JOURNAL OF COGNITIVE PSYCHOLOGY, 2015
http://dx.doi.org/10.1080/20445911.2015.1118106

Routledge
Taylor & Francis Group

Increased training complexity reduces the effectiveness of brief working memory training: evidence from short-term single and dual *n*-back training interventions

Kristina Küper[a,b] and Julia Karbach[b,c]

[a]Aging Research Group, Leibniz Research Centre for Working Environment and Human Factors, Dortmund, Germany; [b]Department of Educational Science, Saarland University, Saarbrücken, Germany; [c]Department of Psychology, Goethe-University, Frankfurt, Germany

ABSTRACT

N-back training has recently come under intense scientific scrutiny due to reports of training-related improvements in general fluid intelligence. As of yet, relatively little is known about the effects of short-term *n*-back training interventions, however. In a pretest-training-posttest design, we compared brief dual and single *n*-back training regimen in terms of training gains and transfer effects relative to a passive control group. Transfer effects indicated that, in the short-term, single *n*-back training may be the more effective training task: At the short training duration we employed, neither training group showed far transfer to specific task switch costs, Stroop inhibition costs or matrix reasoning indexing fluid intelligence. Yet, both types of training resulted in a reduction of general task switch costs indicating improved cognitive control during the sustained maintenance of competing task sets. Single but not dual *n*-back training additionally yielded near transfer to an untrained working memory updating task.

ARTICLE HISTORY
Received 17 February 2015
Accepted 4 November 2015

KEYWORDS
Dual *n*-back training; single *n*-back training; short-term training; working memory training; transfer effects

Introduction

Working memory (WM), the ability to maintain and manipulate a limited amount of information over a limited time span, is crucial for higher order cognition and in turn everyday functioning and academic and occupational success (Miyake & Shah, 1999). The close correlation between WM capacity and general fluid intelligence (e.g. Conway, Cowan, Bunting, Therriault, & Minkoff, 2002) has raised substantial interest in training interventions designed to enhance WM function. Previous research has shown that performance in a variety of WM tasks can be improved even by relatively short-term training on these tasks. Importantly, transfer effects, that is, training-related performance gains in untrained tasks, indicate that these performance benefits reflect an enhancement of underlying cognitive processes rather than mere task-specific practice effects (Au et al., 2014; Karbach & Verhaeghen, 2014).

In an influential study by Jaeggi, Buschkuehl, Jonides, and Perrig (2008), participants completed short- and long-term versions of an adaptive dual *n*-back training. In single *n*-back tasks, participants are presented with a continuous stream of items and have to indicate whether a given stimulus matches an item presented *n* trials beforehand. In the more complex dual *n*-back task, this matching task had to be performed simultaneously on visually presented spatial positions and auditorily presented letters. In addition to performance gains in the trained task, the authors observed far transfer to performance in a matrix reasoning task measuring general fluid intelligence. This far transfer effect to fluid intelligence appears to be somewhat elusive, however. Although it has been replicated in several studies employing the same complex dual *n*-back training (Jaeggi, Buschkuehl, Shah, & Jonides, 2014; Jaeggi, Studer-Luethi, Buschkuehl, Su, Jonides, & Perrig, 2010; Schweizer, Hampshire, & Dalgleish, 2011; Stephenson & Halpern, 2013), it has been notably absent in a number of others (Chooi & Thompson, 2012; Redick et al., 2013; Salminen, Strobach, & Schubert, 2012; Thompson et al., 2013).

In light of the large body of research on the effects of dual *n*-back training, surprisingly little is known about the cognitive processes which

CONTACT Kristina Küper kueper@ifado.de

underlie task performance and which may in turn mediate potential effects of dual *n*-back training on fluid intelligence. The diverse and somewhat inconsistent transfer effects associated with training on the dual *n*-back task indicate that it is a very complex task drawing on various distinct cognitive processes each of which has been discussed as a potential mediating factor in far transfer to fluid intelligence. Near transfer effects have been reported for simple span tasks measuring short-term memory capacity but not for complex span tasks involving additional processing demands (Jaeggi et al., 2008; Lilienthal, Tamez, Shelton, Myerson, & Hale, 2013; Schweizer et al., 2011). Dual *n*-back training has improved updating processes in some cases (Jaeggi et al., 2010; Salminen et al., 2012) but not in others (Lilienthal et al., 2013) and has been shown to either expand the focus of attention (Lilienthal et al., 2013) or improve the efficiency of dividing attention (Salminen et al., 2012). Salminen et al. (2012) additionally reported an improvement in the ability to maintain and select task sets in a task-switching paradigm.

