Intelligence 50 (2015) 196–208

Contents lists available at ScienceDirect

Intelligence

Neural efficiency in working memory tasks: The impact of task demand

Daniela Nussbaumer [a,*], Roland H. Grabner [b], Elsbeth Stern [a]

[a] Institute for Behavioral Sciences, ETH Zurich, Clausiusstr. 59, 8092 Zurich, Switzerland
[b] Department of Psychology, University of Graz, Universitätsplatz 2, 8010 Graz, Austria

ARTICLE INFO

Article history:
Received 17 April 2014
Received in revised form 31 March 2015
Accepted 9 April 2015
Available online xxxx

Keywords:
Intelligence
Working memory task
Training
Task difficulty
Event-related desynchronization (ERD)

ABSTRACT

Studies of human intelligence provide strong evidence for the neural efficiency hypothesis, which suggests more efficient brain functioning (i.e., less or more focused activation) in more intelligent individuals. Recent studies have specified the scope of the neural efficiency hypothesis by suggesting that the relationship between brain activation and intelligence only holds true for problems of moderate difficulty and can be altered through training and is only found in frontal brain regions. We investigated the moderating roles of task difficulty and training on the neural efficiency phenomenon in the context of working memory (WM) training.

In two studies of 54 participants (study 1) and 29 participants (study 2), cortical activation was assessed by means of electroencephalography (EEG), or more precisely by means of event-related desynchronization (ERD) in the upper alpha band. ERD was assessed during the performance of WM tasks in a pre-test – training – post-test design, comparing groups of lower and higher intelligence.

We found supportive evidence for the neural efficiency hypothesis only in moderately difficult WM tasks in frontal brain regions, even in the absence of performance differences. There was no effect of intelligence on the simple or highly demanding, adaptive WM tasks. In the latter task, however, an intelligence-related difference emerged at the behavioral level, but training did not modulate the relationship between intelligence and brain activation.

These results corroborate the moderating role of task difficulty in the neural efficiency hypothesis in the context of WM demands and suggest that training does not impact the neural efficiency phenomenon in the context of WM demands.

1. Theoretical background

According to the neural efficiency hypothesis, differences in intelligence become apparent in the degree of brain activation that occurs during problem solving, i.e., for more intelligent individuals, the correct answer comes with less brain activation than for less intelligent individuals (Haier et al., 1988). This original hypothesis of neural efficiency was introduced in a positron emission tomography (PET) study, the results of which showed less brain glucose metabolism in more intelligent individuals while solving cognitive tasks. Haier and colleagues stated, "Intelligence is not a function of how hard the brain works but rather how efficiently it works … This efficiency may derive from the disuse of many brain areas irrelevant for good task performance as well as the more focused use of specific task-relevant areas" (Haier, Siegel, Tang, Abel, & Buchsbaum, 1992b, pp. 415–416). In addition, with electroencephalography (EEG), it was shown that event-related desynchronization (ERD) in the upper alpha band, considered an index of cortical activation (Klimesch, Doppelmayr, Pachinger, & Ripper, 1997; Pfurtscheller & Aranibar, 1977), is negatively related to intelligence (for a review, cf. Neubauer & Fink, 2009). However, although the

* Corresponding author.
E-mail addresses: nussbaumer@ifv.gess.ethz.ch (D. Nussbaumer), roland.grabner@uni-graz.at (R.H. Grabner), stern@ifv.gess.ethz.ch (E. Stern).

http://dx.doi.org/10.1016/j.intell.2015.04.004
0160-2896/

neural efficiency hypothesis has often been confirmed, moderating factors have been identified, in particular, task difficulty and practice or learning (Neubauer & Fink, 2009).

Various studies have demonstrated that the relationship between neural efficiency and intelligence may be altered by task difficulty (for an overview, see Neubauer & Fink, 2009). For instance, Neubauer, Sange, and Pfurtscheller (1999) did not find differences in brain activation between individuals with higher and lower IQ for simple (i.e., elementary cognitive) problems. The authors therefore concluded that a certain level of task difficulty is required for a corroboration of the neural efficiency effect. A different picture emerged in a study EEG measures while solving the Advanced Progressive Matrices Test (RAPM; Raven, 1990). Specifically, a negative relation between brain activation and intelligence was found for the easier items only, while for the more difficult ones, the opposite relationship was observed (Doppelmayr et al., 2005a). According to Neubauer and Fink (2009), these results do not necessarily contradict each other. The authors conclude, rather, that when more effort is required, more intelligent participants invest their available resources, resulting in both higher cortical activation and better achievement. Thus, it seems that in complex tasks, more intelligent individuals invest more cortical resources, resulting in a positive correlation between cortical activation and performance. In contrast, for moderate tasks, more intelligent individuals require less cortical resources to achieve the same performance as less intelligent individuals, resulting in a negative relation between cortical activation and performance.

