The role of individual differences in cognitive training and transfer

Susanne M. Jaeggi · Martin Buschkuehl · Priti Shah ·
John Jonides

© Psychonomic Society, Inc. 2013

Abstract Working memory (WM) training has recently become a topic of intense interest and controversy. Although several recent studies have reported near- and far-transfer effects as a result of training WM-related skills, others have failed to show far transfer, suggesting that generalization effects are elusive. Also, many of the earlier intervention attempts have been criticized on methodological grounds. The present study resolves some of the methodological limitations of previous studies and also considers individual differences as potential explanations for the differing transfer effects across studies. We recruited intrinsically motivated participants and assessed their need for cognition (NFC; Cacioppo & Petty Journal of Personality and Social Psychology 42:116–131, 1982) and their implicit theories of intelligence (Dweck, 1999) prior to training. We assessed the efficacy of two WM interventions by comparing participants' improvements on a battery of fluid intelligence tests against those of an active control group. We observed that transfer to a composite measure of fluid reasoning resulted from both WM interventions. In addition, we uncovered factors that contributed to training success, including motivation, need for cognition, preexisting ability, and implicit theories about intelligence.

S. M. Jaeggi (✉) · M. Buschkuehl
School of Education, University of California, Irvine, Irvine,
CA 92697-5500, USA
e-mail: smjaeggi@uci.edu

M. Buschkuehl
MIND Research Institute, Irvine, CA, USA

P. Shah · J. Jonides
Department of Psychology, University of Michigan, Ann Arbor,
MI, USA

Keywords Working memory · Reasoning · Skill acquisition · Individual differences · Intelligence

Working memory (WM) is an essential system that underlies the performance of virtually all complex cognitive activities (Shah & Miyake, 1999). Consider mentally multiplying 21×33, reading a complex paragraph, or following a lecture while simultaneously keeping up with the latest posts on a social network community. These tasks all rely on WM, in that they require multiple processing steps and temporary storage of intermediate results in order to accomplish the tasks at hand (Jonides, Lewis, Nee, Lustig, Berman, & Moore, 2008). Thus, WM is the cognitive mechanism that supports active maintenance of task-relevant information during the performance of a cognitive task.

People differ in terms of how much information they can store in WM, as well as of how easily they can store that information in the face of distraction (see, e.g., Engle, Kane, & Tuholski, 1999). Individual differences in WM capacity are highly predictive of scholastic achievement and educational success, and WM is crucial for our ability to acquire knowledge and learn new skills (Pickering, 2006). Given the relevance of WM to educational settings and daily life, it is not surprising that numerous recent studies have attempted to develop interventions that are aimed at improving WM. This research is promising in that evidence is accumulating that some WM interventions result in generalizing effects that go beyond the trained domain, an effect that is termed "transfer" (see, e.g., Lustig, Shah, Seidler, & Reuter-Lorenz, 2009; Morrison & Chein, 2011; Rabipour & Raz, 2012; Zelinski, 2009, for reviews). The most consistent transfer effects have been found on related, but not trained, WM tasks; such effects are commonly termed "near transfer" (e.g., Buschkuehl, Jaeggi, Hutchison, Perrig-Chiello, Dapp, Muller, & Perrig, 2008; Dahlin, Neely, Larsson, Backman, & Nyberg, 2008; Holmes, Gathercole, & Dunning, 2009; Holmes, Gathercole, Place,

Dunning, Hilton, & Elliott, 2010; Li, Schmiedek, Huxhold, Rocke, Smith, & Lindenberger, 2008). In addition to near-transfer effects, some evidence for far-transfer effects has also emerged—that is, generalization to domains that are considerably different from the training task (Barnett & Ceci, 2002). Studies have revealed transfer to executive control tasks (Klingberg, Fernell, Olesen, Johnson, Gustafsson, Dahlstrom and Westerberg, 2005; Klingberg, Forssberg, & Westerberg, 2002; Salminen, Strobach, & Schubert, 2012; Thorell, Lindqvist, Bergman Nutley, Bohlin, & Klingberg, 2009), reading tasks (Chein & Morrison, 2010; García-Madruga, Elosúa, Gil, Gómez-Veiga, Vila, Orjales and Duque, 2013; Loosli, Buschkuehl, Perrig, & Jaeggi, 2012), mathematical performance measures (Witt, 2011), and measures of intelligence (Gf; e.g., Borella, Carretti, Riboldi, & de Beni, 2010; Carretti, Borella, Zavagnin, & de Beni, 2013; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Jaeggi, Buschkuehl, Jonides, & Shah, 2011a; Jaeggi, Studer-Luethi, Buschkuehl, Su, Jonides, & Perrig, 2010; Jausovec & Jausovec, 2012; Klingberg et al., 2005; Klingberg et al., 2002; Rudebeck, Bor, Ormond, O'Reilly, & Lee, 2012; Schmiedek, Lövdén, & Lindenberger, 2010; Schweizer, Hampshire, & Dalgleish, 2011; Stephenson & Halpern, 2013; Takeuchi, Taki, Nouchi, Hashizume, Sekiguchi, Kotozaki and Kawashima, 2013; von Bastian & Oberauer, 2013).

Despite the promise of WM training, the research supporting its effectiveness is not yet conclusive. In particular, the far-transfer effects found in some studies are controversial. First, several studies have reported null effects of training (see, e.g., Craik, Winocur, Palmer, Binns, Edwards, Bridges and Stuss, 2007; Owen, Hampshire, Grahn, Stenton, Dajani, Burns and Ballard, 2010; Zinke, Zeintl, Eschen, Herzog, & Kliegel, 2011), and even studies that have used the same training regimen have sometimes found transfer, and sometimes not (Anguera, Bernard, Jaeggi, Buschkuehl, Benson, Jennett, & Seidler, 2012; Bergman Nutley et al., 2011; Holmes et al., 2009; Jaeggi, Buschkuehl, et al., 2010; Klingberg et al., 2005; Redick, Shipstead, Harrison, Hicks, Fried, Hambrick and Engle, 2013; Thorell et al., 2009). One explanation for these inconsistent results across studies may be individual differences in age, personality or preexisting abilities that limit the effectiveness of training for some individuals (Chein & Morrison, 2010; Shah, Buschkuehl, Jaeggi, & Jonides, 2012; Zinke et al., 2011; Zinke, Zeintl, Rose, Putzmann, Pydde, & Kliegel, 2013; Studer-Luethi, Jaeggi, Buschkuehl, & Perrig, 2012). It is also possible that motivational conditions in a particular study (e.g., the degree to which participants are intrinsically vs. extrinsically motivated to participate) influence the effectiveness of training (Anguera et al., 2012; Jaeggi et al., 2011a). Finally, other experimental conditions—such as training time, experimenter supervision of the training process, group versus single-subject settings, quality of instructions, or feedback—may also have an impact on training outcomes (cf. Basak, Boot, Voss, & Kramer,

2008; Jaeggi et al., 2008; Tomic & Klauer, 1996; Verhaeghen, Marcoen, & Goossens, 1992).

In addition to inconsistent transfer effects across studies, some of the research that has reported evidence of far transfer as a result of WM training has been criticized for methodological flaws and/or potential for alternative explanations of the transfer effects. For example, some studies have not included an active control group, yielding the possibility that the transfer found in those studies may be attributable to a Hawthorne effect (Mayo, 1933) or to placebo effects more generally. Other studies have included an active control group, but the nature of the control task has been criticized as being less demanding, engaging, or believable as an intervention than the task experienced by the WM group. Studies have also been criticized for the use of just one or very few far-transfer tasks, rather than using multiple tasks to represent a cognitive construct such as fluid intelligence. Furthermore, some studies have not reported improvements on near-transfer tasks, making it difficult to assess what the underlying mechanisms of improvement might be and leaving open the possibility of placebo-type factors (cf. Buschkuehl & Jaeggi, 2010; Morrison & Chein, 2011; Rabipour & Raz, 2012; Shipstead, Redick, & Engle, 2012, for further discussions). Finally, it is still unresolved whether transfer effects last beyond the training period, and if so, for how long. Only a handful of studies have tested the long-term effects of training by retesting both the experimental and control groups some time after training completion (Borella et al., 2010; Buschkuehl et al., 2008; Carretti et al., 2013; Jaeggi et al., 2011a; Klingberg et al., 2005; Van der Molen, Van Luit, Van der Molen, Klugkist, & Jongmans, 2010). Indeed, some evidence for long-term effects could be attributed to training, but other effects, such as transfer effects that are only present at a long-term follow-up but not at the posttest ("sleeper effects"), are difficult to interpret (Holmes et al., 2009; Van der Molen et al., 2010).

The aim of our present work was to shed light on some of the unresolved issues outlined above. Specifically, this study was designed with three main goals in mind: (1) to resolve the primary methodological concerns of previous research, (2) to consider how motivation may serve as a moderator of transfer effects and provide a potential explanation for inconsistencies across different training studies, and (3) to assess the long-term effectiveness of training and transfer effects.

We randomly assigned participants to one of two WM interventions or to an active control group. The two WM interventions were similar to ones used by us previously (Jaeggi et al., 2008; Jaeggi et al., 2011a; Jaeggi, Studer-Luethi, et al., 2010). Both interventions were versions of an adaptive n-back task in which participants were asked to indicate whether a stimulus was the same as the one presented n-items previously. If participants succeeded at a particular level of n, the task was made incrementally more difficult by increasing the size of n. One WM intervention was a single

auditory *n*-back task (i.e., using spoken letters as stimuli); the other was a dual *n*-back task in which an auditory *n*-back task was combined with a visuospatial task; that is, spoken letters and spatial locations were presented and had to be processed simultaneously. The control task, which we termed the "knowledge-training task," required participants to answer vocabulary, science, social science, and trivia questions presented in a multiple-choice format (cf. Anguera et al., 2012, Exp. 2, as well as Jaeggi et al., 2011a). This control task was adaptive, in that new items replaced material that was successfully learned. Participants found the control task to be engaging and enjoyable. In that it tapped crystallized knowledge, it served as an effective and plausible training condition that did not engage fluid intelligence or WM. The auditory single *n*-back task was selected because our previous studies had always included a visual training task, and we chose to assess whether a nonvisuospatial *n*-back intervention would lead to improvements in visuospatial reasoning tasks. Since we had reason to believe that the processes underlying *n*-back performance are domain-free (Jaeggi, Seewer, Nirkko, Eckstein, Schroth, Groner, & Gutbrod, 2003; Nystrom, Braver, Sabb, Delgado, Noll, & Cohen, 2000; Owen, McMillan, Laird, & Bullmore, 2005), we hypothesized that transfer to reasoning should not depend on the specific stimuli used in the training task. Finally, and most importantly, we used multiple fluid reasoning tasks that we combined into composite scores as transfer measures in order to investigate whether the effects that we had found previously were test-specific, or whether the effects were more general on a construct level. To that end, we chose three matrix-reasoning tasks, and in addition, we used three visuospatial and three verbal reasoning tasks. The latter selection was based on a study that, among other things, looked into the factor structure of reasoning tasks (Kane, Hambrick, Tuholski, Wilhelm, Payne, & Engle, 2004). On the basis of this study, we selected three tasks with the highest factor loadings on a verbal reasoning factor and three tasks with the highest loadings on a spatial reasoning factor (see Kane et al., 2004, Fig. 5).