It has been argued that far transfer to fluid intelligence is contingent on the increased task demands of the dual *n*-back task (Jaeggi et al, 2010). Yet, direct comparisons of dual and single *n*-back training have yielded comparable transfer effects, at least for long-term training interventions across 20 sessions (Jaeggi et al., 2010, 2014; Stephenson & Halpern, 2013). A recent meta-analysis by Au et al. (2014) similarly found a small but significant transfer effect from *n*-back training to fluid intelligence which was equivalent for single and dual versions of the task. Again, the authors confined their analysis to studies employing relatively long-term *n*-back training regimen across on average 19.5 sessions.

There is evidence to suggest, however, that very brief training interventions may be sufficient to enhance neural efficiency, at least when it comes to training on the single *n*-back task: Vartanian et al. (2013) found lower prefrontal brain activation during a divergent thinking task accompanied by enhanced matrix reasoning ability after only three single *n*-back training sessions. This is especially noteworthy since dual *n*-back training has been shown to yield significant transfer to fluid intelligence only after 17 or more training sessions, but not after short-term training (Jaeggi et al., 2008). This raises the question of whether single *n*-back training may be more effective than dual *n*-back training when training duration is short. With the exception of the two studies mentioned above, however, relatively little research has focused on the effects of short-term *n*-back interventions.

The present study addressed this issue by systematically examining training gains and transfer effects associated with brief *n*-back training interventions of different complexity. Participants in the training groups received five sessions of either dual or single *n*-back training. In order to estimate near and far transfer effects, participants completed untrained cognitive tasks measuring WM updating, inhibition, task switching and fluid intelligence, before and after the intervention. Data from the two training groups were compared to a passive control group allowing us to examine whether dual and single *n*-back training differ in terms of training gains and transfer effects when training duration is short.

Method

Participants

Fifty-four healthy young adults (17–36 years; $M = 23.3$, $SD = 4.1$) participated in this study. They were screened for colour-blindness and psychotropic medication, provided written informed consent and received course credit or payment (8 EUR/hour) for participating in the study. Participants were randomly assigned to a passive control condition, a single *n*-back training condition, or a dual *n*-back training condition. The three groups were comparable in terms of education, self-rated physical and mental health, perceptual speed, semantic knowledge, and fluid intelligence (all *p*-values > .13, see Table 1).

Table 1. Demographic characteristics, well-being, and cognitive baseline performance as a function of group (single *n*-back, dual *n*-back, control group).

Variable	Single *n*-back		Dual *n*-back		Control group	
N	18		18		18	
Age range (years)	20–36		17–26		19–32	
	M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
Years of education	13.8	(1.9)	13.00	(0.0)	13.1	(1.4)
Self-rated physical health (one item; 1 = high, 6 = low)	2.1	(0.7)	2.1	(0.6)	2.4	(0.8)
Self-rated mental health (one item; 1 = high, 6 = low)	2.0	(1.2)	1.9	(0.8)	2.0	(0.8)
Digit-symbol substitution test score	65.6	(11.5)	69.5	(9.2)	64.7	(11.0)
Semantic knowledge test Score	23.1	(3.7)	20.4	(3.1)	22.5	(5.1)
Raven matrices test score	21.3	(4.9)	23.0	(5.0)	19.6	(6.4)

Procedure and tasks

We assessed transfer of WM training by means of a pretest-training-posttest design. Transfer was defined as performance improvement at posttest relative to baseline performance at pretest. Participants in the training groups attended seven sessions: One for the pretest and posttest assessment (60–80 minutes each), respectively, and five training sessions (30 minutes each). The pretest session included a battery of cognitive transfer tasks and the control measures included to test for between-group differences (perceptual speed and semantic knowledge). It was followed by the training sessions that were completed within two weeks (no more than one session per day or three sessions per week). The posttest session included the same test battery as the pretest. Participants in the control condition only completed the pretest and the posttest. The order of tasks at pretest and posttest was the same for all participants (perceptual speed (only at pretest), WM, task switching, inhibition, fluid intelligence, semantic knowledge (only at pretest)).

Pretest and posttest sessions

Near transfer of WM training was estimated using an untrained WM task. To investigate far transfer to other cognitive control tasks, we applied a switching task, an inhibition task, and a test of fluid intelligence.

WM

Participants performed a 3-back task on a sequence of digits successively presented on the computer screen. They were to press the response key if the current digit matched the digit presented three positions back in the sequence. Trials started with the presentation of a fixation-cross (500 ms), followed by the target (1000 ms), and another fixation-cross (500 ms). Subjects were to respond as fast and as accurately as possible. They performed four 3-back blocks with 24 trials each (including 6 targets). Test performance was estimated by calculating the total number of hits as well as false alarms.