Individual task difficulty, however, can be altered by practice, and based on the neural efficiency hypothesis, practice-related changes in brain activation may also be a function of intelligence. This has, in fact, been confirmed in two studies, which found a stronger decrease in activation after training for individuals with higher intelligence (Haier et al., 1992b; Neubauer, Grabner, Freudenthaler, Beckmann, & Guthke, 2004). The role of practice in the neural efficiency phenomenon has also become salient in investigations of experts in different domains who had achieved their expertise level through long-term training (Grabner, Neubauer, & Stern, 2006; Grabner, Stern, & Neubauer, 2003). These studies revealed that neural efficiency (in terms of more focused brain activation) is a function not only of intelligence but also of expertise. For instance, Grabner et al. (2006) compared the brain activation of individuals with lower and higher intelligence as well as with lower and higher expertise in tournament chess while solving chess-related tasks. They found independent impacts of intelligence and expertise level on brain activation. As expected, brighter individuals (independently of their expertise) displayed lower overall brain activity than their less intelligent peers. In addition, experts showed a lower frontal and more focused brain activation pattern compared to novices (i.e., individuals with lower degree of expertise).

Also with regard to brain areas only partial support for the neural efficiency hypothesis has been found. Neubauer and Fink (2009) summarize that effects of neural efficiency, i.e. the expected negative brain–intelligence relationship has been observed for frontal (but not for parietal) brain areas. For instance, Neubauer et al. (2004) found the strongest intelligence-related differences during reasoning tasks in frontal areas, more specifically in the prefrontal cortex, an area most strongly associated with reasoning processes. Similarly, Gray, Chabris, and Braver (2003) reported that for WM tasks prefrontal cortical activation discriminates between subjects with higher and lower intelligence, which is in accordance with findings of a high involvement of frontal areas while solving WM tasks (Smith & Jonides, 1997). A discrepancy between frontal and parietal brain areas in the context of neural efficiency has also been revealed in three studies which reported less frontal activation for more intelligent participants and a tendency for more parietal activation in the same participants (Gevins & Smith, 2000; Jaušovec & Jaušovec, 2004 and Rypma et al., 2006). Thus, even though an interplay of frontal and parietal brain areas is discussed to be important for intelligence (cf. the parieto-frontal integration theory by Jung & Haier, 2007); neural efficiency in terms of a negative brain–intelligence relationship has predominantly been found in frontal brain regions.

To summarize, several studies have provided support for the neural efficiency hypothesis mainly for frontal brain areas, but have also shown that task difficulty and training can moderate the relationship between intelligence and brain activation. There is, however, a paucity of studies in which task difficulty and training were combined in a comprehensive design. We conducted such a study involving WM training.

There is a wide agreement that WM is a core of human intelligence. Numerous studies have demonstrated substantial correlations between achievement on WM tasks and IQ (e.g., Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Conway, Kane, & Engle, 2003; Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004; Kyllonen & Christal, 1990). It can therefore be expected that the use of brain imaging while solving WM tasks will particularly highlight the impact of intelligence on neural activation. Moreover, the difficulty of WM tasks can be varied in a systematic and transparent way, for instance, by modulating demands for interference resolution or the amount of load. This allows the study of how the relationship between brain activation and intelligence may be moderated by task difficulty. Lastly, there is overwhelming evidence for the trainability of many types of WM tasks. As a result of repeated practice, the solution rate increases while the solution time goes down. Whether training effects transfer to other WM tasks, thereby demonstrating the malleability of WM functions, is hotly debated in psychology, and reviews and meta-analyses have provided reasons to doubt broader transfer effects (Chein & Morrison, 2010; Melby-Lervåg & Hulme, 2013; Shipstead, Redick, & Engle, 2012). Taken together, the advantages of WM tasks in investigations of neural efficiency are threefold: First, WM is seen as a basis of human intelligence. Second, the difficulty level can be manipulated gradually in that more or less WM load is incorporated into tasks. Third, WM activities are well represented in cortical activation, i.e., it is known that while solving WM tasks, there is a high involvement of frontal areas (Smith & Jonides, 1997). The present study consists of two training studies in which we assess brain activation (in terms of alpha ERD) in frontal areas before and after a three-week WM training in adult students differing in intelligence. The two studies differ in the level of task difficulty or WM load.

In study 1, we administered a WM-training with moderate complexity focusing on interference resolution. Interference resolution, which is the ability to select information among competing alternatives, is seen as a key function of WM (e.g., Nee, Wager, & Jonides, 2007). Specifically, the participants in the interference group (i.e., experimental group) practiced

