagendorf (North-Holland: Elsevier), 409–422.
- Dahlin, E., Stigsdotter Neely, A. S., Larsson, A., Bäckman, L., and Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science* 320, 1510–1512.
- Diamond, A., Barnett, W. S., Thomas, J., and Munro, S. (2007). Preschool program improves cognitive control. *Science* 318, 1387–1388.
- Fields, R. D. (2008). White matter in learning, cognition and psychiatric disorders. *Trends Neurosci.* 31, 361–370.
- Hair, J. F., Anderson, R. E., Tatham, R. L., and Black, W. C. (1998). *Multivariate Data Analysis*, 5th Edn. Upper Saddle River, NJ: Prentice Hall.
- Hazy, T. E., Frank, M. J., and O'Reilly, R. C. (2006). Banishing the homunculus: making working memory work. *Neuroscience* 139, 105–118.
- Hertzog, C., Kramer, A. F., Wilson, R. S., and Lindenberger, U. (2009). Enrichment effects on adult cognitive development: can the functional capacity of older adults be preserved and enhanced? *Psychol. Sci. Public Interest* 9, 1–65.
- Hoyer, W. J., Stawski, R. S., Wasylshyn, C., and Verhaeghen, P. (2004). Adult age and Digit Symbol Substitution performance: a meta-analysis. *Psychol. Aging* 19, 211–214.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., and Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6829–6833.
- Jäger, A. O., Süß, H.-M., and Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test, BIS-Test. Form 4. Handanweisung*. Göttingen: Hogrefe.
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., and Engle, R. E. (2004). The generality of working-memory capacity: a latent-variable approach to verbal and visuo-spatial memory span and reasoning. *J. Exp. Psychol. Gen.* 133, 189–217.
- Karbach, J., and Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task switching training. *Dev. Sci.* 12, 978–990.
- Kempermann, G. (2008). The neurogenic reserve hypothesis: what is adult hippocampal neurogenesis good for? *Trends Neurosci.* 31, 163–169.
- Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlstrom, K., Gillberg, C. G., Forsberg, H., and Westerberg, H. (2005). Computerized training of working memory in children with ADHD - A randomized, controlled trial. *J. Am. Acad. Child. Psychiatry* 44, 177–186.
- Kramer, A. F., and Willis, S. L. (2002). Enhancing the cognitive vitality of older adults. *Curr. Dir. Psychol. Sci.* 11, 173–177.
- Lehrl, S. (1991). *Manual zum MWT-A [Manual for MWT-A]*. Erlangen: Perimed.
- Li, S.-C., Schmiedek, F., Huxhold, O., Röcke, C., Smith, J., and Lindenberger, U. (2008). Working memory plasticity in old age: practice gain, transfer, and maintenance. *Psychol. Aging* 23, 731–742.
- Lindenberger, U., and Baltes, P. B. (1997). Intellectual functioning in old and very old age: cross-sectional results from the Berlin Aging Study. *Psychol. Aging* 12, 410–432.
- Little, T. D., Lindenberger, U., and Nesselrode, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: when "good" indicators are bad and "bad" indicators are good. *Psychol. Methods* 4, 192–211.
- Lövdén, M., Lindenberger, U., Bäckman, L., Schaefer, S., and Schmiedek, F. (2010). A theoretical framework for the study of adult cognitive plasticity. *Psychol. Bull.* 136, 659–676.
- Lustig, C., Shah, P., Seidler, R., and Reuter-Lorenz, P. A. (2009). Aging, training, and the brain: a review and future directions. *Neuropsychol. Rev.* 19, 504–522.
- Mahncke, H. W., Connor, B. B., Appelman, J., Ahsanuddin, O. N., Hardy, J. L., Wood, R. A., Joyce, N. M., Boniske, T., Atkins, S. M., and Merzenich, M.

- M. (2006). Memory enhancement in healthy older adults using a brain plasticity-based training program: a randomized, controlled study. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12523–12528.