Our second goal was to evaluate the effects of motivation on training and transfer. Our previous research with children had provided evidence that motivation may play a substantial role in the effectiveness of training (Jaeggi et al., 2011a). In addition, our research with young adults provided preliminary evidence that motivational factors mediate training outcomes. We compared the training outcomes across several studies with young adults conducted by our research team, and found that transfer effects to measures of Gf were found only when participants either were not paid at all to participate (Jaeggi et al., 2008) or were paid a very modest amount (i.e., $20; Jaeggi, Studer-Luethi, et al., 2010; see also Stephenson & Halpern, 2013). In contrast, in one study that we conducted, participants were paid a substantial fee for participation (i.e., $150; Anguera et al., 2012, Exp. 2), and we found no far-

transfer effects on measures of Gf,[1] although near transfer did occur to measures of WM (Anguera et al., 2012; see also Kundu, Sutterer, Emrich, & Postle, 2013).[2] Three other research groups that used our training paradigm paid participants ~$130, $352, or about $800, and interestingly, they did not find transfer on any of their outcome measures (Chooi & Thompson, 2012; Redick et al., 2013; Thompson, Waskom, Garel, Cardenas-Iniguez, Reynolds, Winter and Gabrieli, 2013). The motivational literature has repeatedly demonstrated that extrinsic rewards such as monetary incentives can severely undermine intrinsic motivation (Deci, Koestner, & Ryan, 1999) and, ultimately, performance (Burton, Lydon, D'Alessandro, & Koestner, 2006). Consistent with this notion, the training curves of the paid studies that did not find far transfer are indeed considerably shallower than those of the earlier, successful studies: Whereas the training gains in the paid studies were only between 1.6 and 1.8 *n*-back levels (Chooi & Thompson, 2012; Redick et al., 2013; Seidler, Bernard, Buschkuehl, Jaeggi, Jonides, & Humfleet, 2010), the gains on our successful studies were 2.3 and 2.6 *n*-back levels, respectively (Jaeggi et al., 2008; Jaeggi, Studer-Luethi, et al., 2010). Thompson and colleagues claim that their training gains were similar to those we observed in our 2008 study, but note that their participants trained roughly twice as long as our participants had, and consequentially, this comparison is not entirely appropriate (Thompson et al., 2013). On the basis of this observation, for the present study we recruited participants for a four-week "Brain Training Study" without payment. By not paying participants, we expected the participants to be intrinsically motivated, and consequently, we hoped to increase the likelihood of training and transfer.

Given that individual differences in intrinsic motivation may play a role in training and transfer, we included two baseline assessments of motivation. In our study with 7- to 12-year-old children, we had found a positive relationship between training improvement and gain in fluid intelligence (Jaeggi et al., 2011a; see also Schweizer et al., 2011; Zhao, Wang, Liu, & Zhou, 2011, for similar findings). The children who did not improve on the training task reported that it was "too difficult and effortful" and disengaged from training. Thus, we used the Need for Cognition Scale (Cacioppo & Petty, 1982) to assess

---

[1] Note that Gf data are not reported in Anguera et al. (2012, Exp. 2) because that article was primarily addressing issues of WM and motor function, building on the findings of Experiment 1. However, the Gf data are available as a technical report (Seidler et al., 2010).

[2] Kundu et al. (2013) paid participants about $360 for study completion. Note, though, that as in Anguera et al.'s (2012) study, Kundu et al.'s participants showed *n*-back-related improvement in other outcome measures. Furthermore, the lack of differential group effects in the Raven's test could also stem from the fact that their control task consisted of Tetris, a game that had previously resulted in training-related improvements in spatial ability itself (e.g., De Lisi & Wolford, 2002; Okagaki & Frensch, 1994; Terlecki, Newcombe, & Little, 2008).

enjoyment of difficult cognitive activities. Our hypothesis was that individuals who do not enjoy challenging cognitive work may not benefit from training as much as those who do.

Additionally, we included the Theories of Cognitive Abilities Scale (Dweck, 1999). This scale classifies individuals as having "fixed" beliefs about intelligence (i.e., that intelligence is innate) or "incremental" or "malleable" beliefs about intelligence (i.e., that intelligence can be modified by experience). A large body of research has found that individuals who hold fixed beliefs about intelligence are more likely to withdraw or disengage from tasks that are perceived as being too challenging, whereas individuals who hold incremental beliefs are more likely to persist in challenging tasks (e.g., Blackwell, Trzesniewski, & Dweck, 2007; Grant & Dweck, 2003; Hong, Chiu, Dweck, Lin, & Wan, 1999; Mueller & Dweck, 1998). It is also plausible to assume that those who hold incremental beliefs may be more susceptible to placebo effects because of heightened training expectations. Thus, we expected one of three possible outcomes. One was that participants who received WM training would improve on training and transfer tasks, but only if they held incremental beliefs about intelligence. In contrast, participants with fixed beliefs might disengage and not improve, regardless of training condition. A second possible pattern of outcomes was that participants with incremental beliefs would improve on the transfer tasks regardless of training condition (WM or control) due to an unspecific placebo effect, but that individuals with fixed beliefs would not improve, regardless of training task. This would support the idea that earlier training studies with a no-contact control group were successful due to placebo effects. Third, it is possible that participants with incremental beliefs improve regardless of whether they receive WM or control training (general placebo effect), whereas individuals with fixed beliefs only benefit when they receive WM training (WM training effect). This third possibility implies that training has a real effect, as well as a placebo effect for some individuals. Such a pattern might explain why training and transfer effects with good randomized controls are difficult to find (because even some individuals in the control group show a placebo effect), even though WM training may, in fact, be effective.

Finally, our study was intended to address whether or not we could find long-term effects of training. As in our study with children (Jaeggi et al., 2011a), we included a follow-up measurement three months after the completion of training in order to test for long-term transfer effects.

Method

Participants

A group of 175 participants from the University of Michigan and the Ann Arbor community took part in the present study (mean age = 24.12 years, SD = 6.02, range = 18–45; 99 women, 76 men). They volunteered to participate in a study advertised as a "Brain Training Study" and did not receive payment or course credit. They were recruited via flyers and various online resources, such as Facebook. Fifty-four participants (31 %) withdrew from the study after having completed one or two pretest sessions and having trained for no more than three sessions, largely because of time constraints; most of them (N = 36) never trained at all. Forty-three participants (25 %) dropped out at some point during the training period and/or failed to complete the posttest, after having trained for 11.58 sessions on average (SD = 5.45, range = 4–20).[3] Note that the dropout rates did not differ among groups [$\chi^2(2)$ = 2.62; see Table 1]. The final group of participants, who completed pre- and posttesting and a minimum of 17 training sessions, consisted of 78 individuals (mean age = 25.21 years, SD = 6.46, range = 18–45; 36 women, 42 men).

We randomly assigned participants to one of the three groups until we had a sample of 12 participants. All subsequent participants were assigned to a training group, so that the three groups would remain as similar as possible on the following variables: gender, age, and pretest performance on the APM and the CFT, which were assessed in the first pretest session (cf. Jaeggi et al., 2011a). In addition to the participants recruited for the "brain training" study, 34 participants (mean age = 22.79 years, SD = 6.11, range = 18–44; 17 women, 17 men) were recruited via flyers to take part in the baseline measurement sessions only, and they were paid at an hourly rate of $15.

Basic demographic data for the different training groups, as well as for the groups that did not end training, are given in Table 1, as well as in Table S1 (Supplementary Online Material).

Transfer measures

Visual reasoning tests

Raven's Advanced Progressive Matrices (APM; Raven, 1990) This test consists of a series of visual inductive reasoning problems arranged by increasing difficulty. Each problem consists of a 3×3 matrix of patterns in which one pattern in the lower right corner is missing. Participants are required to select the pattern that appropriately fits into the missing slot by choosing from amongst eight response alternatives. After having completed Set I as practice items (12 items), participants worked on half of the test items of Set II (18 items out of 36). Since our data from previous work (Jaeggi, Buschkuehl, et al., 2010; Jaeggi, Studer-Luethi, et al., 2010, Study 1)

---

[3] Even though the present participants were not paid for participation, the 25 % attrition rate is comparable to or lower than the rates reported in other cognitive training studies. For example, Chooi and Thompson (2012) reported a drop-out rate of 28 %, and Redick et al. (2013) reported a drop-out rate of approximately 42 %.

Table 1 Number of training sessions completed as a function of intervention group

	Completed Training		Dropped Out	
	N	Training sessions	N	Training sessions
Dual N-back	25	19.60	18	13.28
(Percentage) (SD; range)	(58 %)	(1.08; 17–22)	(42 %)	(5.45; 5–20)
Single N-back	26	20.00	16	10.13
(Percentage) (SD; range)	(62 %)	(0.00; 20–20)	(38 %)	(4.37; 4–19)
Knowledge Trainer	27	19.89	9	10.78
(Percentage) (SD; range)	(75 %)	(0.58; 18–21)	(25 %)	(7.76; 4–20)

We found no significant differences between the three intervention groups in the amounts of training sessions, in either the group that completed the entire training or the group that did only partial training (both Fs<2.22). Furthermore, drop-out rates did not differ among the groups [χ^2(2)=2.62]

suggested that simply splitting the test into odd and even items yielded slightly imbalanced versions (with the even items being harder, on average), we created more balanced versions on the basis of the individual item performance from our previous studies ($N = 104$). Thus, Version A consisted of Items 3, 4, 5, 8, 9, 10, 15, 17, 18, 19, 20, 25, 27, 29, 31, 32, 33, and 34 of the original APM, whereas Version B consisted of Items 1, 2, 6, 7, 11, 12, 13, 14, 16, 21, 22, 23, 24, 26, 28, 30, 35, and 36. In contrast to our previous studies (Jaeggi et al., 2008; Jaeggi, Studer-Luethi, et al., 2010), we administered the test without any time restrictions. However, we note that using an untimed procedure with relatively few items increases the possibility that a considerable number of participants might perform at ceiling, reducing the possibility of detecting transfer (Jaeggi, Studer-Luethi, et al., 2010). The dependent variable consisted of the number of correct responses.

Cattell's Culture Fair Test (CFT; Cattell & Cattell, 1963) We used Scale 3, Forms A and B consisting of 100 items total (plus 22 practice items). Each version consists of four subtests; the tasks on the subtests include series, classification, matrices, and conditions (topology) (cf. Johnson, te Nijenhuis, & Bouchard, 2008). We took forms A and B and created three parts with an equal number of items from each subtest (8–10), also on the basis of performance on the individual items as obtained in our laboratory (cf. Supplementary Online Material). After completing two to three practice items for each subtest, participants worked on the remaining 34 items without any time restriction. The number of correct solutions served as the dependent variable.

Bochumer Matrizen Test (BOMAT; Hossiep, Turck, & Hasella, 1999) This task is similar to the APM, except that it is more difficult because it was developed for high-ability samples such as university students. The problems consist of a 5×3 matrix of patterns, and the missing pattern can occur in

any slot. The participant has six answer alternatives from which to choose. As in the CFT, we took both parallel versions of the original long version (80 items) and split the test into three equal parts (cf. Supplementary Online Material). After having completed the ten practice items, participants worked on the 27 test items without time restriction. The dependent variable was the number of correctly solved problems (note that the first item in each test version was considered a warm-up item and was not included in the analyses; thus, the maximum score was 26).

