

## The relationship between n-back performance and matrix reasoning – implications for training and transfer<sup>☆</sup>

Susanne M. Jaeggi<sup>a,\*</sup>, Barbara Studer-Luethi<sup>b,c</sup>, Martin Buschkuhl<sup>a</sup>, Yi-Fen Su<sup>c</sup>, John Jonides<sup>a</sup>, Walter J. Perrig<sup>b</sup>

<sup>a</sup> Department of Psychology, University of Michigan, 530 Church Street, Ann Arbor, MI, 48109-1043, USA

<sup>b</sup> Department of Psychology, University of Bern, Switzerland

<sup>c</sup> Department of Educational Psychology and Counseling, National Taiwan Normal University, Taipei, Taiwan

### ARTICLE INFO

#### Article history:

Received 20 March 2010

Received in revised form 8 August 2010

Accepted 6 September 2010

#### Keywords:

Cognitive training

Fluid intelligence

Working memory

### ABSTRACT

We have previously demonstrated that training on a dual n-back task results in improvements in fluid intelligence (*Gf*) as measured by matrix reasoning tasks. Here, we explored the underlying mechanisms of this transfer effect in two studies, and we evaluated the transfer potential of a single n-back task. In the first study, we demonstrated that dual and single n-back task performances are approximately equally correlated with performance on two different tasks measuring *Gf*, whereas the correlation with a task assessing working memory capacity was smaller. Based on these results, the second study was aimed on testing the hypothesis that training on a single n-back task yields the same improvement in *Gf* as training on a dual n-back task, but that there should be less transfer to working memory capacity. We trained two groups of students for four weeks with either a single or a dual n-back intervention. We investigated transfer effects on working memory capacity and *Gf* comparing the two training groups' performance to controls who received no training of any kind. Our results showed that both training groups improved more on *Gf* than controls, thereby replicating and extending our prior results.

© 2010 Elsevier Inc. All rights reserved.

## 1. Introduction

Fluid intelligence (*Gf*) is defined as a complex human ability that allows us to adapt our thinking to new cognitive problems

<sup>☆</sup> This work was supported by a grant of The Michigan Center for Advancing Safe Transportation throughout the Lifespan (MCASTL #DTRT07-G-0058) to MB, a fellowship from the Swiss National Science Foundation (PA001-117473) to SMJ, and a grant from the Office of Naval Research to JJ (N00014-09-0213). Study 2 was conducted as a part of BSL's master thesis with the guidance of the other co-authors and with the support of the University of Bern and the National Taiwan Normal University. The authors wish to thank Chao-Yi Ho and Philip Cheng for their help with the construction of the Mandarin stimuli and the help translating the instructions, as well as Courtney Behnke, Kirti Thummala, and Patrick Bissett, Wu Shan-Yun, Yi Han Chiu, and David Studer for their help with data collection, and finally, Randall Engle's group for providing us with the automated version of their automated operation span task.

\* Corresponding author.

E-mail address: [sjaeggi@umich.edu](mailto:sjaeggi@umich.edu) (S.M. Jaeggi).

or situations for which we cannot rely on previously acquired knowledge (e.g. Carpenter, Just, & Shell, 1990). *Gf* is considered critical for a wide variety of cognitive tasks (Engle, Tuholski, Laughlin, & Conway, 1999; Gray & Thompson, 2004), and it seems to be one of the most important factors in learning (Deary, Strand, Smith, & Fernandes, 2007; Neisser et al., 1996; Rohde & Thompson, 2007; te Nijenhuis, van Vianen, & van der Flier, 2007). There is considerable agreement that a substantial proportion of the variance in *Gf* is hereditary (Baltes, Staudinger, & Lindenberger, 1999; Cattell, 1963; Gray & Thompson, 2004), but it has also been shown that social class and age moderate the heritability of *Gf* (Haworth et al., 2009; Turkheimer, Haley, Waldron, D'Onofrio, & Gottesman, 2003). Although high heritability in principle does not preclude alteration of *Gf* through environmental factors or interventions (Jensen, 1981), evidence of such alteration has been sparse. However, there is now accumulating evidence showing that certain interventions seem to increase performance in *Gf* tasks

(Buschkuhl & Jaeggi, 2010), although the mechanisms that underlie such change are not well understood (Basak, Boot, Voss, & Kramer, 2008; Jaeggi, Buschkuhl, Jonides, & Perrig, 2008; Klingberg et al., 2005; Klingberg, Forssberg, & Westerberg, 2002; Rueda, Rothbart, McCandliss, Saccomanno, & Posner, 2005; Tranter & Koutstaal, 2007). For example, we have shown that the n-back task can be used as a training vehicle to improve performance on matrix reasoning tasks which are commonly used as a typical measure of *Gf* (e.g. Gray & Thompson, 2004; Kane & Engle, 2002; Snow, Kyllonen, & Marshalek, 1984). In our study, subjects were pretested on measures of *Gf*, after which they were given up to four weeks of daily training on a dual n-back task (Jaeggi, Buschkuhl, Jonides, & Perrig, 2008). The dual n-back task consisted of a position that was pseudo-randomly marked on a computer screen in each stimulus frame which subjects had to match for spatial position to the stimulus presented n frames back in the sequence. Simultaneously with the spatial task, subjects had to process an auditory stream of stimuli in which a single letter was presented in each auditory frame that had to be matched to the letter that appeared n items ago. The value of n was matched for the spatial and verbal tasks, both of which required responses. The level of n changed during the experiment according to the participants' performance to keep overall task difficulty approximately constant. Following training, subjects were given non-overlapping items from an instrument measuring *Gf*. The results showed that training on a dual n-back task yielded improvements in *Gf* relative to a control group that did not train.

Why was this training regimen successful? That is, what mechanisms drive such transfer effects? We believe it is critical that the training and the transfer tasks share overlapping cognitive processes for transfer to succeed. Thus, we think that the gain in *Gf* emerges because the processes that are engaged by the training task also mediate performance in *Gf* tasks. We proposed that the framework by Halford, Cowan, and Andrews (2007) might serve as a useful model to understand why *Gf* can be improved by means of a working memory task. Their claim is that working memory and intelligence share a common capacity constraint, which is driven by attentional control processes. Other authors have come to a related conclusion (Gray, Chabris, & Braver, 2003; Kane et al., 2004), and in particular, Carpenter, Just, and Shell (1990) have proposed that the ability to derive abstract relations and to maintain a large set of possible goals in working memory accounts for individual differences in typical tasks that measure *Gf*. The underlying neural circuitries provide additional evidence for the shared variance between working memory and *Gf* in that both seem to rely on similar neural networks, most consistently located in lateral prefrontal and parietal cortices (Gray, Chabris, & Braver, 2003; Kane & Engle, 2002). Thus, it seems plausible that the training of a certain neural circuit might lead to transfer to other tasks that engage similar or at least overlapping neural circuits. Indeed, recent evidence shows that transfer occurs if the training and the transfer task engage overlapping brain regions, but not if they engage different regions (Dahlin, Neely, Larsson, Backman, & Nyberg, 2008; see also Persson & Reuter-Lorenz, 2008).