WM tasks requiring the resolution of proactive interference. These participants were contrasted with a control group who worked on parallelized but very simple tasks that were not expected to challenge WM. Based on current evidence, we expected that the more intelligent participants would show higher neural efficiency (less brain activation) in frontal areas than those with lower intelligence while solving these WM tasks of moderate complexity. In addition, we hypothesized a decrease in brain activation in the course of the training that is expected to be larger for the more intelligent participants. In the control group, in contrast, we did not expect a relationship between intelligence and brain activation before or after training or that the changes in brain activation would depend on the individuals' intelligence level. In other words, we hypothesized that task difficulty influences the corroboration of the neural efficiency hypothesis in WM tasks in that the negative relationship between intelligence and frontal brain activation can be found for moderately difficult tasks only. In study 2, we employed a different WM training with high complexity to investigate the neural efficiency hypothesis in an adaptive task that maximized the individual task difficulty. Specifically, the training in study 2 had an identically high demand for resolution of proactive interference as in the interference group of study 1 but varied in two aspects. First, the training was adaptive and therefore ensured that WM load was constantly high and that the participants trained at their limits at all times. The task ensured the continuous engagement of executive processes and allowed only minimal development of task-specific strategies but encouraged solution strategies that could be applied in varying situations. Second, a dual task component was incorporated so that executive processes were required for managing the two tasks simultaneously as well as each task separately. By this task composition we attempted to ensure that automatic processes were minimized and that the dual task group trained with maximal WM load and with a maximal demand for interference resolution. The participants had to allocate relevant information to one of both tasks (dual task) and to the correct trial (resolution of proactive interference) while inhibiting irrelevant information. In light of the adaptive nature of the task, we did not expect differences in neural activation in the pre-test between participants of differing intelligence levels. Intelligence differences, however, should emerge at the behavioral level in that the more intelligent individuals should display better performance compared to their less intelligent peers. A similar result pattern can be expected to hold true for the post-test. Finding no training-related change (and no differences between the more and less intelligent individuals) in brain activation would be in line with our expectation, as the individual level of cognitive demand remains constantly high throughout training.

2. Study 1

2.1. Method of study 1

2.1.1. Participants

A total of 54 healthy students of science- and humanities-related fields from three Swiss universities completed the study, which included a pre-and a post-test session at the first author's institution and training sessions at home ($n = 54$, $M_{age} = 23.4$, $SD = 3.5$, 24 males, 30 females). From the initial 61 participants, five participants were excluded due to installation problems of the training software on their home computers. Two other participants were excluded due to non-adherence to the training paradigms, resulting in 54 participants for subsequent analysis. All of the participants were right-handed and without any medical or psychological diseases (both determined by self-report). The participants were paid for their participation in the study.

2.1.2. Procedure

Participants were randomly assigned to either the interference group or the control group. The interference group trained at three different non-adaptive tasks with a high number of interference trials and a moderate level of WM load, whereas the control group trained at three different tasks for which only 1 item had to be memorized at a time. The participants trained 5 days a week during a 3-week period for half an hour daily on their home computer. To check for the plausibility of training gains, the participants were required to send their training results immediately after completing the training session at home. The first and last training session took place at the first author's institution, and the tasks were performed while measuring EEG.

2.1.3. Material

2.1.3.1. Training paradigms. Both of the groups trained at three non-adaptive WM tasks, with each task taking approximately 10 min on each training day in a counterbalanced order. The solution time and the proportion of correct answers (i.e., the correct answers divided by the number of trials) were measured for all tasks.

The interference group practiced two recognition tasks (one with letters and one with faces) and one n-back task. All of the tasks of the interference group were characterized by a focus on resolution of proactive interference and by moderate WM load. In the two recognition tasks, a fixation point (1500 ms) was followed by a target set of four items that were arranged in a square configuration (faces were presented for 1500 ms and letters were presented for 500 ms). After a 3000 ms retention interval, during which only a fixation point was presented, a single probe appeared for a maximum of 1500 ms (see Fig. 1). On 50% of the trials, the probes were of the current target set; on the other 50% of the trials, the probes were new, i.e., non-matching. Of these nonmatching probes, two-thirds were interference trials in that they appeared in the target set of the previous trial or the previous two trials, and only one-third of the probes were non-recent. The participants had to press two different buttons to indicate a match vs. a mismatch. The solution times were measured from probe onset to the button press, and the proportion of correct answers was calculated. The stimuli in verbal tasks consisted of 19 consonants (without l and y). As face stimuli, 20 digitized gray-scale portraits (50% males, 50% females) of ordinary people were used. In addition, the interference group practiced one n-back task in a 3-back version in which a series of letters was shown at a rate of 2000 ms per letter, with a 2000 ms interstimulus interval (see Fig. 2). When the current letter matched the letter that appeared three positions earlier, the participants were required to press a button as quickly and accurately as possible. Half of the letters were target trials (matched the letter three positions earlier), and half were non-target trials. Of the non-target trials,