ETS Surface Development Test (Ekstrom, French, Harmon, & Derman, 1976; cf. Kane et al., 2004) In this test, participants are presented with a two-dimensional drawing of a piece of paper that, if folded along the given dotted lines, would result in a three-dimensional shape. Some of the edges of the unfolded paper are marked with letters and others with numbers. Participants are asked to determine which of the paper's lettered edges correspond to each of the shape's numbered edges. The test consists of six paper–shape pairs, and each pair has five numbered edges for which responses are required (yielding 30 responses). Following one practice item, participants were given 6 min to complete the test. Version A consisted of the odd items of the original ETS version, and Version B consisted of the even items. The dependent variable was the number of correct responses given within the time limit.

DAT Space Relations (Bennett, Seashore, & Wesman, 1972; cf. Kane et al., 2004) Here, participants are presented with outlines of patterns that can be folded into a three-dimensional object. From four alternatives, participants select the appropriate object into which the pattern can be folded. In this study, participants were given two practice items followed by 17 test items, Participants were allowed 5 min to complete the test. Version A consisted of the odd items of the original test (leaving out the very last item number 35), whereas Version B consisted of the even items of the original test. The dependent measure was the number of correctly solved items in the given time limit.

ETS Form Board Test (Ekstrom et al., 1976; cf. Kane et al., 2004) Each item consists of a set of five two-dimensional shapes that can be combined into a two-dimensional geometrical shape given at the top of the page. Participants indicate which of the pieces would contribute to the target figure by marking them with a plus sign, and they are also asked to mark unnecessary pieces with a minus sign. Six item sets corresponded to each target figure, and each test version included four target figures (cross, pentagon, square, and triangle). Version A consisted of Items 1–6, 19–24, as well as 31–42 of the original ETS test, whereas Version B consisted of Items 7–18, 25–30, as well as 43–49. After completing two practice item sets, participants were given 8 min to complete 24 items sets consisting of five shapes each, yielding 120

responses. The dependent variable was the number of correct responses given within the time limit.

Verbal reasoning tests

ETS Inferences Test (Ekstrom et al., 1976; cf. Kane et al., 2004) For each item, participants are presented with one or two brief written statements, and they are asked to decide which of five conclusions can be drawn from the statements without assuming any additional information. Following one sample item, participants were allowed 6 min to complete the test. Version A consisted of the odd items of the original ETS version (ten items), and Version B consisted of the even items (ten items). The dependent variable was the number of correct responses given within the time limit.

Air Force Officer Qualifying Test (AFOQT) Reading Comprehension Test (Berger, Gupta, Berger, & Skinner, 1990; Kane et al., 2004; Kane & Miyake, 2007) For each item, participants read a short two- to six-sentence paragraph, and are asked to complete the final sentence of the paragraph with one out of five answer alternatives. Each test version included ten test items, and participants were given 5 min to complete the test.

Verbal Analogies Test (based on Kane et al., 2004; Wright, Thompson, Ganis, Newcombe, & Kosslyn, 2008) In this test, participants are asked to compare relationships between two simultaneously presented word pairs that are on the left and right of the screen; that is, they must judge whether the relationship between the words in the left-hand pair is the same as the relationship between the words in the right-hand pair. In this study, participants responded by pressing the "1" key for "same" and the "0" key for "different" word pairs (example: few–many vs. noisy–quiet; answer: same). The relationship within word pairs varied to reflect synonyms, opposites, categories, function, or linear order. The task was self-paced; however, participants were required to respond within 8 s, and a 500-ms blank screen was presented between trials. After eight practice trials, participants completed 57 unique trials per session (48 items from Wright et al. 2008, and nine from Kane et al., 2004, adapted so they had the same format as the items from Wright et al.). The dependent variable was the proportion of correctly solved items.

Speed

Digit Symbol Test We used the digit–symbol coding test (DST) from the WAIS (Wechsler, 1997). It consists of the presentation of nine digit–symbol pairs, and participants have to fill in the corresponding symbol under a list of 130 digits as quickly as possible. Participants are given 90 s to complete as many items as possible. The dependent measure is the number of correct items completed in the time limit.

Questionnaires

Need for Cognition (NFC; Cacioppo & Petty, 1982) We used this questionnaire to assess how much participants enjoy cognitively challenging tasks. Statements such as "I really enjoy a task that involves coming up with new solutions to problems" were presented and participants were asked to indicate their level of agreement or disagreement on a 9-point Likert scale.

Theories of Cognitive Abilities (TOCA; Dweck, 1999) We assessed the degree to which participants think of intelligence as malleable or fixed. The questionnaire consists of eight statements, such as "You have a certain amount of cognitive ability and you can't really do much to change it," and participants indicate their agreement or disagreement on a 6-point Likert scale.

Cognitive Failure Questionnaire–Memory and Attention Lapses (CFQ-MAL, as used in McVay & Kane, 2009) This questionnaire was used to assess cognitive failures, and consists of a list of 40 questions, such as "Do you read something and find you haven't been thinking about it, so you have to read it again?" Responses are given on a 5-point Likert scale.

Training tasks

Dual N-back task We used the same training task as in previous studies (Jaeggi et al., 2008; Jaeggi, Studer-Luethi, et al., 2010). In short, participants had to process two streams of stimuli (auditory and visuospatial; eight stimuli per modality) that were synchronously presented at the rate of 3 s per stimulus. The task was to decide for each stream whether the present stimulus matched the one that was presented n items back in the series. The task was adaptive so that after each block of trials (= one round), the level of n was varied as a function of performance. If participants made fewer than three errors in both modalities, the level of n increased in the next round by one; if they made more than five errors in either modality, the level of n decreased in the next round by one; in all other cases, n remained the same. Each round included six targets per modality. Participants trained for 15 rounds in each session, each round consisting of $20 + n$ trials.

Single N-back task Participants trained on the auditory stream that was used in the dual n-back task. Six targets were presented per round, and everything else (timing, adaptivity, length, and amount of rounds) was the same as in the dual n-back task. Note that because we used just the auditory stream in this task, we consider it a verbal n-back training task with no spatial component.

Knowledge training task We used an adult variant of the knowledge training task described in Jaeggi et al. (2011a)

and used in Anguera et al. (2012). Participants solved GRE-type general knowledge, vocabulary questions, and trivia questions selected from a pool of approximately 5,000 questions. Each question was presented in the center of the screen, and participants chose one out of four answer alternatives presented below the question. After the participant's response, the correct answer was provided, occasionally along with some additional facts related to the question. Questions answered incorrectly were presented again in the beginning of the next session in order to evoke a learning experience.

Regardless of condition, each training session lasted approximately 20–30 min. After each session, participants rated how engaged they were during the session (responses were given on a Likert scale ranging from 1 to 9). Finally, participants were presented with a curve representing their performance (one point for each session) in relation to a generic curve that was derived from previous data collected in our laboratory.

Procedure

After obtaining informed written consent, participants underwent a first baseline assessment session, consisting of five reasoning tasks (Inferences, Surface Development, Verbal Analogies, APM, CFT; administered in that order). This session lasted approximately 90 min, and participants were allowed to take breaks between successive tests if they wished. Participants were asked to complete questionnaires administered online before coming in to the second session (NFC, TOCA, CFQ-MAL). In the second session, the remainder of the baseline assessments were administered (Space Relations, Reading Comprehension, Form Board, Digit Symbol, BOMAT; in this order); with all but the BOMAT being timed. This session typically lasted approximately 90 min, as well. After having completed all assessments, the participants who signed up for training received one of three intervention programs installed on their computers to train individually at home and underwent a few practice trials to ensure that they knew how their assigned training task worked. They were instructed to train once a day, five times per week for a total of 20 sessions. In order to increase and monitor compliance, participants were asked to e-mail their training data files to the lab after each session, and they received reminder e-mails if they failed to do so. After training completion, participants completed two posttest sessions with the same procedure as outlined above, except that they received parallel versions of the tests (counterbalanced between participants).[4] Finally, three months after training completion, participants came back to perform a subset of assessments (DST, CFT, BOMAT, in this order), and at the very end of the follow-up assessment, they completed a

---

[4] Note that the DST has no parallel test version.

training session on the training task that was relevant for them. This session typically lasted approximately 2 h.

Results

We first compared the four groups (participants who completed the training, drop-outs, participants who completed only the pretest and then withdrew, and participants who completed only the pretest and were paid for it) to determine whether differences would emerge among the groups at pretest. Indeed, we found differential group effects, most notably in all matrix reasoning tasks [APM: $F(3, 205) = 4.11$, $p = .007$, $\eta_p^2 = .06$; BOMAT: $F(3, 195) = 4.77$, $p = .003$, $\eta_p^2 = .07$; CFT: $F(3, 205) = 3.23$, $p = .02$, $\eta_p^2 = .05$], as well as in reading comprehension [AFOQT: $F(3, 196) = 3.27$, $p = .02$, $\eta_p^2 = .05$]. Furthermore, we found significant group differences in the need for cognition scale [NFC: $F(3, 192) = 2.75$, $p = .04$, $\eta_p^2 = .04$], as well as in self-reported cognitive failures [CFQ-MAL: $F(3, 192) = 3.48$, $p = .02$, $\eta_p^2 = .05$]. In general, the group that completed the training obtained the highest scores among all of the groups in the various cognitive measures at pretest, an effect that was most prominent in the matrix reasoning tasks (e.g., in a composite of all three matrix reasoning tasks, illustrated in Fig. 1): $p = .001$ (two-tailed), 95 % CI = [.14, .59] (planned contrast). This group also reported a higher need for cognition than the other three groups: $p = .009$ (two-tailed), 95 % CI = [.10, .67]; (planned contrast). Finally, all participants who signed up for training (including the participants who dropped out or withdrew) reported a higher amount of self-reported cognitive failures than did the participants who signed up for the baseline assessment only: $p = .002$ (two-tailed), 95 % CI = [–.97, –.23]; (planned contrast) (see Fig. 1; a detailed report is given in the Supplementary Online Material).[5]

Training data

The training performance for the three groups is illustrated in Fig. 2.

All training groups significantly improved their performance over the four weeks of training (all ps < .01). The largest improvements were observed in the single *n*-back group (83 %; from an average of *n*-back level 3.55 in the first two sessions to an *n*-back level of 6.40 in the last two sessions), followed by the dual *n*-back group (67 %; from an average of *n*-back level 2.62 in the first two sessions to an *n*-back level of 4.26 in the last two sessions), and the knowledge training group (44 %).