But overlapping processes and neural circuits might not be the only prerequisites for transfer. We believe that the training task has to be very carefully designed in a certain way to promote transfer. First, a successful training task must minimize

the development of strategies that are specific to the task in question because the object of training must be changes in the information processing system, not changes in the way one particular task is performed (cf. Ericsson & Delaney, 1998). Second, we think that it is very important to keep a persistently high level of training demand while also considering inter-individual performance differences. This can be achieved by using an adaptive training method that continuously adjusts the current training difficulty to the actual performance of each subject. Third, we argue that it is necessary to stress the information processing system during training, for example by taxing more than one input modality at a time or by having the subject engage in two tasks simultaneously (Oberauer, Lange, & Engle, 2004). As we have shown in our work, the dual n-back training paradigm is a task that fulfills these requirements and subsequently leads to the predicted transfer effects (Jaeggi et al., 2007; Jaeggi, Buschkuhl, Jonides, & Perrig, 2008). Nevertheless, although various versions of the n-back task are widely used in research, only few studies have examined the processes involved in n-back performance (e.g. Hockey & Geffen, 2004; Jaeggi, Buschkuhl, Perrig, & Meier, 2010; Kane, Conway, Miura, & Colflesh, 2007). Therefore, little knowledge is available about the cognitive processes that mediate performance in this task and consequentially, about the processes underlying n-back training that eventually promote transfer to *Gf*. In addition, although the n-back task is commonly regarded as a measure of working memory, its concurrent validity is still open to question (Jaeggi, Buschkuhl, Perrig, & Meier, 2010; Jarrold & Towse, 2006; Kane, Conway, Miura, & Colflesh, 2007; Oberauer, 2005). For example, research from Kane's lab as well as our own work suggests that the n-back task and more traditional measures of working memory capacity (e.g. reading span or operation span tasks) do not share a great deal of common variance, although they independently predict performance in *Gf* tasks (Jaeggi, Buschkuhl, Perrig, & Meier, 2010; Kane, Conway, Miura, & Colflesh, 2007). This is in line with findings from training on the n-back task which leads to improvements in *Gf* (Jaeggi, Buschkuhl, Jonides, & Perrig, 2008), but not in measures of working memory capacity (Jaeggi, Buschkuhl, Jonides, & Perrig, 2008; Li et al., 2008). Therefore, we do not know whether training on an n-back task results in transfer to *Gf* due to an improvement in basic working memory processes, or whether there are other processes that are better predictive of such transfer.

**Study 1.** The main goal of Study 1 was to document the results of a correlational study investigating the relationship between the n-back task and selected cognitive tasks chosen so that they might reveal factors that underlie the transfer effect that we have observed by training on the dual n-back task. Based on our own work and Kane's work, we included measures of matrix reasoning and a measure of working memory capacity (Jaeggi, Buschkuhl, Perrig, & Meier, 2010; Kane, Conway, Miura, & Colflesh, 2007). Further, as we were interested in investigating the transfer potential of a simpler n-back task version as well as the dual n-back task, we included both single and dual n-back task versions. There are four reasons to investigate the transfer potential of a single n-back task: First, the dual n-back task is relatively new and not much is known about its constituent processes (Jaeggi et al., 2007; Jaeggi, Buschkuhl, Jonides, & Perrig, 2008; Jaeggi, Schmid, Buschkuhl, & Perrig, 2009; Jaeggi

et al., 2003). Second, the dual n-back task is inherently complex, and so it is not easy to disentangle the underlying processes. Third, the dual n-back task includes an obvious task-switching component (i.e., going back and forth between the two stimulus streams that must be tracked). This task-switching component might contribute to increased reasoning performance because in many matrix reasoning problems, it seems important to be able to switch back and forth between different representations. However, it is not at all clear that task-switching processes are an essential component of *Gf*; thus, if task-switching processes are not critical to matrix reasoning, then a single n-back task should correlate just as well with *Gf* as a dual n-back task. Finally, the dual n-back task is very challenging for participants, thereby restricting its range of application mainly to healthy young adults. We know from our previous research that a frequently used and well-established single n-back task recruits similar neural networks to a dual n-back task (Jaeggi et al., 2003), and also, that single n-back tasks share common variance with *Gf* tasks (e.g. Gray, Chabris, & Braver, 2003; Hockey & Geffen, 2004; Jaeggi, Buschkuhl, Perrig, & Meier, 2010; Kane, Conway, Miura, & Colflesh, 2007). Thus, we investigated the relationship between single n-back performance and measures of *Gf*, and whether and how this relationship is different from that of the dual n-back task and *Gf*. We also investigated the role of working memory capacity, hypothesizing that working memory capacity predicts performance in n-back tasks, however, to a lesser extent than *Gf*.

## 2. Method

### 2.1. Subjects

A total of 104 participants (65 women) with a mean age of 21.3 years ( $SD=2.2$ ) were tested. Subjects were recruited from the student population of the University of Michigan and were paid \$14 per hour for participation.

### 2.2. Tasks and procedure

#### 2.2.1. n-back tasks

**2.2.1.1. Single n-back task.** Participants were shown a sequence of visual stimuli and they had to respond each time the current stimulus was identical to the one presented *n* positions back in the sequence. The stimulus material consisted of 8 random shapes (Vanderplas & Garvin, 1959) which we have used previously (Jaeggi et al., 2003). The shapes were all shown in yellow and presented centrally on a black background for 500 ms each, followed by a 2500 ms interstimulus interval. Participants were required to press a pre-defined key for targets, and their response window lasted from the onset of the stimulus until the presentation of the next stimulus (3000 ms); no response was required for non-targets. Participants were tested on 2-, 3-, and 4-back levels in that order, with each level presented for 3 consecutive blocks, resulting in a total of 9 blocks. A block consisted of 20 + *n* stimuli and contained 6 targets and 14 + *n* non-targets each. The dependent measure was the proportion of hits minus false alarms averaged over all n-back levels.

**2.2.1.2. Dual n-back task.** In contrast to the single n-back task, participants were required to respond to two independent

streams of stimuli, a visual one and an auditory one. We used 8 spatial positions for the visual modality, and 8 letters for the auditory modality (cf. Jaeggi et al., 2007; Jaeggi, Buschkuhl, Jonides, & Perrig, 2008; Jaeggi, Schmid, Buschkuhl, & Perrig, 2009). Participants were required to press a key whenever the currently presented square was at the same position as the one *n* stimuli back in the series, and another key whenever the presented letters matched the one that was presented *n* stimuli back in the sequence. No responses were required for non-targets. The value of *n* was always the same for visual and auditory stimuli. There were 6 auditory and 6 visual targets per block of trials (4 appearing in only one modality at a time, and 2 appearing in both modalities at the same time; i.e. targets could occur in either one modality stream only, or in both modality streams simultaneously), and their positions were determined randomly. Otherwise, the procedure, timing, number of blocks, and levels were similar to the single n-back task. The dependent measure was the proportion of hits minus false alarms averaged over both modalities and all n-back levels.

#### 2.2.2. Working memory task

We used the automated version of the operation span task (OSPAN) as a complex measure of WMC (Kane et al., 2004; Unsworth, Heitz, Schrock, & Engle, 2005). The task requires participants to recall a sequence of stimuli in the correct order in addition to completing a distracting processing task (cf. Conway et al., 2005). We presented three sets of stimuli per set size (i.e., the number of stimuli to be recalled), and the set sizes ranged from 3 to 7. The score, i.e. the sum of all perfectly recalled sets, served as a dependent measure representing complex working memory span (Unsworth, Heitz, Schrock, & Engle, 2005).