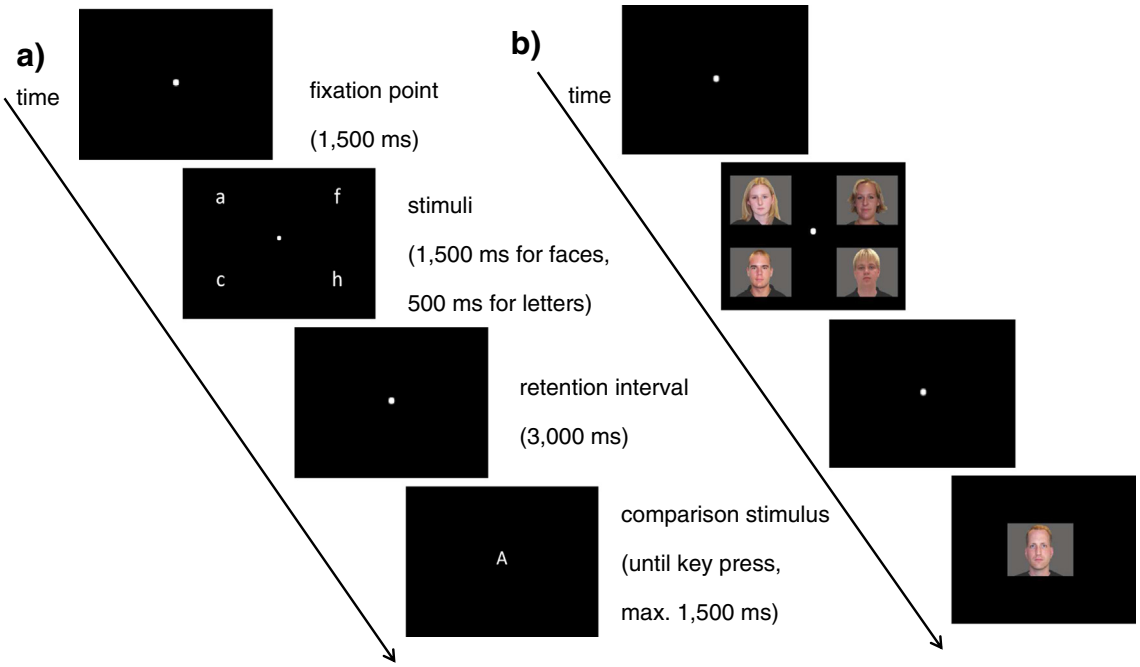
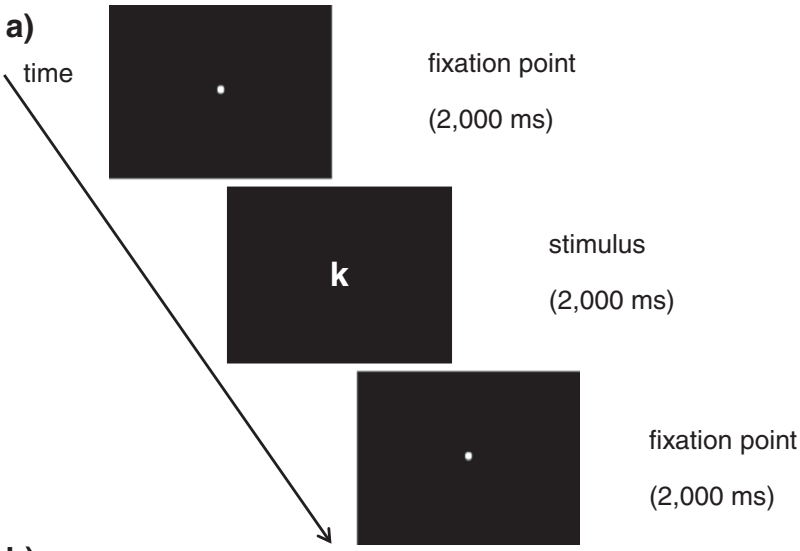


Fig. 1. a) Schematic view of the letter recognition task (4-letters task). Note that the comparison stimulus is written in capital letters to minimize purely perceptual processing. The letter-matching task (1-letter task) is analogous, except that the stimulus only consists of one letter. b) Schematic view of the face recognition task (4-face task). The face-matching task (1-face task) is analogous, except that the stimulus only consists of one face.



b)

Sequence	a	f	f	a	m	m	s	m
3-back	x	x	x	hit	x	x	x	hit
1-back	x	x	hit	x	x	hit	x	x

Fig. 2. Schematic view of the n-back training tasks. a) The same timing applied for the 3-back and the 1-back task b) possible sequence of the 3-back and the 1-back task (upper line), with hits in the 3-back version (middle line) and the 1-back version (lower line).

three quarters were interference trials, i.e., a letter presented matched the letter two, four, or five trials before. All of the tasks of the control group were tightly parallelized to the tasks of the interference group but induced only marginal WM load. In all 3 tasks, only 1 item had to be memorized at a time, and no interference trials were incorporated into the task. The face- and letter-matching tasks used the same stimuli, timing and response format as the face- vs. letter-recognition task in the interference group (see Fig. 1). However, the memory set consisted of one item only, so the participants simply indicated whether or not the probe matched an item. Analogous to the 3-back task, the participants solved a 1-back task with the same timing. In this task, the participants had to indicate whether a letter was repeated or not (see Fig. 2). In agreement with the WM literature (e.g., Shipstead, Redick, & Engle, 2010), a task with a WM load of three items or less is not an actual WM task; this group was therefore considered a control group. The tight parallelization of the interference and the control group allowed us to straightforwardly attribute potential differences developing over the course of the training to the training itself.