---

[5] Note that the comparisons reported in this paragraph contrasted the entire training group (i.e., including all three interventions) with the groups that had dropped out, withdrawn, or just completed the paid pretest. Within the training group, we found no significant intervention group differences for any of those variables (on either the cognitive measures or the questionnaires).
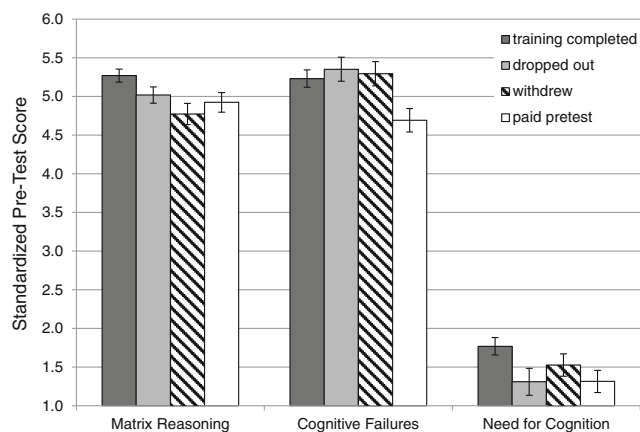
Fig. 1 Baseline assessment data for the four groups of participants (participants who completed the training; participants who completed part of the training, but dropped out before the posttest; participants who initially signed up for the training, but only completed the pretest and no more than three training sessions; and finally, participants who only completed a paid baseline assessment). For illustration purposes, the pretest and questionnaire scores are depicted as standardized measures—that is, each individual score divided by the standard deviation of the whole sample (see Table S1 for the individual scores). Error bars represent standard errors of the means

Transfer data

Descriptive data, as well as the test–retest reliabilities and effect sizes, for all transfer measures are reported in Table 2.

Preliminary analyses Since visuospatial and verbal reasoning abilities are assumed to be correlated, we conducted an exploratory factor analysis with an oblique rotation technique (cf. Fabrigar, Wegener, MacCallum, & Strahan, 1999) on the pretest measures for those participants who completed the training.[6] The analysis revealed two factors that explained 48.6 % of the total variance. The first factor was interpreted as verbal reasoning (represented by four measures accounting for 35.2 % of the variance), and the second factor was interpreted as visuospatial reasoning (represented by five measures accounting for 13.4 % of the variance); see Table S2. We then calculated composite scores consisting of the mean of the standardized gains for each of the measures going into the factors. Standardized gains were calculated as the gain (post minus pre) divided by the standard deviation of

---

[6] Note that by taking the whole sample into account ($N = 195$; this sample size reflects all participants who completed all of the baseline assessments, including the paid pretest sample), the same factor structure emerged, explaining 45 % of the total variance, although the APM loaded equally on both factors.

The DST was included in a first exploratory factor analysis and revealed equal, weak loadings on both factors (<.3). In the reported analysis, DST was not included, and was also discarded from all further analyses involving composite scores.

the whole sample at pretest for each measure (cf. Jaeggi et al., 2011a).

In order to test for potential selective attrition rates across intervention groups that might confound the outcome (i.e., transfer), we calculated multiple logistic regression models with different predictors using drop-out (yes vs. no) as the outcome. In each model, we tested the predictor-by-group interaction term. The specific predictors were pretest score for the visuospatial factor, pretest score for the verbal factor, gender, and age. We did not observe any significant interaction term in any of the models, suggesting that no confound of selective attrition with group was present (see also Table 1).

Overall transfer effects To assess transfer across all three groups (single *n*-back, dual *n*-back, and control), we conducted univariate analyses of covariance (ANCOVAs)[7] for both composite gain scores (with Intervention Type as a between-subjects factor and test version as a covariate), followed by planned contrasts. Note that no significant group differences were apparent at baseline (both *F*s < 0.82). Our analyses revealed no significant intervention effect for Factor 1 [Verbal Reasoning: $F(2, 73) = 0.01$, $p = .99$, $\eta_p^2 = .0004$]. In contrast, we found a significant intervention effect for Factor 2 [Visuospatial Reasoning: $F(2, 74) = 3.51$; $p = .035$, $\eta_p^2 = .09$]; see Fig. 3. Planned contrasts for Visuospatial Reasoning revealed that both *n*-back groups combined outperformed the knowledge training group in terms of performance gain from pre- to posttest: $p = .007$ (one-tailed), 95 % CI = [–.44, –.05]. Comparing the two training tasks, no difference was significant between the single *n*-back group and the dual *n*-back group: $p = .40$ (two-tailed), 95 % CI = [–.33, .13].

Next, we calculated separate ANCOVAs for both composite gain scores (with Intervention Type as a between-subjects factor and test version as a covariate), comparing the single *n*-back group with the control group, and comparing the dual *n*-back group with the control group. For the single *n*-back group versus control group comparison, our analyses revealed a significant intervention effect for Visuospatial Reasoning [$F(1, 50) = 7.20$, $p = .005$, $\eta_p^2 = .13$, one-tailed] and none for Verbal Reasoning [$F(1, 50) = 0.00$, $p = .50$, $\eta_p^2 = .000$, one-tailed]. For the dual *n*-back group versus the control group, we observed a significant intervention effect for Visuospatial Reasoning [$F(1, 49) = 3.07$; $p = .04$, $\eta_p^2 = .06$, one-tailed] and no intervention effect for Verbal Reasoning [$F(1, 48) = 0.02$; $p = .46$, $\eta_p^2 = .000$, one-tailed].

Moderators Next, we investigated whether any individual-difference variables could account for differential transfer effects. Need for cognition did not predict transfer. By

---

[7] Note that the *p* values remained significant if the data were analyzed using a repeated measures analysis of variance with Time (pretest, posttest) as a within-subjects factor and Group (single *n*-back, dual *n*-back, control) as a between-subjects factor.
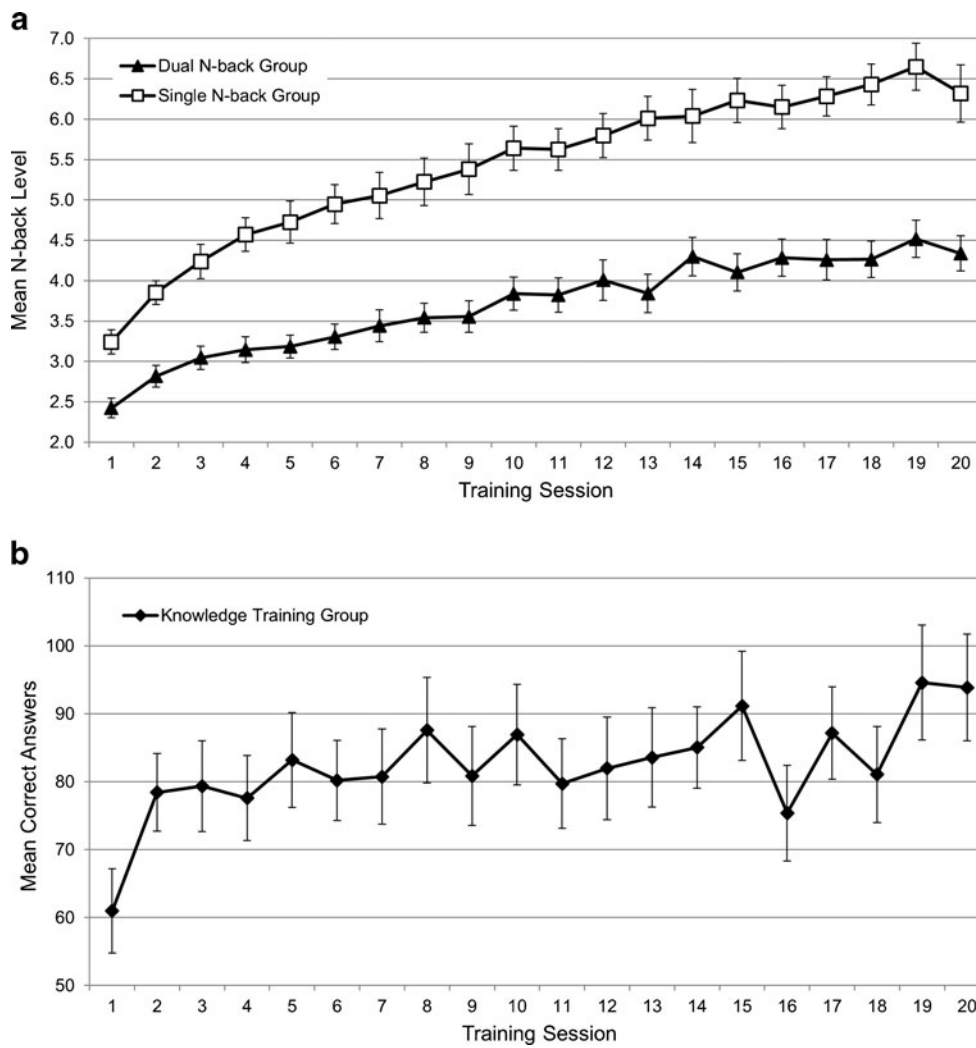
Fig. 2 Training performance for the participants who completed the training, illustrated separately for each intervention. The *y*-axes represent, (**a**) the mean *n*-back level achieved in each training session (*n*-back interventions) or (**b**) the average correct responses given (knowledge training intervention). Error bars represent standard errors of the means

contrast, we observed significant group differences for participants who indicated higher beliefs in the malleability of intelligence, who showed more transfer to visuospatial reasoning than did those who believed that intelligence is fixed [$t(74) = 2.17$, $p = .033$; see Fig. 4; groups were determined by median split]. Although the beliefs-by-intervention interaction was not significant ($F < 0.5$), the effect was most likely driven by the active control group, as this was the only group that showed a reliable correlation of beliefs and transfer ($r = .38$, $p < .05$), whereas the correlation of beliefs and transfer was negligible for both *n*-back interventions ($r < .06$, $p > .75$). Furthermore, the intervention effect to visuospatial reasoning remained significant after controlling for beliefs in intelligence assessed at pretest [$F(2, 71) = 3.33$, $p = .041$, $\eta_p^2 = .09$].

Long-term effects of training Analyses of the follow-up data revealed no differential group effects three months after training completion (by analyzing the gain from either pre or post

to follow-up; none of the analyses of variance were significant, all $p > .18$). However, we note that a considerable number of participants did not come back for the follow-up testing ($N = 24$; 31 %), resulting in a loss of power, which was further aggravated by the fact that the three intervention groups now noticeably differed in sample size. Numerically, the dual *n*-back group showed the largest retention effects, in terms of effect sizes in the CFT and speed (see Table 3).