#### 2.2.3. Fluid intelligence tasks

We assessed *Gf* by using two different matrix reasoning tasks, either the A or B version of the Bochumer Matrices Test (BOMAT; 29 items; Hossiep, Turck, & Hasella, 1999), and the even or the odd items of Raven's Advanced Progressive Matrices (APM; 18 items; Raven, 1990), both counterbalanced. Both tests were given with a time restriction, a procedure adopted by many researchers (e.g. Jaeggi, Buschkuhl, Jonides, & Perrig, 2008; Kane, Conway, Miura, & Colflesh, 2007; Kane et al., 2004; Salthouse, 1993; Salthouse, Atkinson, & Berish, 2003; Unsworth & Engle, 2005; Unsworth, Heitz, Schrock, & Engle, 2005). The reasons for choosing time-restricted versions were two: First, administering the APM with the standard time restriction or untimed usually results in ceiling performance for a considerable number of participants in our labs. With the BOMAT, ceiling performance is less of an issue, but especially for the BOMAT, the standard testing time is rather long; thus, our second reason for short time limits was to keep total testing time as short as possible. It should be noted though that scores in timed versions of the APM are nicely predictive of scores in untimed versions (Freason & Eysenck, 1986; Hamel & Schmittmann, 2006; Heron & Chown, 1967; Salthouse, 1993; Unsworth & Engle, 2005).

After several practice trials (10 items for the BOMAT, 2 items from Set I for the APM), participants were allowed to work for 10 min on the BOMAT, and 10 min on the APM

**Table 1**

Mean, standard deviation (SD), range, and reliability estimates (Cronbach's  $\alpha$ ) for each of the used tasks.

|                                 | Mean  | SD    | Range      | Reliability |
|---------------------------------|-------|-------|------------|-------------|
| n-back                          |       |       |            |             |
| Single n-back<br>(mean 2–4back) | 0.45  | 0.19  | –0.01–0.89 | 0.79        |
| Dual n-back<br>(mean 2–4back)   | 0.45  | 0.16  | 0.07–0.82  | 0.91        |
| Working memory                  |       |       |            |             |
| OSPAN                           | 55.58 | 13.92 | 20–75      | 0.73        |
| Fluid intelligence              |       |       |            |             |
| Raven's APM                     | 10.88 | 2.87  | 4–18       | 0.74        |
| BOMAT                           | 7.44  | 2.42  | 2–12       | 0.58        |

Note:  $N = 104$ .

(Set II). The number of correct solutions provided in this time limit was used as the dependent variable.

#### 2.2.4. Analyses

We used SPSS (Release 15) for all our data analyses. The data were examined to determine whether they fulfilled the assumptions necessary for multiple linear regression: We checked for univariate and multivariate normality, multicollinearity, heteroscedasticity, independent errors, and normally distributed errors and we found all these to be in appropriate ranges. We calculated several multiple linear regression models in order to determine which variables were best suited to predict n-back performance, and also, which variables were most predictive for performance in the BOMAT and APM. We used a direct method by entering all variables into the model and then removing the non-significant predictors one after another until only significant predictors remained. With the significant predictors, we ran a forward stepwise analysis in order to determine the individual contribution of each predictor.

### 3. Results

Means, standard deviations, and reliability estimates for all measures are reported in Table 1. The Pearson's correlations among the variables used for the regression analyses are reported in Table 2.

Overall, our data revealed a strong relationship between the single and dual n-back tasks with a correlation of  $r = .72$ , indicating that the two versions share a considerable amount of common variance (see Table 2). Further, the correlation of both n-back tasks with the matrices tasks were stronger than the correlations of the n-back and the working memory task, although the difference only reached statistical significance

**Table 2**

Pearson correlation coefficients for the measures used in the regression models.

|                      | Single n-back | Dual n-back | OSPAN | APM    | BOMAT |
|----------------------|---------------|-------------|-------|--------|-------|
| <i>Single n-back</i> |               |             |       |        |       |
| Dual n-back          | 0.72**        |             |       |        |       |
| OSPAN                | 0.21*         | 0.26**      |       |        |       |
| Raven's APM          | 0.44**        | 0.41**      | 0.24* |        |       |
| BOMAT                | 0.53**        | 0.40**      | 0.05  | 0.42** |       |

Note:  $N = 104$ ; \*\* $p < .01$ ; \* $p < .05$ .

**Table 3**

Direct multiple regression model for the single n-back task as outcome variable.

|                         | B     | SE B | $\beta$ |
|-------------------------|-------|------|---------|
| Model 1 ( $R^2 = .35$ ) |       |      |         |
| Constant                | –0.05 | 0.08 |         |
| OSPAN                   | 0.00  | 0.00 | 0.14    |
| Raven's APM             | 0.01  | 0.01 | 0.23*   |
| BOMAT                   | 0.03  | 0.01 | 0.42**  |
| Model 2 ( $R^2 = .33$ ) |       |      |         |
| Constant                | 0.03  | 0.06 |         |
| Raven's APM             | 0.02  | 0.01 | 0.26*   |
| BOMAT                   | 0.03  | 0.01 | 0.41**  |

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

for the single n-back task (*Single n-back*: BOMAT vs OSPAN: Steiger's  $Z = 2.92$ ,  $p < .01$ ; APM vs OSPAN:  $Z = 2.04$ ,  $p < .05$ ; *Dual n-back*: BOMAT vs OSPAN:  $Z = 1.20$ ,  $p = \text{ns.}$ ; APM vs OSPAN:  $Z = 1.12$ ,  $p = \text{ns.}$ ).

In order to predict single n-back task performance, we first entered all 3 predictors (OSPAN, APM, and BOMAT) into a regression model. The model is reported in Table 3 showing that single n-back performance was best predicted by the BOMAT. The model resulting from the stepwise forward analysis is reported in Table 4, showing that both the APM and the BOMAT, but not the working memory measure significantly predicted n-back performance.

As shown in Table 5, we entered the same predictors as before, but with dual n-back performance as the outcome measure. Similar to single n-back performance, dual n-back performance was best predicted by the BOMAT. This time however, all three predictors contributed significantly to the variance in the model. The results of the stepwise forward analysis are reported in Table 6.

In Tables 7 and 8, we show the results for the regression analyses in which we predict performance in the two matrix

**Table 4**

Stepwise forward analysis for the single n-back task as outcome variable.

|             | B    | SE B | $\beta$ |
|-------------|------|------|---------|
| Step 1      |      |      |         |
| Constant    | 0.15 | 0.05 |         |
| BOMAT       | 0.04 | 0.01 | 0.52*** |
| Step 2      |      |      |         |
| Constant    | 0.03 | 0.06 |         |
| OSPAN       | 0.03 | 0.01 | 0.41*** |
| Raven's APM | 0.02 | 0.01 | 0.26**  |

Note:  $R^2 = .28$  for Step 1;  $\Delta R^2 = .05$  for Step 2 ( $p$ 's  $< .01$ ); \*\*\* $p < .001$ ; \*\* $p < .01$ .