2.1.3.2. Paper–pencil measures. As a fluid intelligence test, the well-established Advanced Progressive Matrices Test (RAPM, Set II) by Raven (1990) was administered. Half of the items were presented in the first session, and the other half were presented in the session after the training. As all participants solved the RAPM twice, once at pre-testing before and once at post-testing after training, an even-odd split version was presented (participants were randomly assigned to the specific order). As only half the items were presented at each measurement point, no IQ-value could be calculated.

2.1.3.3. Mental effort rating. Participants had to judge the subjective mental effort of each training task on the mental effort rating scale (Paas, 1992; Paas, Tuovinen, Tabbers, & Van Gerven, 2003) in order to determine the subjectively perceived task difficulty. This measure allows to judge whether the level of task complexity in the two training groups differs as intended. Specifically, the differentiation between moderate and high WM load has been established via this rating scale (Paas, 1992, p. 431; minimum 1, maximum 9). Values between 4 and 6 are qualified as moderate, values below as low and values above as high.

2.1.3.4. EEG. The EEG assessment was conducted using an ActiveTwo-System (BioSemi, Amsterdam, The Netherlands). Event-related desynchronization/synchronization (ERD/ERS) was calculated for the upper alpha band (10–13 Hz) (Klimesch, 1999; Neubauer, Fink, & Grabner, 2006). Sixty-four scalp electrodes were placed according to the extended 10–20 system, and four additional electrodes were placed horizontally and vertically around the eyes to measure the electrooculogram (EOG). The EEG and EOG were sampled at 256 Hz. The EEG was recorded during resting state for 3 min and during deliberate, instructed eye movement to allow for the automatic correction of eye movement artifacts while performing tasks. To eliminate contamination artifacts, a band-pass filter was administered (between 0.5 and 45 Hz), EOG artifacts were reduced (automatic regression method; Schloegl et al., 2007), and all sequences were visually inspected for artifacts. For a detailed description of data

Table 1
Mean and standard deviation. a) For solution time per task for pre- and post-testing. b) For proportion of correct answers per task for pre- and post-testing.

	Control group						Interference group					
	1-back task		Item matching (1-face task)		Item matching (1-letter task)		3-back task		Item recognition (4-face task)		Item recognition (4-letter task)	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
a) Solution time (measured in seconds) per task for pre- and post-testing												
Low IQ-group	0.48 (0.05)	0.42 (0.06)	0.72 (0.1)	0.64 (0.13)	0.56 (0.08)	0.48 (0.07)	0.6 (0.12)	0.47 (0.13)	0.86 (0.11)	0.7 (0.1)	0.64 (0.07)	0.6 (0.09)
High IQ-group	0.46 (0.08)	0.4 (0.06)	0.66 (0.12)	0.58 (0.09)	0.54 (0.11)	0.45 (0.07)	0.6 (0.13)	0.46 (0.11)	0.96 (0.16)	0.75 (0.13)	0.7 (0.11)	0.59 (0.1)
Overall	0.47 (0.06)	0.41 (0.06)	0.70 (0.11)	0.61 (0.11)	0.55 (0.09)	0.47 (0.07)	0.60 (0.12)	0.47 (0.11)	0.92 (0.15)	0.73 (0.12)	0.67 (0.09)	0.60 (0.09)
b) Proportion of correct answers per task for pre- and post-testing												
Low IQ-group	0.86 (0.21)	0.97 (0.04)	0.97 (0.02)	0.95 (0.06)	0.96 (0.05)	0.96 (0.03)	0.71 (0.12)	0.93 (0.1)	0.65 (0.07)	0.70 (0.09)	0.84 (0.14)	0.91 (0.03)
High IQ-group	0.99 (0.01)	0.97 (0.03)	0.96 (0.02)	0.96 (0.05)	0.97 (0.03)	0.95 (0.06)	0.81 (0.15)	0.94 (0.08)	0.7 (0.11)	0.76 (0.14)	0.89 (0.05)	0.91 (0.07)
Overall	0.91 (0.17)	0.97 (0.04)	0.97 (0.02)	0.95 (0.05)	0.96 (0.04)	0.96 (0.05)	0.77 (0.15)	0.93 (0.09)	0.68 (0.10)	0.73 (0.12)	0.86 (0.10)	0.91 (0.06)

Table 2
MANOVA results for task performance in study 1 with the between-subject factor intelligence-group (IQ-group: lower vs. higher intelligence) and the within-subject factor time (pre- and post-testing) for all trained tasks.