Discussion

This study incorporated several methodological advances over previous WM training studies ,and nonetheless replicated transfer to measures of fluid intelligence (e.g., Jaeggi et al., 2008; Jaeggi, Studer-Luethi, et al., 2010). In particular, this study showed transfer to a composite score representing five visuospatial reasoning measures. Thus, transfer effects do not

Table 2 Descriptive data for the transfer measures as a function of group

	Pretest					Posttest					Pre vs. Post		
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max	p	r	ES
Dual N-back Training													
APM	25	15.32	2.63	8	18	25	14.96	2.70	8	18		.61	−.15
BOMAT	25	18.24	3.26	12	24	25	18.48	3.90	11	25		.68	.08
Surface Development	25	19.96	8.50	5	30	25	22.88	8.10	4	30	**	.80	.55
Space Relations	25	12.40	4.17	4	17	25	13.36	4.08	4	17		.73	.32
Form Board	25	61.92	26.20	0	107	25	75.68	25.20	25	123	***	.77	.79
CFT	25	19.84	3.95	11	27	25	20.56	3.25	14	26		.48	.19
Inferences	25	7.16	1.91	2	10	25	7.68	1.68	3	10		.31	.25
Reading Comprehension	25	7.52	2.02	4	10	25	7.60	2.57	1	10		.69	.04
Verbal Analogies	24	.74	.10	.55	.92	25	.75	.08	.56	.90		.40	.10
DST	24	67.88	13.26	40	88	25	70.36	12.77	42	91		.86	.36
Single N-back Training													
APM	26	14.88	2.30	7	18	26	15.23	2.44	10	18		.59	.20
BOMAT	26	17.31	3.03	11	23	26	18.96	2.84	13	23	*	.42	.52
Surface Development	26	19.96	8.34	2	30	26	21.85	8.30	7	30	*	.85	.41
Space Relations	26	12.54	3.97	3	17	26	14.27	3.24	7	17	**	.78	.70
Form Board	26	73.19	28.37	3	114	26	80.54	24.79	30	116		.61	.31
CFT	26	19.50	3.37	12	25	26	20.15	2.51	16	24		.45	.21
Inferences	26	8.00	1.79	3	10	26	8.04	1.59	4	10		.09	.02
Reading Comprehension	26	6.88	2.47	2	10	26	7.69	2.69	1	10		.59	.35
Verbal Analogies	26	.74	.09	.52	.88	26	.74	.10	.57	1.00		.26	.00
DST	25	72.80	12.15	54	100	26	75.88	11.09	53	98	**	.91	.61
Knowledge Training													
APM	27	14.81	2.79	7	18	27	14.74	2.80	8	18		.60	−.03
BOMAT	27	18.00	4.10	7	23	27	17.63	3.78	11	26		.72	−.12
Surface Development	27	21.41	8.23	1	30	27	22.56	7.65	0	30		.84	.25
Space Relations	27	13.48	3.13	7	17	27	13.81	3.11	4	17		.43	.10
Form Board	27	67.41	27.26	5	115	27	66.63	24.49	22	117		.73	−.04
CFT	27	19.63	3.12	14	26	27	19.63	3.12	12	24		.44	.00
Inferences	27	7.22	1.91	2	10	27	7.96	1.89	2	10	*	.66	.47
Reading Comprehension	27	6.04	2.77	0	10	27	5.85	2.78	1	10		.74	−.09
Verbal Analogies	27	.72	.10	.54	.87	27	.75	.09	.56	.90		.29	.26
DST	27	68.67	11.26	37	91	27	71.85	11.41	57	93	*	.77	.41

r = test–retest reliability (partial correlations accounting for test version); ES = effect size that accounts for the correlation between the pre- and posttest measures, $\mu_2 - \mu_1 / \sqrt{\sigma_1^2 + \sigma_2^2 - 2r_{12}\sigma_1\sigma_2}$ $^*p < .05$; $^{**}p < .01$; $^{***}p < .001$

seem to be restricted to a specific task such as the BOMAT; rather, they seem to be more general, in that they emerged with respect to a visuospatial reasoning factor that did not consist of matrix reasoning tasks alone. Second, this transfer was observed despite the use of an active control group that trained on a knowledge-based task (which showed no improvements in visuospatial reasoning).

In addition to replicating previous research on WM training using the n-back task with several methodological improvements, this study also went beyond previous research by assessing the breadth of transfer as well as other factors that might have determined the outcome. Of particular interest is that transfer to visuospatial reasoning emerged as a function of auditory–verbal n-back training; that is, it emerged as a function of a training task that did not involve stimuli that were visuospatial at all, indicating that the n-back training effect on visuospatial reasoning is modality independent. Thus, it is likely that both n-back versions share similar underlying processes that drive these effects (cf. Jaeggi, Studer-Luethi, et al., 2010). Candidate processes might include updating, and
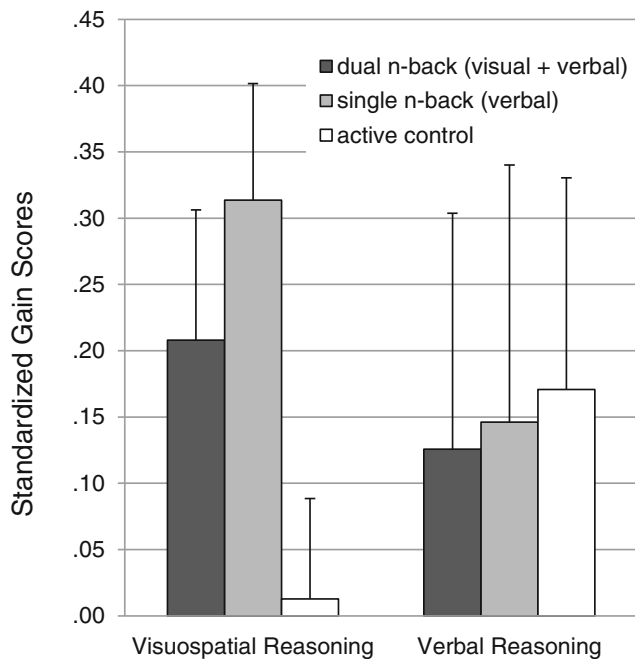
Fig. 3 Transfer effects in the factor-analytically derived measures of visuospatial and verbal reasoning. The scores for each intervention group are given as standardized gain scores—that is, each individual score divided by the standard deviation of the pretest score of the whole sample (see Table 2 for the individual scores). Error bars represent standard errors of the means

especially discriminating between relevant and irrelevant stimuli, which was essential in our versions of the n-back tasks because they contained a considerable number of lures (e.g., a two-back match when a participant was engaged in a three-back task). Discriminating between targets and nontargets might also be an important process that was required by almost all of our transfer tasks, because the answer alternatives usually contained features that were close to the solution but that lacked one or two important details (Wiley & Jarosz, 2012).
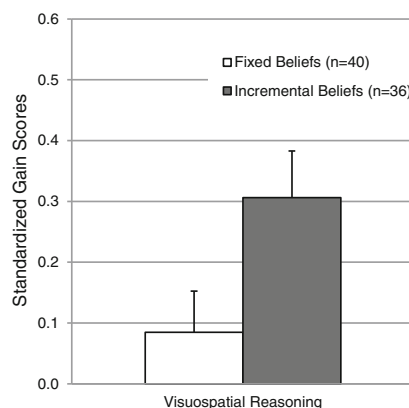

Fig. 4 Transfer effects to visuospatial reasoning as a function of self-reported beliefs in theories about intelligence, irrespective of training intervention. Transfer effects are illustrated by standardized gain scores, and error bars represent standard errors of the means

Unfortunately, we can only speculate that interference resolution is the process driving the effect, since we did not directly assess any measures of interference resolution in the present study. Nevertheless, in other studies, we and others have shown evidence for transfer from WM training to measures in which efficient discrimination between relevant and irrelevant targets was crucial, suggesting that processes such as inhibition and interference resolution share overlapping resources with WM (Jaeggi, Buschkuehl, Jonides, & Shah, 2011b; Klingberg et al., 2005; Klingberg et al., 2002).

Interestingly, although participants showed small retest effects in the verbal reasoning factor, we found no differential group effects, which could suggest that transfer might be restricted to the visuospatial domain. However, the data are inconclusive, as most measures contributing to the verbal reasoning factor turned out to show weaker test-retest reliability than did those of the visuospatial factor (see Table 3); thus, it might be that reliability issues masked some of the transfer effects. However, we acknowledge that our test–retest correlations were lower-bound estimates of reliability, because they might have been reduced by the intervention, and furthermore, the correlations between tasks at pretest (cf. Table S3) suggest that the actual reliability of the verbal reasoning tasks could be higher. Nonetheless, it is important to note that, in general, the lower the reliability, the lower the chances for transfer—that is, effect size (cf. Fig. S1). Furthermore, the reliabilities of the verbal tasks were overall significantly lower than the reliabilities of the spatial tasks [average verbal ($r = .45$) vs. average visuospatial ($r = .68$); $t(25) = 3.59$, $p = .001$], and as such, the verbal tasks might suffer from more error variance than the spatial tasks. Poor reliability might be a common problem in other studies assessing changes in Gf: Very few fluid reasoning measures come with reliable parallel test versions, and no measures have three versions that could have been used for the three assessment times used here (pre, post, and follow-up). The commonly used method of splitting the tests in half (or even thirds) to present distinct items at pretest, posttest, and perhaps follow-up has the disadvantage of reducing the reliability and validity of each test and reducing the range of values in the scoring, due to the restricted range of items available. This might also decrease sensitivity, which might contribute to the null effects reported by others (e.g. Redick et al., 2013). A solution for future studies may be to use computer-generated tasks that have been developed by some researchers in recent years, providing a virtually infinite number of usable items (e.g., Arendasy & Sommer, 2005; Freund & Hotting, 2011; Matzen et al., 2010). However, the construct validity of these computer-generated tasks is still largely unresolved; additionally, very few tasks are available for which such an approach would be possible, and those tasks are currently restricted to the visuospatial domain.

Of course, alternative explanations might account for the lack of transfer to verbal abilities. One possible explanation, proposed

Table 3 Descriptive data and effect sizes for those participants who completed all three assessments

	Pre		Post		Follow-Up		Pre vs. Follow-Up			Post vs. Follow-Up		
	Mean	SD	Mean	SD	Mean	SD	p	r	ES	p	r	ES
Dual N-Back												
BOMAT	17.71	3.16	18.29	4.10	18.76	4.98		.61	.27		.48	.10
CFT	19.59	3.66	20.71	3.20	21.35	3.10	*	.30	.44		.83	.35
DST	64.25	13.64	68.00	14.16	73.18	15.04	***	.89	1.30	***	.94	1.01
Single N-Back												
BOMAT	17.36	3.10	19.21	2.89	18.14	4.05		.73	.28		.62	−.33
CFT	20.36	3.71	20.29	2.43	21.29	3.24		.62	.30		.43	.32
DST	72.29	12.13	75.36	11.32	79.29	8.82	***	.86	1.09	.058	.78	.55
Knowledge Training												
BOMAT	17.74	4.31	17.13	3.82	17.78	4.36		.82	.02		.76	.23
CFT	19.30	3.07	19.57	3.26	20.13	2.75		.38	.26		.27	.15
DST	68.91	12.17	71.83	11.51	73.39	12.76	***	.89	.76		.76	.18

Dual N-back group: $N=17$ (except for DST pre, where $N=16$). Single N-back group: $N=14$. Knowledge Training: $N=23$. r = retest reliability (partial correlations accounting for test version); ES = effect size, as Cohen's d. * $p<.05$; *** $p<.001$.

by Miyake, Friedman, Rettinger, Shah, and Hegarty (2001), is that people generally have less practice on spatial than on verbal tasks, and thus may have more room for improvement on spatial tasks. An additional explanation is that some of the verbal ability tasks taxed both crystallized and fluid abilities; thus, improvement in only fluid abilities might have been be less telling for transfer. Interestingly, greater transfer of training to the spatial than to the verbal domain has been found in other labs, as well as in other studies from our own research group (Buschkuehl et al., 2008; Klauer, Willmes, & Phye, 2002; Rueda, Rothbart, McCandliss, Saccomanno, & Posner, 2005). We realize that these possibilities are speculative and that further research will be necessary to clarify the mechanisms of the differential effect observed here.

Of interest is the finding that belief in the malleability of intelligence affects the degree of transfer (Fig. 4). In particular, individuals who held beliefs about the malleability of intelligence showed greater improvement on the visuospatial reasoning factor than did those who held fixed beliefs about intelligence. This effect was driven primarily by the active control group. That is, individuals in the active control group showed a placebo effect.[8] However, it is important to emphasize that the training-by-session interaction for visuospatial reasoning was still significant when initial beliefs about intelligence were controlled. Thus, the placebo effect was in

addition to the effect of training condition. Our finding of the influence of theories of intelligence may actually provide some insight into why some studies may find transfer whereas others do not, if by chance an active control group included a greater number of participants who had beliefs in the malleability of abilities. This finding highlights the importance of using an active control group, and also of assessing motivation and beliefs when conducting training studies.