Task switching

We applied a task-switching paradigm including single-task (task A or B only) and mixed-task blocks (switching between both tasks, cf. Karbach & Kray, 2009). In mixed-task blocks, subjects had to switch tasks on every second trial. In task A, they were to decide whether a picture showed a fruit or a vegetable ("food" task), and in task B whether a picture was small or large ("size" task). Stimuli consisted of 16 fruit and 16 vegetable pictures, each presented in a large and a small version (yielding a total of 64 stimuli). Trials started with the presentation of a fixation-cross (500 ms), followed by the target (1500 ms) and an inter-trial interval (1000 ms). Participants were to respond as fast and as accurately as possible. They performed 10 single-task and 10 mixed-task blocks (17 trials per block) in an alternating sequence. We calculated two different measures: General switch costs, measuring the ability to maintain and select between two task sets, were defined as the difference in performance between single-task blocks and mixed-task blocks. Specific switch costs, indexing the ability to flexibly switch between tasks on trial-to-trial transitions, were defined as the difference in performance between stay (AA, BB) and switch (AB, BA) trials within mixed blocks.

Inhibition

In a computerised Stroop task, participants had to indicate the font colour of words presented in red, blue, green, or yellow font by pressing one of four response keys as quickly as possible. Words were presented in uppercase 18-point font against a black background. In congruent trials, colour names were identical to the font colour (e.g. "red" presented in red). In incongruent trials, word meaning interfered with the font colour (e.g. "blue" presented in yellow) and in neutral trials words were not semantically linked to colours (e.g. "hat" presented in green). Trials started with the presentation of the stimulus for 2000 ms or until the subject responded, followed by a response-stimulus interval of 700 ms. Participants first performed two practice blocks (12 trials each) followed by four experimental blocks (24 trials each).

Fluid intelligence

We applied 36 items from Raven's Advanced Progressive Matrices (Raven, Raven, & Court, 2003; internal consistency = .85). Subjects were instructed to select one of eight figures that best completed a given pattern shown on the computer screen. The test score referred to the number of correctly solved items within 20 minutes.

Perceptual speed

In the digit-symbol substitution test (Wechsler, 1982; internal consistency = .87), nine digit-symbol mappings were presented at the top of a test sheet. Below, 100 digits without the corresponding symbols were displayed and participants were instructed to fill in the missing symbols. The test score was the number of correctly added symbols within 90 seconds.

Semantic knowledge

In the spot-a-word test (cf. Lehrl, 1977; internal consistency = .93), 35 items were successively presented on the computer screen, each containing one word and four pronounceable nonwords, numbered from one to five. Participants had to identify the word and press the corresponding number on the keyboard. The test score was the number of correctly solved items within five minutes.

Training sessions

Single n*-back training*

Participants saw successive letters on the screen and were to press the response key if the letter on the screen was identical to the one n positions back in the sequence. Trials started with the presentation of a fixation-cross (500 ms), followed by the target (1000 ms), and another fixation-cross (500 ms). At the beginning of each block, an instruction window appeared, indicating the *n*-level for the next block. Task difficulty was individually adapted to performance: Subjects started the first training session with one block on the 2-back level. Afterwards, the difficulty level was adjusted at the beginning of each new block based on the performance score of the preceding one: If performance accuracy was above 67%, the *n*-back level increased by one, if it was between 67% and 33%, the *n*-back level did not change, and if it was below 33%, the *n*-back level decreased by one. At the beginning of each training session, participants resumed training with the *n*-back level of the last block in the preceding training session. Each session consisted of 15 experimental blocks (20 + *n*-back level trials including six targets per block).

Dual n*-back training*

Participants performed the dual *n*-back task introduced by Jaeggi et al. (2008), including an auditory and a visual task. In the auditory task, letters (C, G, H, K, P, Q, T, W) were presented via headphones. In the visual task, blue squares were presented successively at eight different locations on the screen in random order. Stimuli were presented for 500 ms, followed by an inter-stimulus interval of 2500 ms. Participants responded by pressing left (visual task) and right (auditory task) response keys when the current stimulus matched the one presented *n* trials before in the sequence. The *n*-level was always identical for both tasks. Training started with one block at the 2-back level; the difficulty level of the following blocks was adjusted based on the performance score of the preceding block. If participants performed at least 90% of trials accurately in both modalities, the n-level was increased by one. If response accuracy was between 90% and 70%, the *n*-level did not change and if accuracy was below 70% the *n*-level decreased by one.[1] At the beginning of each training session, participants resumed training with the *n*-back level of the last block in the preceding training session. Each session consisted of 15 experimental blocks (20 + *n*-back level trials per block) with six auditory and six visual targets per block.