	Main effect time	Main effect IQ-group	Interaction time * IQ-group
Solution time			
Control group	$F(2,20) = 10.52$	$F(2,20) = 0.45$	$F(2,20) = 0.18$
	$p < .001$	n.s. ($p = .72$)	n.s. ($p = .90$)
	$\eta^2_p = .62$	$\eta^2_p = .07$	$\eta^2_p = .03$
Interference group	$F(2,21) = 22.34$	$F(2,21) = 2.08$	$F(2,21) = 2.52$
	$p < .001$	n.s. ($p = .13$)	n.s. ($p = .09$)
	$\eta^2_p = .76$	$\eta^2_p = .23$	$\eta^2_p = .27$
Proportion of correct answers			
Control group	$F(2,20) = 0.96$	$F(2,20) = 0.91$	$F(2,20) = 2.07$
	n.s. ($p = .43$)	n.s. ($p = .45$)	n.s. ($p = .14$)
	$\eta^2_p = .13$	$\eta^2_p = .13$	$\eta^2_p = .25$
Interference group	$F(2,21) = 11.17$	$F(2,21) = 0.97$	$F(2,21) = 1.20$
	$p < .001$	n.s. ($p = .43$)	n.s. ($p = .34$)
	$\eta^2_p = .62$	$\eta^2_p = .12$	$\eta^2_p = .15$

analyses, see Grabner and De Smedt (2011) and De Smedt, Grabner, and Studer (2009). The ERS/ERD ratio was calculated for correctly solved trials for the upper alpha frequency band (10–13 Hz). A reference interval (R) comprised the time during the fixation interval (from 500 to 2500 ms after trial onset). The activation interval (A) is defined as the period from problem presentation the response. The 125 ms before response were excluded to eliminate motor artifacts. The band powers of A and R were obtained by squaring and averaging the artifact-free EEG signal separately for the sequences R and A. The ERS/ERD ratio was computed as follows: %ERS/ERD = [(A * R) / R]100. Negative values (ERD) indicate desynchronization and a decrease in power. Positive values (ERS) indicate synchronization and an increase in power.

For the statistical analyses, the %ERS/ERD was topographically aggregated by averaging electrodes. In light of the crucial role of frontal brain areas in the context of neural efficiency (Neubauer & Fink, 2009) our hypotheses focus on frontal areas, i.e. we expect differences to be limited to these parts of the brain. We therefore calculated a frontal %ERS/ERD-measure by averaging the following electrodes: Fp1, AF7, AF3, F7, F5, F3, F1, FC5, FC3, FC1 Fp2, AF8, AF4, F8, F6, F4, F2, FC6, FC4, and FC2 (for the distribution of the electrodes see, e.g., Grabner & De Smedt, 2011). Since parietal areas are also critical in WM tasks (Gevins & Smith, 2000; Jaušovec & Jaušovec, 2004 and Rypma et al., 2006) and have been considered to be important in the parieto-frontal integration theory of intelligence (Jung & Haier, 2007), we also calculate a parietal %ERS/ERD-measure by averaging the following electrodes: CP5, CP3, CP1, P7, P5, P3, P1, PO7, PO3, O1 CP6, CP4, CP2, P8, P6, P4, P2, PO8, PO4, and O2. Finding no differences between more and less intelligent participants in the parietal areas, however, is compatible with the neural efficiency hypothesis. Therefore, running separate analyses for both brain areas without correction for multiple comparisons is justified.

2.1.3.5. Statistical analysis. Solution time on the three training tasks was first analyzed in a multivariate analyses of variance for repeated measures (MANOVA, separately for the interference group and the control group, as tasks did not overlap), with

Table 3
Mean and standard deviation. a) For ERD frontal per task for pre- and post-testing. b) For ERD parietal per task for pre- and post-testing.

IQ-group	Control group						Interference group					
	1-back task		Item matching (1-face task)		Item matching (1-letter task)		3-back task		Item recognition (4-face task)		Item recognition (4-letter task)	
	T1	T13	T1	T13	T1	T13	T1	T13	T1	T13	T1	T13
a) ERD frontal per task for pre- and post-testing												
Low	−2.06 (8.9)	0.46 (8.83)	−0.02 (8.81)	1.21 (13.68)	−11.82 (11.48)	−2.43 (16.66)	−2.85 (8.23)	−1.21 (7.94)	−9.33 (9.33)	−7.41 (9.54)	−6.73 (17.5)	1.81 (15.09)
High	−7.07 (6.99)	−3.47 (8.48)	−9.64 (14.43)	−1.57 (10.15)	0.07 (23.13)	−15.92 (14.37)	2.81 (7.42)	−0.85 (6.31)	−1.12 (5.78)	−0.67 (5.96)	−5.48 (13.99)	−4.51 (16.48)
b) ERD parietal per task for pre- and post-testing												
Low	−3.07 (9.94)	2.96 (14.84)	−0.04 (12.06)	1.09 (15.09)	−5.75 (15.68)	14.46 (26.72)	−2.02 (10.82)	5.92 (10.67)	−5.88 (11.5)	−7.76 (12.05)	−3.72 (20.93)	1.71 (10.85)
High	−4.01 (10.05)	−0.41 (7.2)	−8.23 (15.25)	−1.29 (11.06)	−2.45 (20.3)	−7.79 (15.79)	6.34 (11.51)	4.11 (10.18)	0.27 (7.88)	−0.24 (11.63)	−3.41 (17.21)	−0.18 (18.94)

intelligence group (lower intelligence, higher intelligence) as the between-subjects factor and solution time as the dependent variable. In the case of a significant MANOVA, we further analyzed each training task separately with univariate analyses (ANOVAs for repeated measures, with intelligence group as the between-subjects factor and solution time as the dependent variable). The same procedure was chosen to analyze the proportion of correct answers, frontal cortical activation (ERD-frontal) and parietal cortical activation (ERD-parietal).