An additional important result of the present study is that we observed no significant retention at the three-month follow-up. Although a few reports have shown long-term effects after cognitive training (Borella et al., 2010; Carretti et al., 2013; Jaeggi et al., 2011a; Klingberg et al., 2005; Van der Molen et al., 2010), other studies have failed to show such effects (Buschkuehl et al., 2008). If we consider WM training as being analogical to cardiovascular training, occasional practice or booster sessions may be needed in order to maximize retention (e.g., Ball, Berch, Helmers, Jobe, Leveck, Marsiske, & Willis, 2002; Bell, Harless, Higa, Bjork, Bjork, Bazargan, & Mangione, 2008; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Haskell, Lee, Pate, Powell, Blair, Franklin, & Bauman 2007; Whisman, 1990). Unfortunately, at the current stage of knowledge, we do not have a good sense of how often such booster sessions would have to take place. Ultimately, the definition of factors that promote the longevity of training and the investigation of how cognitive training affects real-life outcomes would be essential from an applied point of view; yet, overall, we remain skeptical about the interpretation of our follow-up data, because of the small and uneven sample sizes.

One potential concern about the present study is that, although we found significant transfer for the visuospatial-

---

[8] A slightly different interpretation is that those with malleable mindsets are better learners in both the active control and intervention groups, and thus improve on transfer tasks regardless of the intervention; this would also yield a placebo-like effect.

Engagement
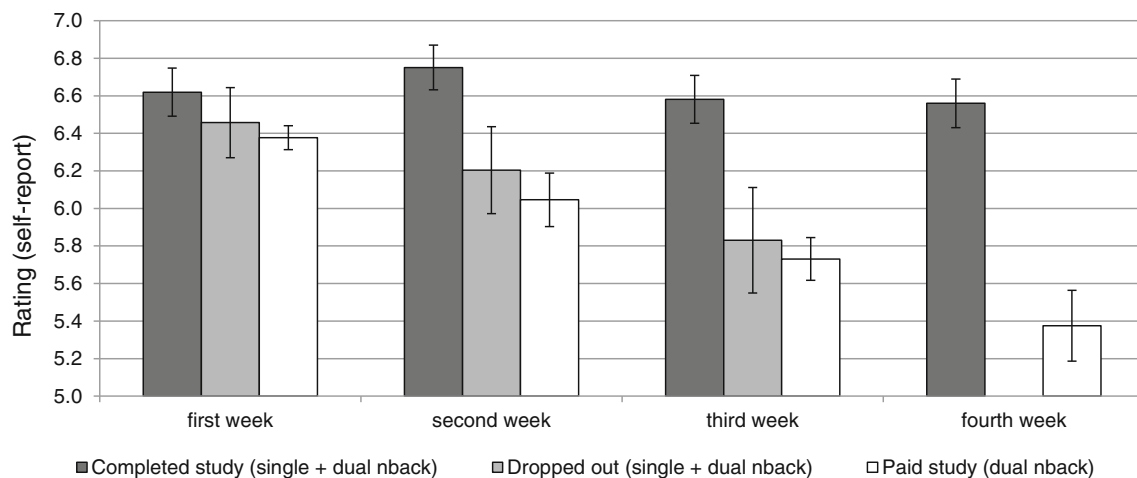### (assessed after each training session)

Fig. 5 Self-reported engagement ratings indicated after each training session, averaged for each of the four weeks of training. Depicted are the ratings for the two unpaid n-back groups from the present study [dark gray bars, completed training ($N=48$); light gray bars, dropped out ($N=$ 27)], as well as those of the paid dual n-back group from Anguera et al. (2012; white bars, $N=26$). Error bars are standard errors of the means. Note that, due to the very small N in the fourth week for drop-outs, data are not shown

task composite measure, the effects did not emerge on each individual measure (Table 2). The pretest data in Table 2, however, indicate that performance was near ceiling at pretest for some tasks. Furthermore, some tasks had relatively low levels of reliability. The use of a composite score based on factor analysis has the benefit of reducing the effect of such measurement problems. Indeed, one possible explanation for the null findings of cognitive training in some studies (e.g., Redick et al., 2013) is the undue influence of such measurement issues. Another potential reason for the relatively lower effect sizes, as compared to our earlier studies, could be the somewhat limited training gain observed in the present sample. As we pointed out in the introduction, participants in our successful studies improved over 2 n-back levels in the dual n-back task, whereas the present improvement was slightly below that (1.8 levels).[9] This apparent discrepancy might have occurred for a reason, though: Unlike in our previous studies, the participants in the present study trained at home in an unsupervised and uncontrolled environment, which most likely had an impact on training fidelity, even though the participants were intrinsically motivated in general (see below). For example, we could not control whether the training took place in an undistracted environment, and training times varied

greatly and included quite a few late-night sessions. Thus, the magnitude of task-specific improvement seems to be an important factor contributing to transfer (Jaeggi et al., 2011a). Finally, of additional potential concern could be the numerically different drop-out rates for the n-back and knowledge trainer groups (25 % vs. about 40 %). Note, however, that although numerical differences were present, these differences were not significant (cf. Table 1). Nevertheless, it is possible that individuals in the knowledge trainer group enjoyed the intervention regardless of whether they felt that they were improving, and thus continued to participate. By contrast, individuals in the n-back group may have become frustrated due to lack of improvement, and thus dropped out. This interpretation is highly speculative, but future research in which participants are interviewed about why they dropped out of the study might shed light on this issue.

All of these considerations aside, the present results add to our previous demonstration of a far-transfer effect as a consequence of WM training with different populations and age groups (Buschkuehl, Jaeggi, Hutchison, Perrig-Chiello, Dapp, Muller, & Perrig, 2008; Jaeggi et al., 2008; Jaeggi et al., 2011a; Jaeggi, Studer-Luethi, et al., 2010; Loosli et al., 2012). We have also provided testable hypotheses to investigate why results sometimes conflict, in that researchers sometimes find transfer to measures of fluid intelligence, and sometimes not. Our work also highlights intrinsic motivation as an important factor (Burton et al., 2006; Deci et al., 1999). For example, whereas participants who completed the study reported engagement levels that remained relatively stable throughout the four weeks of training, participants who did

---

[9] In the previous single n-back intervention, the improvement was 3.6 n-back levels (Jaeggi et al., 2010) versus 3.0 levels in the present study. However, note that the previous sample trained on a visuospatial version, whereas the present sample trained on an auditory version; thus, the improvements might not be entirely comparable.

not complete the study reported engagement levels that gradually decreased over the course of the intervention period—a decrease that was comparable with the self-reports from our earlier paid participants (cf. Anguera et al. 2012; see Fig. 5). Furthermore, we found a modest but reliable correlation between self-reported engagement and training gain ($r = .27, p < .05$), which was especially pronounced in the group that dropped out of the training ($r = .41, p < .05$). Thus, whereas the participants in our earlier paid experiment did complete the study, presumably because they were paid for completion, they might not have shown as much transfer due to a lack of intrinsic motivation. In contrast, in the present study, the participants who lost interest in the training might simply have dropped out, as they had no incentive to complete the training and come back for the posttest.

This leads us to the important issue of who actually signs up for such a study, and who sticks to an intervention over the course of a full four-week period: Our data show that participants who signed up to participate in the training study reported more cognitive failures than did those participants who completed just the pretest, without the intention to train (Fig. 1). That is, the participants who signed up for training seem to have had some self-perceived deficit that may have influenced their interest in improving their memory and cognitive performance in the first place. Interestingly, although these participants reported cognitive failures, they did not perform worse on the baseline tests administered in the present study. At the same time, the participants with the highest pretest scores combined with the highest need-for-cognition scores ended up being the ones who actually completed the training. Thus, a combination of high intelligence paired with self-perceived cognitive failures and high need for cognition seems to constitute the kind of person who is motivated and shows consistent engagement to complete a training study; this is a combination of personality traits that might also be related to persistence to engage with a regimen that might not always be fun and interesting. This trait has been termed "grit" (Duckworth, Peterson, Matthews, & Kelly, 2007).

On the basis of the pattern of traits discussed above, it may be that the individuals who completed our study were not the individuals who needed cognitive training the most. This suggests that the challenge in cognitive intervention research will be to get the training to those individuals who might show greater profit from cognitive improvement and to keep them engaged throughout the whole intervention period. In our work with children, we have improved our n-back regimen in order to make it more fun and interesting by borrowing features that are known from the video game literature to enhance motivation (see Jaeggi et al., 2011a). It might well be that we need to optimize our adult versions in a similar way, in order to capture and keep participants who lack intrinsic motivation otherwise.

Conclusion

To conclude, this study has served to replicate earlier findings of transfer from training using the n-back task, even after resolving potential methodological concerns that have been cited as problems in interpreting previous training studies. We observed transfer to a composite score of visuospatial reasoning consisting of various measures, providing evidence for broader generalization effects than we have demonstrated in our previous studies. Interestingly, the transfer effects seem to be restricted to the visuospatial domain; however, we temper this conclusion with the observation that the verbal tasks may have had poorer reliability. More importantly, the processes underlying n-back training seem to be domain-free, in that training on a verbal n-back task resulted in transfer to measures of visuospatial reasoning. Finally, boundary conditions seemed to modulate the effect of training. Intrinsic motivation, preexisting ability, and the need for cognition affect whether or not one chooses to participate in a cognitive training intervention and to stick with the training regimen. In addition, beliefs in the malleability of intelligence moderate transfer effects. All of these issues require further exploration, to further delimit when cognitive training will yield effective benefits to other cognitive skills.

References

Anguera, J. A., Bernard, J. A., Jaeggi, S. M., Buschkuehl, M., Benson, B. L., Jennett, S., & Seidler, R. D. (2012). The effects of working memory resource depletion and training on sensorimotor adaptation. *Behavioural Brain Research, 228,* 107–115.

Arendasy, M., & Sommer, M. (2005). The effect of different types of perceptual manipulations on the dimensionality of automatically generated figural matrices. *Intelligence, 33,* 307–324. doi:10.1016/j.intell.2005.02.002

Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., & Willis, S. L. (2002). Effects of cognitive training interventions with older adults: A randomized controlled trial. *JAMA: Journal of the American Medical Association, 288,* 2271–2281.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128,* 612–637.

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.

Bell, D. S., Harless, C. E., Higa, J. K., Bjork, E. L., Bjork, R. A., Bazargan, M., & Mangione, C. M. (2008). Knowledge retention after an online tutorial: A randomized educational experiment among resident physicians. *Journal of General Internal Medicine, 23,* 1164–1171.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1972). *Differential Aptitude Test: Space relations*. New York: Psychological Corp.