Data analysis

Performance gains on the training tasks (single *n*-back, dual *n*-back) were analysed based on the mean *n*-level per training session. The transfer measures were analysed with individual analyses of variance (ANOVA) and with a multiple ANOVA to protect against type 1 errors. Both analyses yielded the same pattern of results. Analyses for the switching task and the inhibition task were restricted to mean reaction times (RT, ms) for correct responses and accuracy. Practice blocks, the first trial in each block, and responses faster than 200 ms were not considered. In order to control for potential differences in general response speed, we also ran control analyses based on log-transformed RT. After this transformation, costs were expressed as differences between logarithms,

[1]The two *n*-back training regimen differed with respect to the threshold used for the adaptive adjustment of task difficulty. The present data indicate, however, that this did not have a substantial effect on training and transfer gains. The single *n*-back group who trained with a more lax adjustment criterion did not show a particularly steep increase in *n*-level across training sessions, but levelled up at a comparable rate as the dual *n*-back group. This is due to the fact that a more lax criterion for levelling up would come to bear primarily once training gains have reached a plateau: a lax criterion could then potentially allow for more additional level-ups than a stricter criterion. Figure 1 indicates, however, that training gains were not yet at ceiling after the five training sessions of the present study.

Table 2. Means (*SD*) of the performance on the transfer measures as a function of session (pretest, posttest) and group (single *n*-back, dual *n*-back, control).

	Single *n*-back		Dual *n*-back		Control group	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
WM transfer (*N* hits)	15.8 (4.7)	20.8 (2.6)	19.3 (4.0)	21.4 (1.7)	14.6 (4.0)	17.2 (4.2)
WM transfer (*N* false alarms)	6.2 (5.0)	6.8 (4.6)	2.8 (2.0)	2.9 (2.5)	7.6 (3.3)	6.7 (4.4)
Task switching (RT)						
Switch (ms)	738.3 (183.8)	649.3 (143.6)	627.2 (161.5)	598.5 (131.3)	765.2 (117.6)	744.9 (137.5)
Stay (ms)	634.3 (133.3)	569.7 (107.9)	540.37 (113.8)	537.1 (106.4)	658.6 (137.0)	651.0 (140.9)
Single (ms)	557.1 (95.7)	525.7 (91.3)	507.0 (79.8)	501.6 (67.2)	614.2 (77.1)	587.1 (97.6)
Inhibition (RT)						
Incongruent (ms)	666.1 (124.8)	590.8 (73.0)	613.9 (99.8)	567.8 (105.1)	693.6 (190.5)	629.5 (133.2)
Neutral (ms)	631.8 (100.2)	564.2 (63.5)	580.4 (77.6)	538.1 (73.6)	636.7 (146.0)	581.6 (97.1)
Congruent (ms)	615.5 (101.6)	559.9 (53.3)	561.5 (68.7)	523.5 (53.4)	623.3 (147.9)	565.8 (97.3)
Fluid intelligence (*N* correct responses)	21.3 (4.8)	23.6 (6.1)	22.9 (5.0)	24.9 (5.5)	19.6 (6.4)	20.7 (7.3)

which is equivalent to ratio scores (Ratcliff, 1993) that are less sensitive to differences in baseline performance (Meiran, 1996). Given that the pattern of results did not change, we only report analyses based on mean RT. General and specific switch costs were defined as two orthogonal contrasts for the factor trial type (single, stay, switch). General switch costs can also be calculated as the difference in performance between single and stay trials ("mixing cost") rather than the difference between single and mixed-task blocks. Therefore, we additionally ran a set of control analyses on mixing costs, which yielded the same result pattern as reported below. Means and standard deviations for all tasks are provided in Table 2.

Results

Testing for group differences at baseline

In order to test for preexisting group differences at baseline, we analysed pretest data for each of the transfer measures. Repeated-measures ANOVA with the between-subjects factor Group (single *n*-back, dual *n*-back, control) and the within-subjects factor Trial type for the task switching (single, stay, switch) and the inhibition task (congruent, incongruent, neutral) revealed no significant interactions (all $p > .29$), indicating that there were no differences in performance at pretest. One-way ANOVA with the between-subjects factor Group (single *n*-back, dual *n*-back, control) for the measures of fluid intelligence, perceptual speed, and semantic knowledge similarly revealed no significant group effects (all $p > .11$). We did, however, find significant differences in WM performance both on the level of hits, $F(2, 53) = 5.96$, $p < .01$, $\eta^2_p = .65$, as well as false alarms, $F(2, 53) = 8.25$, $p < .001$, $\eta^2_p = .25$, with the dual *n*-back group performing better than the remaining groups (all $p < .05$). When analysing WM transfer effects, we therefore controlled for these baseline differences by computing the pretest–posttest difference relative to the pretest performance (i.e. (post-pre)/pre). For these analyses, the pattern of results was identical to the one reported below.