**Table 5**

Direct multiple regression model for the dual n-back task as outcome variable.

|               | B    | SE B | $\beta$ |
|---------------|------|------|---------|
| $(R^2 = .26)$ |      |      |         |
| Constant      | 0.03 | 0.08 |         |
| OSPAN         | 0.00 | 0.00 | 0.19*   |
| Raven's APM   | 0.01 | 0.01 | 0.24*   |
| BOMAT         | 0.02 | 0.01 | 0.29**  |

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



**Table 6**  
Stepwise forward analysis for the dual n-back task as outcome variable.

|             | B    | SE B | $\beta$ |
|-------------|------|------|---------|
| Step 1      |      |      |         |
| Constant    | 0.20 | 0.06 |         |
| Raven's APM | 0.02 | 0.01 | 0.41*** |
| Step 2      |      |      |         |
| Constant    | 0.13 | 0.06 |         |
| Raven's APM | 0.02 | 0.01 | 0.29**  |
| BOMAT       | 0.02 | 0.01 | 0.28**  |
| Step 3      |      |      |         |
| Constant    | 0.03 | 0.07 |         |
| Raven's APM | 0.01 | 0.01 | 0.24*   |
| BOMAT       | 0.02 | 0.01 | 0.29**  |
| OSPAN       | 0.00 | 0.00 | 0.18*   |

Note:  $R^2 = .16$  for Step 1;  $\Delta R^2 = .06$  for Step 2;  $\Delta R^2 = .03$  for Step 3 ( $p < .01$ ); \*\*\* $p < .001$ ; \*\* $p < .01$ ; \* $p < .05$ .

reasoning measures. For both measures, it was the single n-back task alone that accounted for the variance in the *Gf* tasks.

#### 4. Discussion

The findings of Study 1 confirm other findings from the literature (Jaeggi, Buschkuhl, Perrig, & Meier, 2010; Kane, Conway, Miura, & Colflesh, 2007): Consistent with our hypotheses, both n-back task variants were highly correlated, and both were best predicted by *Gf*.

In general, matrix reasoning tasks seem to be better predictors for both the single and the dual n-back tasks than a measure of working memory capacity. As the reliability estimates were appropriate for the n-back tasks, the lack of correlation between the n-back tasks and the measure of working memory capacity cannot be attributed to insufficient reliability (Jaeggi, Buschkuhl, Perrig, & Meier, 2010). Rather, it seems that performance for the two tasks relies on different sources of variance, which might result from the different memory processes that are involved in the two tasks: whereas the n-back task relies on passive recognition processes, performance in working memory capacity tasks requires active and strategic recall processes (Kane, Conway, Miura, & Colflesh, 2007).

Despite the apparent process overlap in single and dual n-back performance, we still observed differential cognitive processes mediating performance in single or dual n-back tasks: whereas single n-back performance was mostly predicted

**Table 7**  
Direct multiple regression model for Raven's APM as outcome variable.

|                         | B    | SE B | $\beta$ |
|-------------------------|------|------|---------|
| Model 1 ( $R^2 = .23$ ) |      |      |         |
| Constant                | 6.05 | 1.14 |         |
| Single n-back           | 4.56 | 1.96 | 0.29*   |
| OSPAN                   | 0.03 | 0.02 | 0.13    |
| Dual n-back             | 2.79 | 2.24 | 0.16    |
| Model 2 ( $R^2 = .21$ ) |      |      |         |
| Constant                | 6.33 | 1.12 |         |
| Single n-back           | 6.28 | 1.40 | 0.41*   |
| OSPAN                   | 0.03 | 0.02 | 0.15    |
| Model 3 ( $R^2 = .19$ ) |      |      |         |
| Constant                | 7.83 | 0.67 |         |
| Single n-back           | 6.78 | 1.38 | 0.44*** |

Note: \*\*\* $p < .001$ ; \* $p < .05$ .

**Table 8**  
Direct multiple regression model for the BOMAT as outcome variable.

|                         | B     | SE B | $\beta$ |
|-------------------------|-------|------|---------|
| Model 1 ( $R^2 = .28$ ) |       |      |         |
| Constant                | 4.82  | 0.93 |         |
| Single n-back           | 6.50  | 1.59 | 0.50*** |
| OSPAN                   | -0.01 | 0.02 | -0.07   |
| Dual n-back             | 0.87  | 1.82 | 0.06    |
| Model 2 ( $R^2 = .28$ ) |       |      |         |
| Constant                | 4.91  | 0.90 |         |
| Single n-back           | 7.04  | 1.13 | 0.54*** |
| OSPAN                   | -0.01 | 0.02 | -0.07   |
| Model 3 ( $R^2 = .28$ ) |       |      |         |
| Constant                | 4.35  | 0.54 |         |
| Single n-back           | 6.86  | 1.10 | 0.52*** |

Note: \*\*\* $p < .001$ .

by matrix reasoning, dual-task performance was mediated by working memory capacity in addition to *Gf*. Further, the single n-back task was the only predictor for both matrix reasoning measures. Considering the rationale that transfer is more likely to happen for tasks that share considerable variance, we can conclude that training on both single and dual n-back tasks should yield transfer to matrix reasoning, but that transfer to working memory capacity should be less likely, especially in the case of single n-back training. Considering the variance explained in the matrix reasoning tasks, our data suggest that the single n-back task might be an even better training vehicle than the dual n-back task.

**Study 2.** In Study 2, we tested the implication of the findings from Study 1 by investigating the transfer potential of a single n-back task to measures of *Gf* as compared to training with a dual n-back task, and also whether transfer occurs to a measure of working memory capacity. Based on the rationale and the results from Study 1, we hypothesized that both training regimens should yield transfer to both matrices tasks, but that the effect on working memory capacity should be smaller than the effect on *Gf* due to the smaller inter-correlations evident in Study 1. We trained 46 undergraduate students with either a single or a dual n-back task over the course of one month, assessing their performance on trained tasks, on variants of these tasks using different stimulus material, on a measure of working memory, and on the two matrices tasks that we used in Study 1. To control for re-test effects, the performance of the trained groups was compared to a control group ( $N = 43$ ) that completed the same transfer tasks in a pre- and post-test session, but that was not trained between the two testing sessions.

## 5. Method

### 5.1. Participants

Ninety-nine undergraduates (mean age = 19.4 years;  $SD = 1.5$ ; 76 women) from the National Taiwan Normal University in Taipei volunteered to take part in the study. Fifty-two (41 women) were assigned to the control group and 47 (35 women) were assigned to the experimental group. In return for participation, participants earned course credit. In addition, the training groups received NT\$ 600 (about US\$20) as well as the training software after study completion. After

the pretest, participants in the experimental group were assigned to either the single or the dual n-back training intervention. The two groups were matched by using the software 'Match' (Van Casteren & Davis, 2007) based on the following criteria: age, gender, pre-test performance in the n-back baseline tasks (single and dual), as well as pre-test performance in one of the matrices tasks (BOMAT). One participant from the dual-task training regimen dropped out after a few training sessions and these data were discarded from further analyses. The final single n-back training group consisted of 21 participants (mean age = 19.0 years,  $SD = 1.5$ ; 17 women), and the dual n-back group consisted of 25 participants (mean age = 19.1 years,  $SD = 1.2$ ; 18 women). Since there were also drop-outs in the no-contact control group, we had a final  $n$  of 43 participants (mean age = 19.4,  $SD = 1.0$ ; 34 women) in this group.

## 5.2. Training and transfer tasks

### 5.2.1. Training

We used two n-back interventions, an adaptive dual n-back task that we used previously (Jaeggi, Buschkuhl, Jonides, & Perrig, 2008), and an adaptive single n-back task using only visospatial material (Jaeggi, Schmid, Buschkuhl, & Perrig, 2009).