2.2. Results of study 1

2.2.1. Behavioral data

The interference group and the control group did not differ in their initial intelligence level ($t(52) = -.82$, $p = .42$), and their intelligence level after completion of the training also did not differ ($t(52) = -.46$, $p = .65$). Taking into account all participants of the two groups, the median of the RAPM raw scores at pre-testing was ($n = 54$, $M = 12.26$, $SD = 2.32$) and at post-testing was ($n = 54$, $M = 12.65$, $SD = 2.38$). Therefore, two intelligence groups were formed by a median split of the RAPM raw scores[1] at pre-testing (splitting at 12.26 points; lower intelligence-group: $n = 28$, $M = 10.46$, $SD = 1.45$; higher intelligence-group: $n = 26$, $M = 14.19$ $SD = 1.27$; resulting in an effect size of $d = 2.74$). It should be noted that our sample consisted of university students and that, with a mean of 14 out of 18 maximal points for the higher intelligence group, they performed at a high level. In addition, the lower intelligence group, with a mean performance of 10.5 out of 18, performed at a respectable level, which was lower than the other group but still at an average level.

Table 1 depicts descriptive statistics for solution times and proportion of correct answers, and Table 2 presents the results of MANOVAs. All of the simple tasks of the control group (1-letter, 1-face, 1-back) show ceiling effects for the proportion of correct answers already in the pre-test. This result indicates that these tasks were – as intended – not challenging for the entire sample. For the proportion of correct answers in the interference group, the MANOVA revealed a main effect of time, and follow-up analyses showed that this was due to gains in the 3-back and the 4-face tasks. Contrary to our intention, ceiling effects also partially appeared in the interference group: For the 4-letter task, the proportion of correct answers exceeded .80 in the pre-test and .90 in the post-test. A similar ceiling effect was observed for the 3-back task in the post-test.

For the solution times, a MANOVA revealed a main effect of time for the control group as well as for the interference group (see Table 2; for descriptive statistics, see Table 1). Follow-up ANOVAs for each task revealed that the mean solution time decreased significantly for all tasks of the interference and the control group as a result of practice. No significant effects of "intelligence" and no interactions between time and intelligence appeared. Thus, as intended, in neither group were intelligence differences related to any behavioral measure. Therefore, no trade-off between performance and neural efficiency can be said to be responsible for significant differences in ERD between intelligence groups.

However, the goal of finding tasks of medium difficulty for the interference group was only partly approached. For the 4-letter task, ceiling effects could be seen to emerge already in the pre-test for solution rates (>.80) as well as for solution times, which only decreased from 0.67 s in the pre-test to 0.60 in the post-test. While the 4-letter task turned out to be too easy already in the pre-test, the 3-back task appeared to be particularly amendable by practice and therefore less challenging in the post-test than assumed. The mean solution rate exceeded .90 in the post-test. A comparison of the solution times for the 1-back task (control group) and the 3-back task (intervention group) underlined the effects of practice: In the pre-test, it took clearly longer to solve the 3-back task (0.60 s) than the 1-back task (0.47 s), $d = 1.4$. In the post-test, however solution times grew closer together (1-back task: 0.41 s, 3-back task 0.46 s, $d = .7$). Comparing the solution times between the 1-face task and the 4-face task revealed effect sizes of $d > 1$ for the pre- as well as the post-test. The same was true for a comparison of the 1-letter task and the 4-letter task.

The proportion of correct answers and the solution times for the 4-face task, however, showed that we were able to create problems of medium difficulty for the entire intelligence range that can be expected to highlight neural efficiency effects. It is therefore justified to investigate whether differences in brain activation can be traced back to intelligence and training, which will be done in the next section.

2.2.2. Neural efficiency

The MANOVA on frontal activation of the interference group revealed differences in cortical activation between the intelligence groups (main effect of intelligence-group ($F(3,20) = 3.37$, $p = .039$, $\eta^2_p = .336$), see Table 4; for descriptive statistics, see Table 3). Subsequent ANOVAs showed that these differences occurred due to a significant main effect of intelligence group in the recognition task 4-faces ($F(1, 24) = 7.38$, $p = .012$, $\eta^2_p = .24$); for the other tasks (4-letters and 3-back we found $p > .10$.). As we are doing multiple comparisons at the level of the follow-up ANOVAs we are taking this into consideration by a Bonferroni adjustment. This lowers the accepted p-value from $p < 0.05$ to $p < 0.016$, which is higher than the obtained value of $p = .012$ in the ANOVA of the 4-face task.

Finding neural efficiency effects for both the 4-letter and the 3-back tasks was not very likely, given the ceiling effects discussed in the previous section. It was therefore thoroughly plausible that the effect solely occurred in the 4-face task. In order to ensure that we do not rely on an artifact we decided to corroborate this finding by two additional analyses.