Training gains

A repeated-measures ANOVA including the within-subjects factor Training session (1–5) and the between-subjects factor Group (single *n*-back, dual *n*-back) yielded a significant main effect of Training session, $F(4, 136) = 63.07$, $p < .0001$, $\eta^2_p = .65$, showing an increase of *n*-level as a function of training (Figure 1), with a significant linear increase from the first ($M = 2.9$, $SD = 1.0$) to the last ($M = 4.8$, $SD = 1.3$) session, $F(1, 34) = 114.80$, $p < .0001$, $\eta^2_p = .77$. Both the main effect of Group ($p = .23$) and the interaction between Group and Training session ($p = .17$) were not significant.

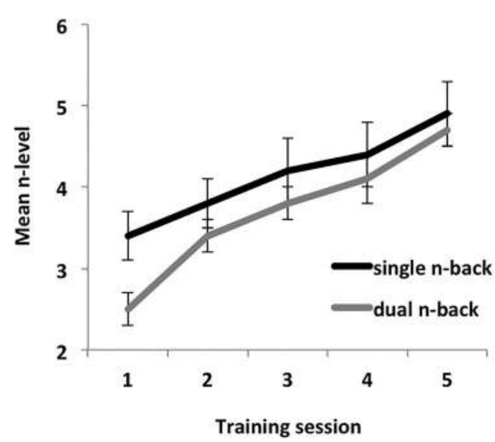

Figure 1. Performance on the training tasks as a function of training session (1–5) and group (single *n*-back, dual *n*-back). Error bars refer to standard errors of the mean.

Transfer effects

WM

A repeated-measures ANOVA of the mean number of hits with the within-subjects factor Session (pretest, posttest) and the between-subjects factor Group yielded a main effect of Session, $F(1, 51) = 46.73$, $p < .0001$, $\eta_p^2 = .48$, pointing to general performance improvements from pretest to posttest. The main effect of Group was also significant, $F(2, 51) = 8.71$, $p < .01$, $\eta_p^2 = .26$, indicating that the control group performed worse than the training groups ($p < .01$), and that the dual *n*-back group performed marginally better than the single *n*-back group ($p = .06$). The interaction between Session and Group, $F(1, 51) = 3.39$, $p < .05$, $\eta_p^2 = .12$, revealed that the improvement from pretest to posttest was larger in the single *n*-back group than in the remaining groups ($p < .05$), but there was no difference between the control group and the dual *n*-back training group ($p = .32$). The analyses of false alarms also yielded a significant main effect of Group, $F(2, 51) = 14.17$, $p < .0001$, $\eta_p^2 = .36$, indicating that the control group performed worse than the training groups ($p < .01$), and that the dual *n*-back group performed better than the single *n*-back group ($p < .001$). We neither found a significant main effect of Session nor an interaction between Session and Group (both $p > .71$).

Task switching—reaction times

A repeated-measures ANOVA with the between-subjects factor Group and the within-subjects factors Session and Trial type (single, stay, switch) yielded a main effect of Session, $F(1, 51) = 11.22$, $p < .01$, $\eta_p^2 = .18$, indicating that participants responded faster at posttest than at pretest. A main effect of Trial type, $F(2, 102) = 111.59$, $p < .0001$, $\eta_p^2 = .69$, revealed significant general and specific switch costs ($F(1, 51) = 103.67$, $p < .0001$, $\eta_p^2 = .67$ and $F(1, 51) = 126.04$, $p < .0001$, $\eta_p^2 = .71$, respectively). A main effect of Group, $F(2, 51) = 5.39$, $p < .01$, $\eta_p^2 = .17$, showed that the training groups responded faster than the control group ($p < .05$), without a difference between the training groups ($p = .10$). The interaction between Group and Trial type was not significant ($p = .27$), but we found an interaction between Session and Trial type, $F(2, 102) = 3.88$, $p < .05$, $\eta_p^2 = .07$, as well as marginally significant interactions between Session and Group, $F(2, 51) = 2.84$, $p = .07$, $\eta_p^2 = .10$, and Session, Group, and Trial type, $F(4, 102) = 2.21$, $p = .07$, $\eta_p^2 = .67$. Within-subjects contrasts indicated that the interaction between Session and Group was significant for general switch costs, $F(2, 51) = 3.42$, $p < .05$, $\eta_p^2 = .12$, but not for specific switch costs ($p = .79$). Follow-up analyses showed a reduction of general switch costs from pretest to posttest in the training groups, $F(1, 34) = 4.77$, $p < .05$, $\eta_p^2 = .12$, but not in the control group ($p = .48$). Although the reduction of general switch costs was larger in the single *n*-back group than in the dual *n*-back group, this difference did not reach significance ($p = .18$) (Figure 2).