**5.2.1.1. Dual n-back task.** For the adaptive dual n-back task, we used the same visual stimuli as used in Study 1. However, for the auditory stimuli, we used 8 syllables of the Mandarin phonetic system instead of letters from the Latin alphabet. In order to match the task to each participant's ability, the level of difficulty was varied by changing the level of  $n$  (Jonides et al., 1997): After each block, each participant's individual performance was analyzed, and in the following block, the level of  $n$  was adapted accordingly: If the participant made fewer than 3 mistakes per modality, the level of  $n$  increased by 1. It was decreased by 1 if more than 5 mistakes were made, and in all other cases,  $n$  remained unchanged. One training session comprised 15 blocks consisting of  $20 + n$  trials resulting in a daily training time of approximately 17–20 min. Participants were given feedback concerning their performance after each block (percent correct for each modality). In addition, participants received feedback at the end of each training session consisting of their performance score for each session that had been completed, as well as a curve representing the scores of a reference group consisting of all participants who completed comparable training in our laboratory in other experiments prior to this one.

**5.2.1.2. Single n-back task.** As a second intervention, we used an adaptive single-task version of the n-back task requiring the processing of the visuospatial modality only. Everything else (i.e. training length, adaptivity of the level of the n-back task based on subjects' performance, and feedback) was the same as in the dual n-back intervention described above.

### 5.2.2. Transfer tasks

**5.2.2.1. n-back.** In order to assess baseline n-back performance, we used the same single n-back task with random shapes that we used in Study 1 ( $n$ -back levels 2, 3, and 4). As

none of the training groups had trained with stimuli of this type, we used this task to assess near transfer.

**5.2.2.2. Working memory span.** We used the automated OSPAN as used in Study 1.

**5.2.2.3. Matrix reasoning.** As in Study 1, we administered two standard matrix reasoning tests in order to measure  $G_f$ , the short version of the Bochumer Matrizen-Test (BOMAT; Hossiep, Turck, & Hasella, 1999), and the Raven's Advanced Progressive Matrices (APM; Raven, 1990). Parallel versions (A and B in the BOMAT consisting of 29 items each, as well as the odd and even items in the APM consisting of 18 items each) were used in order to prevent participants from getting the same items in pre- and post-test (te Nijenhuis, van Vianen, & van der Flier, 2007). The order of the versions was counter-balanced. We had slightly more testing time available than in Study 1, thus, after some practice trials (10 items for the BOMAT, 3 items for the APM, i.e. from set I), participants were allowed to work for 16 min on the BOMAT, and for 11 min on the APM. The dependent measure was the number of correctly solved problems within this time limit.

### 5.2.3. Procedure

To assess the change in cognitive performance, all participants were pre- and post-tested at the same interval (3 days before the start of training, and 3 days after training completion; with a 5-week interval for the no-contact control group) with the tests described above. Participants were tested in groups of 30–40 individuals, and they were divided into two groups of 15–20 students each. One group (consisting of participants from all 3 intervention groups) first completed the n-back task and the OSPAN followed by the matrix reasoning tasks (APM, BOMAT), whereas the second group started with the matrices tasks first and completed n-back and OSPAN afterwards.

After pre-testing, participants in the two training groups trained on a daily basis, five times per week (not on weekends) for a period of 4 weeks. Participants trained in small groups of 10–15 students in a computer laboratory located at the University.

## 6. Results

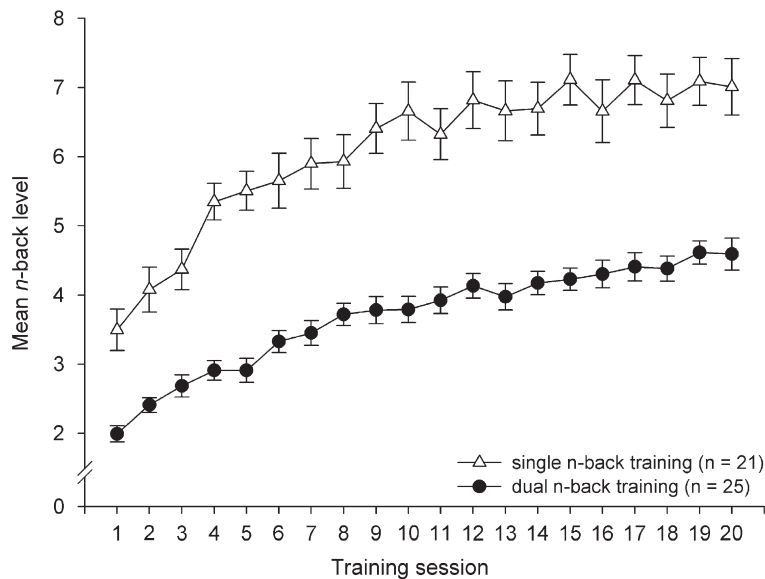
Descriptive data for each of the intervention groups and test session are reported in Table 9. Note that there were no significant group differences at pre-test in any of the criterion measures.

### 6.1. Training data

First, we investigated specific training effects and tested whether there are differential training effects as a function of training group. As illustrated in Fig. 1, both training groups improved their performance over the four weeks of training, but the single-task group trained at a higher n-back level, reflecting the lower complexity of the single task compared to the dual task.

**Table 9**  
Descriptive data for the transfer measures as a function of group.

|                              | Pre-test |       |       |      |     | Post-test |       |      |      | Effect size (Cohen's d) |
|------------------------------|----------|-------|-------|------|-----|-----------|-------|------|------|-------------------------|
|                              | N        | Mean  | SD    | Min  | Max | Mean      | SD    | Min  | Max  |                         |
| Single n-back training group |          |       |       |      |     |           |       |      |      |                         |
| Single n-back                | 20       | .42   | .15   | .18  | .68 | .64       | .18   | .19  | .91  | 1.33                    |
| Operation span               | 21       | 57.60 | 12.83 | 13   | 75  | 55.14     | 13.91 | 7    | 75   | -0.18                   |
| Raven's APM                  | 21       | 11.33 | 2.28  | 7    | 14  | 12.81     | 2.27  | 9    | 18   | 0.65                    |
| BOMAT                        | 21       | 11.48 | 3.11  | 6    | 16  | 13.67     | 3.17  | 8    | 20   | 0.70                    |
| Dual n-back training group   |          |       |       |      |     |           |       |      |      |                         |
| Single n-back                | 25       | .37   | .17   | .00  | .72 | .64       | .18   | .30  | 1.00 | 1.54                    |
| Operation span               | 25       | 57.79 | 14.46 | 14   | 75  | 56.92     | 9.50  | 37   | 71   | -0.07                   |
| Raven's APM                  | 25       | 11.32 | 1.93  | 8    | 15  | 13.36     | 2.22  | 9    | 18   | 0.98                    |
| BOMAT                        | 25       | 10.88 | 2.60  | 5    | 17  | 12.28     | 3.09  | 8    | 19   | 0.49                    |
| No-contact control group     |          |       |       |      |     |           |       |      |      |                         |
| Single n-back                | 41       | .33   | .17   | -.10 | .63 | .37       | .22   | -.12 | .76  | 0.20                    |
| Operation span               | 40       | 52.73 | 11.93 | 21   | 75  | 55.50     | 12.36 | 29   | 75   | 0.23                    |
| Raven's APM                  | 43       | 11.58 | 2.60  | 2    | 17  | 11.81     | 2.27  | 6    | 17   | 0.09                    |
| BOMAT                        | 43       | 10.79 | 2.50  | 6    | 16  | 11.44     | 2.58  | 8    | 19   | 0.26                    |



**Fig. 1.** Specific training effects. Performance increase in the trained task shown separately for each training group. For each session, the mean n-back level achieved by the participants is presented. Error bars represent the standard error of the mean.