Berger, F. R., Gupta, W. B., Berger, R. M., & Skinner, J. (1990). *Air Force Officer Qualifying Test (AFOQT) Form P: Test manual (AFHRL-TR-89-56)*. Brooks Air Force Base, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Bergman Nutley, S., Soderqvist, S., Bryde, S., Thorell, L. B., Humphreys, K., & Klingberg, T. (2011). Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: A controlled, randomized study. *Developmental Science, 14,* 591–601.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78,* 246–263. doi:10.1111/j.1467-8624.2007.00995.x

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

Burton, K. D., Lydon, J. E., D'Alessandro, D. U., & Koestner, R. (2006). The differential effects of intrinsic and identified motivation on well-being and performance: Prospective, experimental, and implicit approaches to self-determination theory. *Journal of Personality and Social Psychology, 91,* 750–762.

Buschkuehl, M., & Jaeggi, S. M. (2010). Improving intelligence: A literature review. *Swiss Medical Weekly, 140,* 266–272.

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

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42,* 116–131.

Carretti, B., Borella, E., Zavagnin, M., & de Beni, R. (2013). Gains in language comprehension relating to working memory training in healthy older adults. *International Journal of Geriatric Psychiatry, 28,* 539–546. doi:10.1002/gps.3859

Cattell, R. B., & Cattell, A. K. S. (1963). *Test of "g": Culture Fair Scale 3*. Champaign: Institute for Personality and Ability Testing.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132,* 354–380. doi:10.1037/0033-2909.132.3.354

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

Craik, F. I. M., Winocur, G., Palmer, H., Binns, M. A., Edwards, M., Bridges, K., & Stuss, D. T. (2007). Cognitive rehabilitation in the elderly: Effects on memory. *Journal of the International Neuropsychological Society, 13,* 132–142. doi:10.1017/S1355617707070166

Dahlin, E., Neely, A. S., Larsson, A., Backman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science, 320,* 1510–1512.

De Lisi, R., & Wolford, J. L. (2002). Improving children's mental rotation accuracy with computer game playing. *Journal of Genetic Psychology, 163,* 272–282. doi:10.1080/00221320209598683

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125,* 627–668. doi:10.1037/0033-2909.125.6.627. disc. 692–700.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92,* 1087–1101. doi:10.1037/0022-3514.92.6.1087

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Philadelphia: Psychology Press.

Ekstrom, R. B., French, J. W., Harmon, H. H., & Derman, D. (1976). *ETS Kit of Factor-Referenced Cognitive Tests*. Princeton: Educational Testing Service.

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: Cambridge University Press.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272–299.

Freund, P. A., & Hotting, H. (2011). How to get really smart: Modeling retest and training effects in ability testing using computer-generated figural matrix items. *Intelligence, 39,* 233–243. doi:10.1016/j.intell.2011.02.009

García-Madruga, J. A., Elosúa, M. R., Gil, L., Gómez-Veiga, I., Vila, J. Ó., Orjales, I., & Duque, G. (2013). Reading comprehension and working memory's executive processes: An intervention study in primary school students. *Reading Research Quarterly, 48,* 155–174.

Grant, H., & Dweck, C. S. (2003). Clarifying achievement goals and their impact. *Journal of Personality and Social Psychology, 85,* 541–553.

Haskell, W. L., Lee, I. M., Pate, R. R., Powell, K. E., Blair, S. N., Franklin, B. A., & Bauman, A. (2007). Physical activity and public health—Updated recommendation for adults from the American college of sports medicine and the American heart association. *Circulation, 116,* 1081–1093. doi:10.1161/Circulationaha.107.185649

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

Holmes, J., Gathercole, S. E., Place, M., Dunning, D. L., Hilton, K. A., & Elliott, J. G. (2010). Working memory deficits can be overcome: Impacts of training and medication on working memory in children with ADHD. *Applied Cognitive Psychology, 24,* 827–836.

Hong, Y. Y., Chiu, C. Y., Dweck, C. S., Lin, D. M. S., & Wan, W. (1999). Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality and Social Psychology, 77,* 588–599.

Hossiep, R., Turck, D., & Hasella, M. (1999). *Bochumer Matrizentest: BOMAT advanced*. Göttingen: 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:10.1073/pnas.0801268105

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011a). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences, 108,* 10081–10086. doi:10.1073/pnas.1103228108

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011b). *Working memory training in typically developing children and children with Attention Deficit Hyperactivity Disorder: Evidence for plasticity in executive control processes.* Paper presented at the Eighteenth Annual Cognitive Neuroscience Society Meeting, San Francisco, CA.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the *N*-back task as a working memory measure. *Memory, 18,* 394–412.

Jaeggi, S. M., Seewer, R., Nirkko, A. C., Eckstein, D., Schroth, G., Groner, R., & Gutbrod, K. (2003). Does excessive memory load attenuate activation in the prefrontal cortex? Load-dependent processing in single and dual tasks: Functional magnetic resonance imaging study. *NeuroImage, 19,* 210–225.

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.

Jausovec, N., & Jausovec, K. (2012). Working memory training: Improving intelligence—Changing brain activity. *Brain and Cognition, 79,* 96–106. doi:10.1016/j.bandc.2012.02.007

Johnson, W., te Nijenhuis, J., & Bouchard, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence, 36,* 81–95. doi:10.1016/j.intell.2007.06.001

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology, 59,* 193–224. doi:10.1146/annurev.psych.59.103006.093615

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

Kane, M. J., & Miyake, T. M. (2007). The validity of "conceptual span" as a measure of working memory capacity. *Memory & Cognition, 35,* 1136–1150. doi:10.3758/BF03193484

Klauer, K. J., Willmes, K., & Phye, G. D. (2002). Inducing inductive reasoning: Does it transfer to fluid intelligence? *Contemporary Educational Psychology, 27*, 1-25.

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlstrom, 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.

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

Kundu, B., Sutterer, D. W., Emrich, S. M., & Postle, B. R. (2013). Strengthened effective connectivity underlies transfer of working memory training to tests of short-term memory and attention. *Journal of Neuroscience, 33,* 8705–8715. doi:10.1523/JNEUROSCI.5565-12.2013

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

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

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

Matzen, L. E., Benz, Z. O., Dixon, K. R., Posey, J., Kroger, J. K., & Speed, A. E. (2010). Recreating Raven's: Software for systematically generating large numbers of Raven-like matrix problems with normed properties. *Behavior Research Methods, 42,* 525–541. doi:10.3758/BRM.42.2.525

Mayo, E. (1933). *The human problems of an industrial civilization*. New York: Macmillan.

McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 196–204. doi:10.1037/a0014104

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130,* 621–640. doi:10.1037/0096-3445.130.3.621

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

Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality and Social Psychology, 75,* 33–52.

Nystrom, L. E., Braver, T. S., Sabb, F. W., Delgado, M. R., Noll, D. C., & Cohen, J. D. (2000). Working memory for letters, shapes, and locations: fMRI evidence against stimulus-based regional organization in human prefrontal cortex. *NeuroImage, 11,* 424–446.

Okagaki, L., & Frensch, P. A. (1994). Effects of video game playing on measures of spatial performance: Gender effects in late adolescence. *Journal of Applied Developmental Psychology, 15,* 33–58.

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,* 775–778.

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

Pickering, S. (Ed.). (2006). *Working memory and education*. Oxford: Elsevier.

Rabipour, S., & Raz, A. (2012). Training the brain: Fact and fad in cognitive and behavioral remediation. *Brain and Cognition, 79,* 159–179. doi:10.1016/j.bandc.2012.02.006

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

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D., Hambrick, D. Z., & Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General, 142,* 359–379. doi:10.1037/a0029082

Rudebeck, S. R., Bor, D., Ormond, A., O'Reilly, J. X., & Lee, A. C. (2012). A potential spatial working memory training task to improve both episodic memory and fluid intelligence. *PLoS ONE, 7,* e50431. doi:10.1371/journal.pone.0050431

Rueda, M. R., Rothbart, M. K., McCandliss, B. D., Saccomanno, L., Posner, M. I. (2005). Training, maturation, and genetic influences on the development of executive attention. *Proceedings of the National Academy of Sciences of the United States of America, 102*(41), 14931-14936.

Salminen, T., Strobach, T., & Schubert, T. (2012). On the impacts of working memory training on executive functioning. *Frontiers in Human Neuroscience, 6,* 166. doi:10.3389/fnhum.2012.00166

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

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

Seidler, R. D., Bernard, J. A., Buschkuehl, M., Jaeggi, S., Jonides, J., & Humfleet, J. (2010). *Cognitive training as an intervention to improve driving ability in the older adult (Technical Report No. M-CASTL 2010-01)*. Ann Arbor: University of Michigan.

Shah, P., Buschkuehl, M., Jaeggi, S. M., & Jonides, J. (2012). Cognitive Training for ADHD: The Importance of Individual Differences. *Journal of Applied Research in Memory and Cognition, 1,* 204–205.

Shah, P., & Miyake, A. (1999). Models of working memory: An introduction. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanism of active maintenance and executive control* (pp. 1–26). New York: Cambridge University Press.

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

Stephenson, C. L., & Halpern, D. F. (2013). Improved matrix reasoning is limited to training on tasks with a visuospatial component. *Intelligence, 41,* 341–357.

Studer-Luethi, B., Jaeggi, S. M., Buschkuehl, M., & Perrig, W. J. (2012). Influence of neurotisicm and conscientiousness on working memory

training outcome. *Personality and Individual Differences, 53*(1), 44-49. doi:10.1016/j.paid.2012.02.012

Takeuchi, H., Taki, Y., Nouchi, R., Hashizume, H., Sekiguchi, A., Kotozaki, Y., & Kawashima, R. (2013). Effects of working memory training on functional connectivity and cerebral blood flow during rest. *Cortex, 49,* 2106–2125. doi:10.1016/j.cortex.2012.09.007

Terlecki, M. S., Newcombe, N. S., & Little, M. (2008). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Cognitive Psychology, 22,* 996–1013. doi:10.1002/Acp.1420

Thompson, T. W., Waskom, M. L., Garel, K. L., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., & Gabrieli, J. D. (2013). Failure of working memory training to enhance cognition or intelligence. *PLoS ONE, 8,* e63614. doi:10.1371/journal.pone.0063614

Thorell, L. B., Lindqvist, S., Bergman Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science, 12,* 106–113.

Tomic, W., & Klauer, K. J. (1996). On the effects of training inductive reasoning: How far does it transfer and how long do the effects persist? *European Journal of Psychology of Education, 11,* 283–299.

Van der Molen, M. J., Van Luit, J. E. H., Van der Molen, M. W., Klugkist, I., & Jongmans, M. J. (2010). Effectiveness of a computerised working memory training in adolescents with mild to borderline intellectual disabilities. *Journal of Intellectual Disability Research, 54,* 433–447.

Verhaeghen, P., Marcoen, A., & Goossens, L. (1992). Improving memory performance in the aged through mnemonic training: A meta-analytic study. *Psychology and Aging, 7,* 242–251. doi:10.1037/0882-7974.7.2.242

von Bastian, C. C., & Oberauer, K. (2013). Distinct transfer effects of training different facets of working memory capacity. *Journal of Memory and Language, 69,* 36–58. doi:10.1016/j.jml.2013.02.002

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale* (3rd ed.). San Antonio: Psychological Corp.

Whisman, M. A. (1990). The efficacy of booster maintenance sessions in behavior therapy: Review and methodological critique. *Clinical Psychology Review, 10,* 155–170.

Wiley, J., & Jarosz, A. F. (2012). How working memory capacity affects problem solving. *Psychology of Learning and Motivation, 56,* 185–227.