Task switching—accuracy

The same analysis based on accuracy (% errors) only revealed a significant main effect for Trial type, $F(2, 102) = 11.12$, $p < .0001$, $\eta_p^2 = .18$, reflecting general and specific switch costs ($F(1, 51) = 6.24$, $p < .05$, $\eta_p^2 = .11$ and $F(1, 51) = 35.06$, $p < .0001$, $\eta_p^2 = .41$, respectively). No further effects reached significance.

Inhibition—reaction times

A repeated-measures ANOVA with the between-subjects factor Group and the within-subjects factors Session and Trial type (congruent, incongruent, neutral) showed a main effect of Session, $F(1, 51) = 29.84$, $p < .0001$, $\eta_p^2 = .37$, revealing faster responses at posttest than at pretest. A main effect of Trial type, $F(2, 102) = 27.55$, $p < .0001$, $\eta_p^2 = .35$, revealed significant interference costs (congruent vs. incongruent), $F(1, 51) = 34.72$, $p < .0001$, $\eta_p^2 = .41$. No further effects reached significance (all *p*-values > .27).

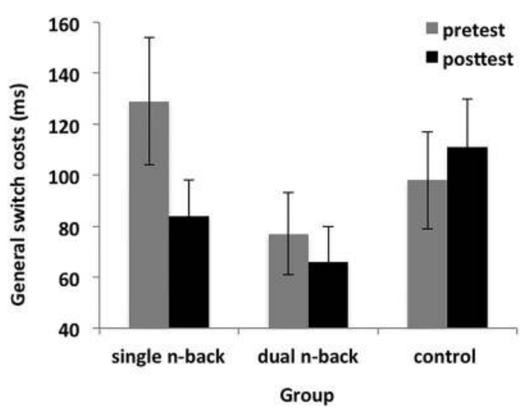

Figure 2. General switch costs as a function of training session (pretest, posttest) and group (single *n*-back, dual *n*-back). Error bars refer to standard errors of the mean.

Inhibition—accuracy

The analysis based on accuracy (% errors) yielded a main effect for Trial type, $F(2, 102) = 18.92$, $p < .0001$, $\eta_p^2 = .27$, revealing significant interference costs ($F(1, 51) = 19.68$, $p < .0001$, $\eta_p^2 = .28$. No further effects reached significance.

Fluid intelligence

A repeated-measures ANOVA with the between-subjects factor Group and the within-subjects factor Session revealed a significant effect of Session, $F(1, 51) = 15.47$, $p < .0001$, $\eta_p^2 = .23$, indicating that performance improved from pretest to posttest. The main effect of Group was marginally significant, $F(1, 51) = 2.00$, $p = .07$, $\eta_p^2 = .07$, pointing to slightly better performance in the training groups than in the control group ($p = .07$). The interaction between Group and Session failed to reach significance ($p = .58$).

Discussion

The aim of the present study was to examine the effectiveness of brief *n*-back training interventions of different complexity. In a pretest-training-posttest design, we compared short-term dual and single *n*-back training regimen in terms of training gains and transfer effects relative to a passive control group.

Across five training sessions, both training groups showed an equivalent linear increase in *n*-back level. Near transfer effects were subject to training complexity: We observed transfer to an untrained 3-back WM updating task only in the single but not in the dual *n*-back training group. Regarding far transfer, we found a reduction of general task switch costs from pretest to posttest in both training groups, but not in the control group. Neither training group showed far transfer to specific switch costs, inhibition costs in a Stroop task or matrix reasoning ability indexing fluid intelligence.

Of the two training tasks employed in the present study, the dual *n*-back task is considered the significantly more complex one. Adaptive dual *n*-back training is thus usually performed at a consistently lower *n*-back level than single *n*-back training (Jaeggi et al., 2010, 2014). In the present study, however, both the single and the dual *n*-back training group trained at comparable *n*-back levels across all five training sessions. For both groups, starting *n*-back levels in the first session were similar to those reported in previous research (e.g. Chooi & Thompson, 2012; Jaeggi et al., 2010, 2014; Smith, Stibric, & Smithson, 2013; see also Au et al., 2014). Baseline differences between studies thus cannot account for these contrasting training patterns. Similarly, years of education, perceptual speed, semantic knowledge, and fluid intelligence were comparable for both training groups and the control group. Preexisting individual differences may nevertheless have been a contributing factor since WM updating performance at pretest was better in the dual *n*-back group compared to the single *n*-back group and the control group. It is thus feasible that updating mechanisms were particularly effective in these participants to begin with enabling them to consistently perform the dual *n*-back training task at a higher difficulty level than reported previously.