6.2. Near transfer effects

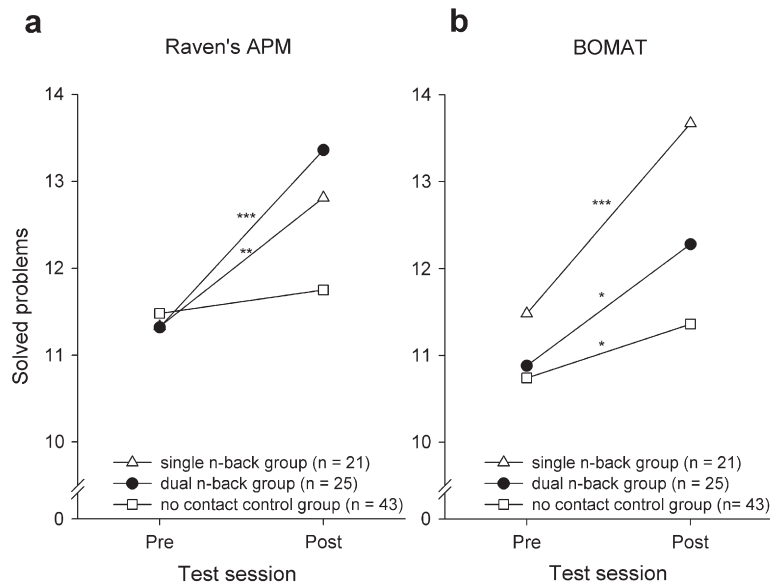
In order to assess near transfer effects, we calculated repeated-measures ANOVAs with session (pre vs post) as a within-subjects factor, and intervention (dual n-back, single n-back, control) as a between-subject factor for the near transfer measure, i.e. for the non-trained single n-back task. For logistical reasons, not all participants were able to complete this task, and the final sample size is indicated in Table 9. Performance ( $P_r$ ) was calculated as a composite score consisting of the averaged 2-back, 3-back, and 4-back accuracy. Our results showed a highly significant session  $\times$  intervention interaction ( $F(2,82) = 15.74; p < .001, \eta_p^2 = .28$ )<sup>1</sup>.

<sup>1</sup> Although there were no significant group differences at pre-test, we ran additional analyses of covariance controlling for pre-test performance for each of the transfer measures which yielded similar results.

Pairwise comparisons showed that the performance gain was largest in the dual n-back training group ( $t(24) = 8.42; p < .001; d = 1.54$ ), followed by the single n-back training group ( $t(18) = 4.61; p < .001; d = 1.23$ ), however there was no difference in gain between the two training groups ( $d = .34$ ). The training groups improved more than the no-contact control group ( $d > 1$ ) which showed no significant performance increase in this task ( $t(40) = 1.72; p = .09; d = .19$ ).

6.3. Operation span task

There was no significant session  $\times$  intervention interaction ( $F(2,82) = 2.11; p = ns., \eta_p^2 = .05$ ). Note that none of the three groups showed significant performance differences between pre- and post-test (all  $t < 2$ ).



**Fig. 2.** Transfer effects on matrix reasoning. Mean problems solved in each session illustrated for each group and *Gf* test. Note that there were no significant group differences at pretest.

#### 6.4. Matrix reasoning tasks

In order to assess transfer effects on matrix reasoning, we calculated repeated-measures ANOVAs with session (pre vs post) as a within-subject factor, and intervention (dual n-back, single n-back, control) as a between-subject factor separately for each matrix task (BOMAT and APM; see Fig. 2). The results yielded significant intervention  $\times$  session interactions for each task (BOMAT:  $F(2,85) = 3.45$ ;  $p < .05$ ,  $\eta_p^2 = .08$ ; APM:  $F(2,85) = 5.03$ ;  $p < .01$ ,  $\eta_p^2 = .11$ )<sup>2</sup>. Pairwise comparisons showed that both training groups significantly improved performance in both tasks (*dual n-back training group*: BOMAT:  $t(24) = 2.38$ ;  $p < .05$ ;  $d = 0.49$ ; APM:  $t(24) = 4.58$ ;  $p < .001$ ;  $d = .98$ ; *single n-back training group*: BOMAT:  $t(20) = 5.04$ ;  $p < .001$ ;  $d = 0.70$ ; APM:  $t(20) = 3.13$ ;  $p < .01$ ;  $d = 0.65$ ), and there was no difference in gain between the two groups in either of the two tasks ( $d < .32$ ). In contrast, the control group only showed a marginally significant re-test effect on the BOMAT, but not in the APM (BOMAT:  $t(42) = 2.08$ ;  $p < .05$ ;  $d = 0.26$ ; APM:  $t(42) = 0.61$ ;  $p = ns.$ ;  $d = 0.10$ ).

## 7. Discussion

The goal of Study 2 was to investigate whether a single n-back intervention is a useful alternative to the complex dual n-back task that we used previously to demonstrate a transfer effect on tests of *Gf*. We based our assumption regarding the effectiveness of the single n-back task on our earlier findings showing that dual and single n-back tasks recruit similar neural networks (Jaeggi et al., 2003), and on the fact that single n-back performance correlates with *Gf* as well as dual n-back performance (Experiment 1; Jaeggi, Buschkuhl, Perrig, & Meier, 2010).

Concerning the near transfer results, both intervention groups improved their performance almost equally in the

random-shape variant of the baseline single n-back task in spite of the fact that neither of the training groups had trained with these stimuli. In contrast, there was only a negligible performance increase for the control group. Thus, both intervention groups were able to generalize their n-back training performance to stimulus material and presentation format which was unfamiliar to them, providing evidence that the intervention had an effect on some general underlying processes involved in n-back performance, rather than just building up a very task- and material-specific skill.

As predicted, we found no transfer effects to a measure of working memory capacity for the single n-back task group. We also found no transfer for the dual n-back task group even though Study 1 showed that dual n-back performance was partially predicted by working memory capacity. We note, however, that the correlation of OSPAN with dual-task performance was considerably smaller than the correlation of this task with both matrices tasks.

This finding might be surprising given that both the OSPAN and the n-back tasks are considered WM tasks. However, previous research has shown that these tasks do not share considerable common variance, although they both seem to predict variance in *Gf* tasks (e.g. Jaeggi, Buschkuhl, Perrig, & Meier, 2010; Kane, Conway, Miura, & Colflesh, 2007). The lack of correlation between the two WM tasks most likely results from the fact that there are different processes involved in the two tasks: whereas the main processes that drive performance in the n-back tasks are familiarity- and recognition-based discrimination processes (Oberauer, 2005; Smith & Jonides, 1998), complex WM span tasks, such as the OSPAN, require active recall processes rather than recognition. This pattern is also consistent with our prior findings of no effect on complex span (Jaeggi et al., 2008). Indeed, Li et al. (2008) have even reported a significant performance decrease after single n-back training in a related complex span measure (rotation span). Thus, although there was no significant performance difference between pre- and post-test in either of the groups in the present study, one

<sup>2</sup> The task order (A/odd in pre then B/even in post, or the other way round) was entered as covariate in the ANOVAs.



could speculate that n-back training somehow interferes with performance in complex span measures as participants might rely more on recognition instead of recall processes at post-test which might prevent any performance gain. Since both WM tasks seem to be related via their relationship to *Gf* measures, though, one might argue that one could also train on complex span measures in order to get transfer to *Gf*. However, [Chein and Morrison \(2010\)](#) trained their subjects on complex-span measures, but they did not find any transfer to matrix reasoning.