- McArdle, J. J., and Nesselroade, J. R. (1994). “Using multivariate data to structure developmental change,” in *Life-span Developmental Psychology: Methodological Contributions*, eds S. H. Cohen and H. W. Reese (Hillsdale, NJ: Erlbaum), 223–267.
- McArdle, J. J., and Prindle, J. J. (2008). A latent change score analysis of a randomized clinical trial in reasoning training. *Psychol. Aging* 23, 702–719.
- McNab, F., Varrone, A., Farde, L., Jucaite, A., Bystritsky, P., Forsberg, H., and Klingberg, T. (2009). Changes in cortical dopamine D1 receptor binding associated with cognitive training. *Science* 323, 800–802.
- Moody, D. E. (2009). Can intelligence be increased by training on a task of working memory? *Intelligence* 37, 327–328.
- Noack, H., Lövdén, M., Schmiedek, F., and Lindenberger, U. (2009). Cognitive plasticity in adulthood and old age: gauging the generality of cognitive intervention effects. *Restor. Neurol. Neurosci.* 27, 435–453.
- Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., Howard, R. J., and Ballard, C. G. (2010). Putting brain training to the test. *Nature* 465, 775–778.
- Park, D. C., Smith, A. D., Lautenschlager, G., Earles, J. L., Frieske, D., Zwahr, M., and Gaines, C. L. (1996). Mediators of long-term memory performance across the life span. *Psychol. Aging* 11, 621–637.
- Rebok, G. W., Carlson, M. C., and Langbaum, J. B. S. (2007). Training and maintaining memory abilities in healthy older adults: traditional and novel approaches. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 62B, 53–61.
- Salthouse, T. A., Babcock, R. L., and Shaw, R. J. (1991). Effects of adult age on structural and operational capacities in working memory. *Psychol. Aging* 6, 118–127.
- Schmiedek, F., Bauer, C., Lövdén, M., Brose, A., and Lindenberger, U. (2010). Cognitive enrichment in old age: web-based training programs. *GeroPsych.*
- Scholz, J., Klein, M. C., Behrens, T. E., and Johansen-Berg, H. (2009). Training induces changes in white-matter architecture. *Nat. Neurosci.* 12, 1370–1371.
- Schott, B. H., Minuzzi, L., Krebs, R. M., Elmenhorst, D., Lang, M., Winz, O. H., Seidenbecher, C. I., Coenen, H. H., Heinze, H. J., Zilles, K., Duzel, E., and Bauer, A. (2008). Mesolimbic functional magnetic resonance imaging activations during reward anticipation correlate with reward-related ventral striatal dopamine release. *J. Neurosci.* 28, 14311–14319.
- Shing, Y. L., Werkle-Bergner, M., Li, S.-C., and Lindenberger, U. (2008). Associative and strategic components of episodic memory: a life-span dissociation. *J. Exp. Psychol. Gen.* 137, 495–513.
- Sternberg, R. J. (2008). Increasing fluid intelligence is possible after all. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6791–6792.
- Thorndike, E. L. (1906). *Principles of Teaching*. New York: Seiler.
- Verhaeghen, P., Marcoen, A., and Goossens, L. (1992). Improving memory performance in the aged through mnemonic training: a meta-analytic study. *Psychol. Aging* 7, 242–251.
- Wilhelm, O., and Oberauer, K. (2006). Why are reasoning ability and working memory capacity related to mental speed? An investigation of stimulus-response compatibility in choice reaction time tasks. *Eur. J. Cogn. Psychol.* 18, 18–50.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 20 April 2010; paper pending published: 20 May 2010; accepted: 16 June 2010; published online: 13 July 2010.
 Citation: Schmiedek F, Lövdén M and Lindenberger U (2010) Hundred days of cognitive training enhance broad cognitive abilities in adulthood: findings from the COGITO study. *Front. Ag. Neurosci.* 2:27. doi: 10.3389/fnagi.2010.00027
 Copyright © 2010 Schmiedek, Lövdén and Lindenberger. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.