In addition to performance gains in the trained tasks, we observed near transfer to an untrained WM updating task. Performance improvements from pretest to posttest were, however, limited to the single *n*-back group and did not emerge as a result of dual *n*-back training, even when controlling for this group's superior WM updating performance at pretest (see above). Previously, dual *n*-back training has improved WM updating in some cases (Jaeggi et al., 2010; Salminen et al., 2012) but not in others (Lilienthal et al., 2013). In a comparison of dual and single *n*-back training regimen, Jaeggi et al. (2010) found comparable near transfer effects to updating after both types of training. These equivalent near transfer effects were, however, observed after 20 training sessions rather than the five sessions participants completed in the present study. In keeping with this, Salminen et al. (2012) who also reported transfer to updating had participants complete 14 dual *n*-back training sessions whereas Lilienthal et al. (2013) who failed to find this near transfer effect trained participants for only eight sessions. Taken together with our data, these previous findings indicate that dual *n*-back training may positively affect WM updating processes only after a sufficient number of training sessions. Adequate training length appears to be crucial for the effectiveness of dual *n*-back training on a more general scale as well since Jaeggi et al. (2008) observed that the magnitude of far transfer from dual *n*-back training to fluid intelligence increased as a function of training length.

For single *n*-back training, on the other hand, five training sessions were sufficient to improve updating processes. This discrepancy between the two training groups may indicate that, at least at short training durations, dual and single *n*-back training initially affect distinct underlying processes: Single *n*-back training directly enhances WM updating whereas the dual task demands of dual *n*-back training chiefly put strain on cognitive control mechanisms which are in turn selectively exercised. Updating processes may thus only benefit from dual *n*-back training after participants have gained sufficient cognitive control to accommodate the two competing tasks, that is, after a higher number of training sessions. For the present study, an exploratory correlational analysis could, however not directly confirm an association between updating performance and training gains in either the single ($p = .21$) or the dual *n*-back training group ($p = .16$). We can thus not rule out alternative interpretations of the between-group differences in WM updating transfer effects. Participants in the single *n*-back group could, for example, have shown better posttest performance in the WM updating task because they found it to be structurally more similar to their training task than participants in the dual *n*-back group did: As opposed to the dual *n*-back task, participants had to attend to only a single stream of stimuli in both the single *n*-back and the WM transfer task. In addition, although all three tasks featured visual stimulus presentation, the visual stimuli used in the WM transfer task were more similar to those used in the single *n*-back training compared to the visually presented spatial positions used in the dual *n*-back training. We can thus not rule out the possibility that task-specific training benefits played a role in the between-group differences in WM updating transfer effects. The easier single *n*-back task could potentially also offer a more positive task experience which could in turn facilitate transfer to WM updating. The question remains, however, why a more positive task experience or indeed any other difference between single and dual *n*-back training which is not directly related to WM updating processes would selectively affect only these processes but not the cognitive abilities tested in the other transfer tasks.

Regarding the effects of *n*-back training on cognitive control processes, we observed a reduction of general but not specific switch costs from pretest to posttest in both training groups, but not in the control group, a result pattern previously reported by Salminen et al. (2012) after 14 sessions of dual *n*-back training. The authors concluded that dual *n*-back training selectively improved cognitive control mechanisms but not inhibition processes since efficient task performance requires attending to both *n*-back tasks simultaneously rather than inhibiting one or the other. The present data corroborate this notion, since dual *n*-back training was neither associated with a reduction of specific task switch costs nor with performance gains in a Stroop task measuring inhibition. What is more, our data extend previous findings as they indicate that transfer to cognitive control can emerge after as few as five training sessions. Importantly, the dual task demands of the dual *n*-back task are apparently not a necessary prerequisite for this far transfer effect since equivalent performance gains were observed after single *n*-back training. In fact, the reduction of general switch costs was numerically even larger in the single than in the dual *n*-back group. This may, however, simply reflect ceiling effects in the performance of the dual *n*-back group who showed numerically reduced general switch costs relative to the other two groups even at pretest.

Neither of our two training groups showed far transfer to fluid intelligence as measured by a matrix reasoning task. This was to be expected for the dual *n*-back group since far transfer to fluid intelligence has been shown to be contingent on a relatively high number of dual *n*-back training sessions (Au et al, 2014; Jaeggi et al., 2008). In contrast, Vartanian et al. (2013) reported that a mere three sessions of single *n*-back training resulted in far transfer to fluid intelligence as measured by a modified version of Raven's Advanced Progressive Matrices (Raven et al., 2003) featuring a 10-minute time limit. Operationalising fluid intelligence in this manner has been criticised on the grounds that it unduly prioritises the ability to quickly solve the easier visual analogies included in a matrix reasoning test at the expense of measuring the more relevant ability to solve the test's more difficult items (Moody, 2009). The present data are based on a more conservative measure of matrix reasoning and indicate that, while dual and single *n*-back training may improve fluid intelligence in the long-term (Jaeggi et al., 2010, 2014; Stephenson & Halpern, 2013), they cannot accomplish this at a very short training duration.