But most interestingly, our results show transfer effects in both matrix reasoning tasks after training. This replicates our prior results ([Jaeggi, Buschkuhl, Jonides, & Perrig, 2008](#)), but it also extends our findings by showing that a) the transfer effect was present in more than just one *Gf* task, and b), that it was also obtained by training on a single n-back task. Although, matrices tasks like the APM and the BOMAT are regarded as prototypical tasks to measure *Gf*, with the APM representing the task with the highest *Gf* loading (e.g. [Gray & Thompson, 2004](#); [Kane & Engle, 2002](#); [Snow, Kyllonen, & Marshalek, 1984](#)), they are only an approximation of *Gf*. Thus, we acknowledge that in order to capture the full range of *Gf*, there should be testing with a more exhaustive battery of tasks. That is, the current data do not allow us to firmly determine whether the gains in matrix reasoning in our study represent real gains in *Gf*, or whether they emerge because some aspects of the training allowed participants to better deal with the specific content of the matrices tasks themselves. A related issue is whether we captured *Gf* with the timed versions of the matrices tasks we used, or whether we “just” improved some task-specific abilities. In a time-limited version of these tests, most subjects do not reach the end of the test; this is especially true of the BOMAT. [Moody \(2009\)](#) has argued that restricting the tests to just the early items leaves out the items that have higher *Gf* loadings. This issue has been addressed before by other researchers who investigated whether there are differential age effects or working memory involvement in the different parts of the APM ([Salthouse, 1993](#); [Unsworth & Engle, 2005](#)). These studies found no evidence for differential processes in the various items of the APM, at least for the first three quartiles of the task; thus, it seems unlikely that a subset of items in the APM measures something different than *Gf*. In our own data, the transfer effects were actually *more* pronounced for the second half of the test in the APM, which is reflected in a significant 3-way interaction (session  $\times$  APM-part  $\times$  intervention;  $F(2,86) = 5.31$ ;  $p < .01$ ;  $\eta_p^2 = .11$ ). In the BOMAT, we observed no differential transfer effects for the earlier vs later items ( $F(2,86) = .64$ ;  $p = ns.$ ;  $\eta_p^2 = .02$ ). Thus, if there are any differences in *Gf* loading in the various parts of the matrices tasks, the present data suggest that the transfer effects are roughly equivalent for the parts of the test that are claimed to have higher vs lower *Gf* loadings.

Because we limited our intervention to versions of the n-back task, one might wonder whether one could improve matrix reasoning with any kind of cognitive training. That is, is there something specific to components of the n-back task that makes it unique or unusual? Of course, this remains an empirical issue, but the sparse reports of transfer after cognitive training in general suggest that the transfer effects obtained in the present study do not represent a general effect of transfer no matter the training task (e.g. [Barnett & Ceci, 2002](#); [Salomon](#)

& [Perkins, 1989](#); [Verhaeghen, Marcoen, & Goossens, 1992](#); [Zelinski, 2009](#)).

A limitation of Study 2 is that we used a no-contact control group. One might argue that the training groups were simply more motivated because they received more experimenter attention and therefore showed more transfer (Hawthorne effect). However, if the transfer was just due to motivational factors, the training groups should have outperformed the control group in all transfer measures. Thus, the lack of improvement in WM capacity for the training groups might be taken as a case against an unspecific effect of mere motivation or arousal. Nevertheless, future studies should replicate the present effects by carefully selecting an appropriate active control group in order to rule out any placebo or Hawthorne effects ([Buschkuhl & Jaeggi, 2010](#); [Shipstead, Redick, & Engle, 2010](#)).

### 7.1. General discussion

In the two studies we aimed to explore the relationship between n-back tasks and ability measures such as WMC and *Gf*, and we also evaluated the transfer potential of a single n-back task to these outcome measures. In the first study, we demonstrated that dual and single n-back task performance is equally well correlated with performance on two different tests measuring *Gf*, whereas the correlation of these n-back tasks with a task assessing working memory capacity was much smaller. Based on these results, we were led to test the hypotheses that training on a single n-back task might yield the same improvement in *Gf* as training on a dual n-back task, and that there should be less transfer to a measure of working memory capacity. Thus, in Study 2, we investigated transfer effects on working memory capacity and *Gf* by training participants on either a single or on a dual n-back task. Consistent with our hypotheses, our results showed that although there was no transfer on a measure of working memory capacity, both training groups improved more on *Gf* than the no-contact control group. This pattern replicates our prior results ([Jaeggi, Buschkuhl, Jonides, & Perrig, 2008](#)), but also goes beyond them by demonstrating that there is transfer on two different matrix reasoning tasks, and that single n-back training seems to be equally effective as dual n-back training.

As there are still many unknown factors concerning training and transfer ([Willis & Schaie, 2009](#)), one aim of the present study was to shed light on the underlying mechanisms that drive the transfer effects that we previously have found. We outlined several features that we think are important for training in order to get transfer. First of all, we proposed that the training and transfer task should engage overlapping processes. Our differential results showing transfer to matrix reasoning but not OSPAN support this assumption, because the data from Study 1 as well as earlier studies show that n-back and measures of working memory capacity do not share much common variance ([Jaeggi, Buschkuhl, Perrig, & Meier, 2010](#); [Kane, Conway, Miura, & Colflesh, 2007](#)). In contrast, n-back performance and matrix reasoning share a great deal of common variance; consequently, there was transfer as a result of n-back training. We also proposed that it is important that participants only minimally learn task-specific strategies in order to prevent specific skill acquisition. We think that besides

the transfer to matrix reasoning, the improvement in the near transfer measure provides additional evidence that the participants trained on task-underlying processes rather than relying on material-specific strategies. Further, we proposed that it is important to maximally stress the information processing system by using a very complex training paradigm, in our case, a dual variant of the n-back task. However, the results of both studies indicate that this is not necessary in order to get transfer: Study 1 showed that the single n-back task is equally related to matrix reasoning, and Study 2 showed that single n-back training yields transfer to our *Gf* tasks as well. Thus, using a dual task for training does not seem to be necessary in order to obtain transfer to matrix reasoning.

Of course, there are many more research questions to be addressed. For example, we still do not know whether it is necessary to tailor the level of difficulty of the training task to the changing performance level of the subjects as they improve. We assume that adaptivity might be a crucial factor for transfer because it ensures that the participants' executive control system is sufficiently taxed to prevent automatic processing. Further, we do not know how general the transfer effects are, whether they extend to measures of daily living or academic or professional success, how long-lasting the effects are, what executive components are the critical ones to train, and whether there are inter-individual differences that moderate training and transfer. These and other questions await further research. Nonetheless, our current work has made progress in examining one of the critical issues raised by our initial study: There is no need to rely on a dual task to achieve improvement in matrix reasoning. Therefore, our findings lead to an important message for training research: In that the dual n-back task is quite complex in the processes it engages, it is not an ideal task for many participant groups other than young adults. The present results provide empirical support that it is not necessary to train with this complex task in order to improve *Gf*; thus, our results open a wider range of application for our training approach in that the single n-back task can be used for participants such as children or older adults who would find the dual n-back task too complex and too taxing. Furthermore, it makes the investigation of the processes in training and transfer more accessible because the processes engaged by the single n-back task are better understood than the ones in dual n-back tasks.