Witt, M. (2011). School based working memory training: Preliminary finding of improvement in children's mathematical performance. *Advances in Cognitive Psychology, 7,* 7–15. doi:10.2478/v10053-008-0083-3

Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training generalized spatial skills. *Psychonomic Bulletin & Review, 15,* 763–771. doi:10.3758/PBR.15.4.763

Zelinski, E. M. (2009). Far transfer in cognitive training of older adults. *Restorative Neurology and Neuroscience, 27,* 455–471.

Zhao, X., Wang, Y. X., Liu, D. W., & Zhou, R. L. (2011). Effect of updating training on fluid intelligence in children. *Chinese Science Bulletin, 56,* 2202–2205.

Zinke, K., Zeintl, M., Eschen, A., Herzog, C., & Kliegel, M. (2011). Potentials and limits of plasticity induced by working memory training in old-old age. *Gerontology, 58,* 79–87. doi:10.1159/000324240

Zinke, K., Zeintl, M., Rose, N. S., Putzmann, J., Pydde, A., & Kliegel, M. (2013). Working memory training and transfer in older adults: Effects of age, baseline performance, and training gains. *Developmental Psychology*. doi:10.1037/a0032982

Supplementary Online Material for

The Role of Individual Differences in Cognitive Training and Transfer

Susanne M. Jaeggi*, Martin Buschkuehl, Priti Shah, and John Jonides

*_Author for correspondence. E-mail:_ smjaeggi@uci.edu

Items per version for the BOMAT

Version A:
From original Form A: Item numbers 1 (warm-up item), 3, 10, 14, 15, 28, 33, 34
From original Form B: Item numbers 2, 4, 5, 6, 8, 11, 16, 18, 19, 20, 25, 27, 28, 29, 30, 34, 37, 38, 39

Version B:
From original Form A: Item numbers 2, 4, 5, 8, 9, 13, 17, 18, 20, 21, 23, 29, 31, 35, 36, 38, 40
From original Form B: Item numbers 1 (warm-up item), 7, 10, 14, 15, 21, 23, 24, 31, 36

Version C:
From original Form A: 6, 7, 11, 12, 16, 26, 27, 25, 19, 24, 22, 32, 30, 37, 39
From original Form B: 3, 17, 13, 9, 12, 22, 26, 35, 33, 32, 40

In addition: warm-up item (created by us analogous to Item 1 in the original forms)

Items per version for the CFT

Version A:
Subtest 1: Items 1, 3, 5, 8, 9, 11, 12, 13 (all from original Form A).
Subtest 2: Items 1, 2, 4, 5, 7, 9, 11, 13 (all from original Form A).
Subtest 3: Items 2, 6, 8, 9, 10, 11, 13 (from original Form A), and Item 2 from original Form B
Subtest 4: Items 2, 4, 5, 8, 9 (from original Form A), and Item 9 from original Form B

Version B:
Subtest 1: Items 3, 5, 6, 7, 9, 11, 13 (from original Form B), and Item 10 from original Form A
Subtest 2: Items 1, 5, 6, 7, 9, 10, 12, 13 (all from original Form B)
Subtest 3: Items 4, 6, 7, 8, 10, 12, 13 (from original Form B), and Item 7 from original Form A
Subtest 4: Items 1, 4, 5, 8, 10 (from original Form B), and Item 7 from original Form A

Version C:
Subtest 1: Items 4, 6, 7, 12 from original Form A, and Items 1, 4, 8, 10 from original Form B
Subtest 2: Items 3, 6, 8, 10, 12, 14 from original Form A, and Items 8, 11 from original Form B
Subtest 3: Items 1, 3, 4, 5, 12 from original Form A, and Items 5, 9, 11 from original Form B
Subtest 4: Items 3, 6, 10 from original Form A, and Items 3, 6, 7 from original Form B

Note that we left out Item 2 of each form in Subtest 1, Items 2, 3, 4, and 14 (Form B) for Subtest 2, Items 1 and 2 (Form b) for Subtest 3, and 1a and 1b for Subtest 4.

Table S1 Demographic and descriptive data for the scores obtained at pre-test shown for the four groups of participants.

	Training Completed			Dropped out			Withdrew (minimal training)			Paid Pretest (no training)		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Age	78	25.21	6.46	43	24.81	6.45	54	22.83	4.54	34	22.79	6.11
No. of women	36			23			23			17		
Completed training sessions	78	19.83	0.71	43	11.58	5.45	54	0.63	0.99	34	-	-
APM	78	15.00	2.56	43	13.95	2.61	54	13.35	3.34	34	14.00	2.19
BOMAT	78	17.85	3.48	43	16.05	3.06	44	15.43	4.63	34	16.38	3.52
CFT	78	19.65	3.44	43	19.98	3.19	54	18.17	3.22	34	18.65	3.63
Surface Development	78	20.46	8.28	43	18.09	8.53	53	17.64	8.01	34	19.26	6.56
Space Relations	78	12.82	3.75	43	12.26	3.92	45	12.49	3.49	34	12.09	3.62
Form Board	78	67.58	27.34	42	60.43	19.14	45	67.58	22.19	34	66.29	18.57
Inferences	78	7.46	1.88	43	6.70	2.25	53	7.08	1.99	33	7.36	1.52
Reading Comprehension	78	6.79	2.49	43	5.86	2.18	45	5.73	2.51	34	5.47	2.65
Verbal Analogies	77	0.73	0.09	43	.73	.11	52	.71	.11	34	0.72	0.09
DST	76	69.78	12.24	43	69.21	10.97	45	68.33	12.28	33	73.79	8.82
Cognitive Failures	77	112.21	21.17	42	114.81	21.57	43	113.58	21.91	34	100.68	18.94
Need for Cognition	77	68.56	38.22	43	50.79	44.07	42	59.21	36.36	34	50.97	32.19
Theories of Intelligence	76	33.71	8.19	39	32.49	8.07	44	33.09	8.28	34	32.82	7.05

Note: There are different sample sizes in the different tests because some participants either did not complete the test (especially in the group that withdrew from training some of which completed only one of the two pre-test sessions), or the data were not recoreded due to technical errors.

Table S2 Exploratory factor analysis based on pretest scores (pattern matrix).

	Factor	
	Verbal	Visuospatial
APM	.208	**.610**
BOMAT	.098	**.684**
CFT	**.512**	.144
Surface Development	.097	**.750**
Space Relations	-.010	**.562**
Form Board	-.150	**.641**
Inferences	**.843**	-.083
Reading Comprehension	**.657**	.019
Verbal Analogies	**.772**	-.005

N=77. Extraction Method: Maximum Likelihood. Rotation Method: Direct Oblimin with Kaiser Normalization. Two components had eigenvalues over Kaiser's criterion of 1 and in combination explained 48.65% of the variance.

Table S3 Pearson's correlations of indiviudal measures at the pre-test.

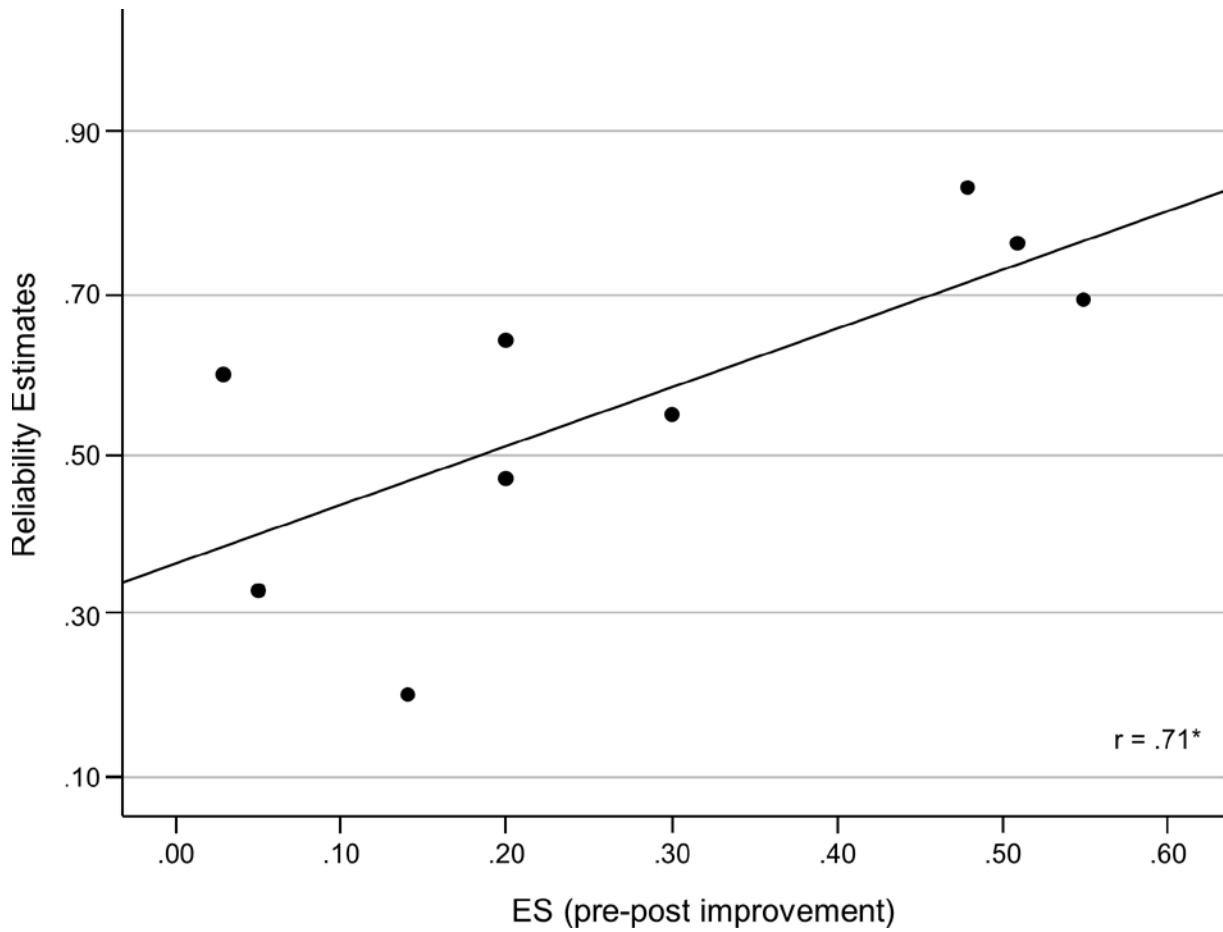|	1	2	3	4	5	6	7	8	9
1 APM	-	.60**	.31**	.55**	.33**	.33**	.34**	.30**	.30**
2 BOMAT		-	.43**	.54**	.34**	.38**	.19	.26*	.26*
3 CFT			-	.30**	.06	.08	.47**	.31**	.44**
4 Surface Development				-	.49**	.47**	.26*	.25*	.31**
5 Space Relations					-	.41**	.16	.19	.17
6 Form Board						-	.04	.07	.09
7 Inferences							-	.54**	.63**
8 Reading Comprehension								-	.54**
9 Verbal Analogies									-

Note: *p<.05; **p<.01.

Figure S1 Effect sizes (transfer effects for the nine outcome measures) plotted as a function of

reliability; illustrated for the two n-back training groups (averaged data from Table 2).