As we used a passive control group we cannot rule out the possibility that the increased computer-based and/or general cognitive activity of the training groups contributed to some of the training and transfer gains we observed (cf. Redick et al., 2013). Like the majority of *n*-back training research (Au et al., 2014), the present study featured a relatively small sample size of 18 participants per group. Small sample sizes not only increase the risk of type 1 and type 2 errors (Bogg & Lasecki, 2014) but may also lead to an overestimation of effect sizes (Halsey, Curran-Everett, Vowler, & Drummond, 2015). It follows that the results of the present study need to be treated with caution and should be substantiated by replicating the present data pattern in a larger sample.

To summarise, it appears that, in the short-term, single *n*-back training may be the more effective training task compared to dual *n*-back training. At least in the present study, brief single *n*-back training improved WM updating processes in addition to cognitive control mechanisms associated with the sustained maintenance of competing task sets. Transfer effects from brief dual *n*-back training, on the other hand, were limited to these cognitive control processes probably owing to the increased complexity of this training task.

Acknowledgement

The authors would like to thank Sarah Brieber for her help during data acquisition.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by Saarland University under a start-up grant.

References

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2014). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review*, 22(2), 366–377. doi:10.3758/s13423-014-0699-x

Bogg, T., & Lasecki, L. (2014). Reliable gains? Evidence for substantially underpowered designs in studies of working memory training transfer to fluid intelligence. *Frontiers in Psychology*. *5*, doi:10.3389/fpsyg.2014.01589

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

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

Halsey, L.G., Curran-Everett, D., Vowler, S. L., & Drummond, G.B. (2015). The fickle P value generates irreproducible results. *Nature Methods*, *12*(3), 179–185. doi:10.1038/nmeth.3288

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829–6833. doi:10.1073/pnas.0801268105

Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition*, *42*(3), 464–480. doi:10.3758/s13421-013-0364-z

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

Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science*, *12*(6), 978–990. doi:10.1111/j.1467-7687.2009.00846.x

Karbach, J., & Verhaeghen, P. (2014). Making working memory work: A meta-analysis of executive-control and working memory training in older adults. *Psychological Science*, *25*(11), 2027–2037. doi:10.1177/0956797614548725

Lehrl, S. (1977). *Mehrfach-Wahl-Wortschatz-Test B (MWT-B)*. Erlangen: Straube.

Lilienthal, L., Tamez, E., Shelton, J. T., Myerson, J., & Hale, S. (2013). Dual n-back training increases the capacity of the focus of attention. *Psychonomic Bulletin & Review*, *20*(1), 135–141. doi:10.3758/s13423-012-0335-6

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1423–1442. doi:10.1037/0278-7393.22.6.1423

Miyake, A., & Shah, P. (Eds.) (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York, NY: Cambridge University Press.

Moody, D.E. (2009). Can intelligence be increased by training on a task of working memory? *Intelligence*, *37*(4), 327–328. doi:10.1016/j.intell.2009.04.005

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532. doi:10.1037/0033-2909.114.3.510

Raven, J., Raven, J. C., & Court, J. H. (2003). *Manual for Raven's progressive matrices and vocabulary scales. Section 1: General Overview*. San Antonio, TX: Harcourt Assessment.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., … Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, *142*(2), 359–379. 10.1037/a0029082

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

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*(9), 1–7. doi:10.1371/journal.pone.0024372

Smith, S. P., Stibric, M., & Smithson, D. (2013). Exploring the effectiveness of commercial and custom-built games for cognitive training. *Computers in Human Behavior*, *29*(6), 2388–2393. doi:10.1016/j.chb.2013.05.014

Stephenson, C. L., & Halpern, D. F. (2013). Improved matrix reasoning is limited to training on tasks with a visuospatial component. *Intelligence*, *41*(5), 341–357. doi:10.1016/j.intell.2013.05.006

Thompson, T. W., Waskom, M. L., Garel, K. L. A., 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*(5), 1–15. doi:10.1371/journal.pone.0063614

Vartanian, O., Jobidon, M. E., Bouak, F., Nakashima, A., Smith, I., Lam, Q., & Cheung, B. (2013). Working memory training is associated with lower prefrontal cortex activation in a divergent thinking task. *Neuroscience*, *236*, 186–194. doi:10.1016/j.neuroscience.2012.12.060

Wechsler, W. (1982). *Manual for the Hamburg-Wechsler intelligence test for adults (HAWIE)*. Bern: Huber.