## References

- Baltes, P. B., Staudinger, U. M., & Lindenberger, U. (1999). Lifespan psychology: theory and application to intellectual functioning. *Annual Review of Psychology*, 50, 471–507.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 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(4), 765–777.
- Buschkuhl, M., & Jaeggi, S. M. (2010). Improving intelligence: a literature review. *Swiss Medical Weekly*, 140(19–20), 266–272.
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97(3), 404–431.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: a critical experiment. *Journal of Educational Psychology*, 54(1), 1–22.
- Chein, J. M., & Morrison, A. B. (2010). Expanding the mind's workspace: training and transfer effects with a complex working memory task. *Psychonomic Bulletin & Review*, 17(2), 193–199.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: a methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769–786.
- Dahlin, E., Neely, A. S., Larsson, A., Backman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science*, 320(5882), 1510–1512.
- Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35(1), 13–21.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, 128(3), 309–331.
- Ericsson, K. A., & Delaney, P. F. (1998). Working memory and expert performance. In R. Logie, & K. J. Gilhooly (Eds.), *Working Memory and Thinking* (pp. 93–114). Hillsdale, NJ: Erlbaum.
- Frearson, W., & Eysenck, H. J. (1986). Intelligence, reaction time (RT), and a new odd-man-out RT paradigm. *Personality and Individual Differences*, 7, 807–818.
- Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, 6(3), 316–322.
- Gray, J. R., & Thompson, P. M. (2004). Neurobiology of intelligence: science and ethics. *Nature Reviews Neuroscience*, 5(6), 471–482.
- Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: a new hypothesis. *Trends in Cognitive Sciences*, 11(6), 236–242.
- Hamel, R., & Schmittmann, V. D. (2006). The 20-minute version as a predictor of the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement*, 66(6), 1039–1046.
- Haworth, C. M., Wright, M. J., Luciano, M., Martin, N. G., de Geus, E. J., van Beijsterveldt, C. E., et al. (2009). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry*, 1–9.
- Heron, A., & Chown, S. M. (1967). *Age and Function*. Boston: Little-Brown.
- Hockey, A., & Geffen, G. (2004). The concurrent validity and test-retest reliability of a visuospatial working memory task. *Intelligence*, 32, 591–605.
- Hossiep, R., Turck, D., & Hasella, M. (1999). *Bochumer Matrizen-test. BOMAT-advanced-short version*. Göttingen: Hogrefe.
- Jaeggi, S. M., Buschkuhl, M., Etienne, A., Ozdoba, C., Perrig, W. J., & Nirkko, A. C. (2007). On how high performers keep cool brains in situations of cognitive overload. *Cognitive, Affective & Behavioral Neuroscience*, 7(2), 75–89.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 105(19), 6829–6833.
- Jaeggi, S. M., Buschkuhl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, 18(4), 394–412.
- Jaeggi, S. M., Schmid, C., Buschkuhl, M., & Perrig, W. J. (2009). Differential age effects in load-dependent memory processing. *Neuropsychology, Development, and Cognition. Section B: Aging, Neuropsychology and Cognition*, 16(1), 80–102.
- Jaeggi, S. M., Seewer, R., Nirkko, A. C., Eckstein, D., Schroth, G., Groner, R., et al. (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(2), 210–225.
- Jarrold, C., & Towse, J. N. (2006). Individual differences in working memory. *Neuroscience*, 139(1), 39–50.
- Jensen, A. R. (1981). Raising the IQ: The Ramey and Haskins study. *Intelligence*, 5(1), 29–40.
- Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., et al. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience*, 9(4), 462–475.
- Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the N-back task: a question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3), 615–622.
- Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychonomic Bulletin & Review*, 9(4), 637–671.
- 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(2), 189–217.
- Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlstrom, K., et al. (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(2), 177–186.
- Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, 24(6), 781–791.

- 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(4), 731–742.
- Neisser, U., Boodoo, G., Bouchard, T. J. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: knowns and unknowns. *The American Psychologist*, 51(2), 77–101.
- Oberauer, K. (2005). Binding and inhibition in working memory: individual and age differences in short-term recognition. *Journal of Experimental Psychology: General*, 134(3), 368–387.
- Oberauer, K., Lange, E., & Engle, R. W. (2004). Working memory capacity and resistance to interference. *Journal of Memory and Language*, 51(1), 80–96.
- Persson, J., & Reuter-Lorenz, P. A. (2008). Gaining control: training executive function and far transfer of the ability to resolve interference. *Psychological Science*, 19(9), 881–889.
- Raven, J. C. (1990). *Advanced Progressive Matrices. Sets I, II*. Oxford: Oxford University Press.
- Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence*, 35, 83–92.
- 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.
- Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: rethinking mechanisms of a neglected phenomenon. *Educational Psychologist*, 24(2), 113–142.
- Salthouse, T. A. (1993). Influence of working memory on adult age differences in matrix reasoning. *British Journal of Psychology*, 84(Pt 2), 171–199.
- Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General*, 132(4), 566–594.
- Shipstead, Z., Redick, T. S., & Engle, R. W. (2010). Does working memory training generalize? *Psychologica Belgica*, 50, 245–276.
- Smith, E. E., & Jonides, J. (1998). Neuroimaging analyses of human working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 95(20), 12061–12068.
- Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence, Vol. 2*. (pp. 47–103) Hillsdale, NJ: Lawrence Erlbaum Associates.
- te Nijenhuis, J., van Vianen, A. E. M., & van der Flier, H. (2007). Score gains on g-loaded tests: no g. *Intelligence*, 35, 283–300.
- Tranter, L. J., & Koutstaal, W. (2007). Age and flexible thinking: an experimental demonstration of the beneficial effects of increased cognitively stimulating activity on fluid intelligence in healthy older adults. *Neuropsychology, Development, and Cognition. Section B: Aging, Neuropsychology and Cognition*, 1–24.
- Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science*, 14(6), 623–628.
- Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: examining the correlation between Operation Span and Raven. *Intelligence*, 33, 67–81.
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behav Res Methods*, 37(3), 498–505.
- Van Casteren, M., & Davis, M. H. (2007). Match: a program to assist in matching the conditions of factorial experiments. *Behavior Research Methods*, 39(4), 973–978.
- Vanderplas, J. M., & Garvin, E. A. (1959). The association value of random shapes. *Journal of Experimental Psychology*, 3, 147–154.
- Verhaeghen, P., Marcoen, A., & Goossens, L. (1992). Improving memory performance in the aged through mnemonic training: a meta-analytic study. *Psychology and Aging*, 7(2), 242–251.
- Willis, S. L., & Schaie, K. W. (2009). Cognitive training and plasticity: theoretical perspective and methodological consequences. *Restorative Neurology and Neuroscience*, 27(5), 375–389.
- Zelinski, E. M. (2009). Far transfer in cognitive training of older adults. *Restorative Neurology and Neuroscience*, 27(5), 455–471.