First, an extreme group approach was chosen for running an analysis of variance. The groups were formed by including the third (33%) with highest and with lowest intelligence instead of forming the groups by a median split. The findings we got from median split were confirmed in a MANOVA with a considerably higher effect size ($F(3, 12) = 6.33$, $p = .008$, $\eta^2_p = .613$). Subsequent ANOVAs showed again differences due to a significant main effect of intelligence group in the recognition task 4-faces ($F(1, 15) = 12.63$, $p = .003$, $\eta^2_p = .46$); for the other tasks (4-letters and 3-back) we found $p > .10$ Again, the result holds true for Bonferroni corrections which require a $p < 0.016$.

Second, due to the high standard deviations of the EEG-data (which are comparable to other studies; e.g. Neubauer et al.,

[1] For the method of median split to form intelligence groups see e.g. Neubauer, Freudenthaler, and Pfurtscheller (1995) for the RAPM and Neubauer and Fink (2003) or Doppelmayr et al. (2005b) for other intelligence tests.

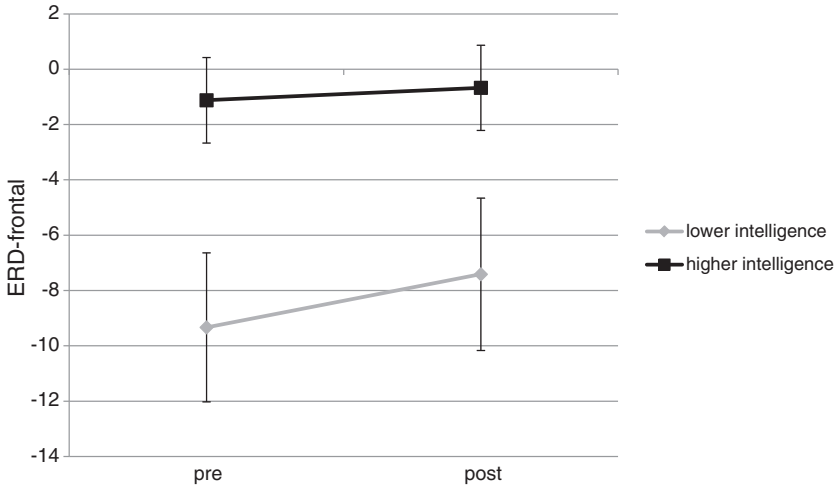


Fig. 3. Graph of the face recognition task (4-faces). Values for ERD-frontal on the y-axis, time on the x-axis. An ANOVA for the face recognition task revealed a main effect of group ($F(1, 24) = 7.38$; $p = .012$; $\eta^2_p = .24$) with no main effect of time and no interaction. The error bars represent the standard error of the mean.

2004) we decided to double-check the findings by running a nonparametric test. The Kruskal–Wallis test revealed a significant result for the face recognition task ($H(1) = 5.85$, $p = .016$ with a rank of 9.6 for the lower intelligence group and with 16.9 for the higher intelligence group). Thus we confirmed a robust neural efficiency effect for the 4-face task, which is highlighted in Fig. 3.

However, while more intelligent participants showed less cortical activation while solving the 4-face task than the less intelligent ones, these activation differences between the intelligence groups did not change after training. They rather remained stable from pre- to post-testing, as there was no interaction between intelligence group and time. There also was no main effect of time. For the control group, none of the effects in the multivariate analyses reached significance (see Table 3 and for descriptive statistics see Table 4), which means there was no difference between the intelligence groups and no training-related change.

Table 4
MANOVA results for alpha ERD in study 1 with the between-subject factor intelligence-group (lower vs. higher intelligence) and the within-subject factor time (pre- and post-testing), including all trained tasks.

	Main effect time	Main effect IQ-group	Interaction time * IQ-group
ERD frontal			
Control group	$F(2,19) = 1.36$ n.s. ($p = .28$) $\eta^2_p = .18$	$F(2,19) = 0.99$ n.s. ($p = .41$) $\eta^2_p = .14$	$F(2,19) = 2.23$ n.s. ($p = .19$) $\eta^2_p = .26$
Interference group	$F(2,20) = 0.70$ n.s. ($p = .57$) $\eta^2_p = .09$	$F(2,20) = 3.37$ $p < .05$ $\eta^2_p = .37$	$F(2,20) = 1.01$ n.s. ($p = .41$) $\eta^2_p = .13$
ERD parietal			
Control group	$F(2,17) = 2.27$ n.s. ($p = .08$) $\eta^2_p = .33$	$F(2,17) = 2.22$ n.s. ($p = .12$) $\eta^2_p = .28$	$F(2,17) = 1.70$ n.s. ($p = .20$) $\eta^2_p = .23$
Interference group	$F(2,20) = 2.20$ *n.s.*($p = .12$) $\eta^2_p = .25$	$F(2,20) = 1.00$ n.s. ($p = .41$) $\eta^2_p = .13$	$F(2,20) = 2.38$ n.s. ($p = .10$) $\eta^2_p = .26$

With respect to the parietal cortical activation of the interference group, multivariate analyses revealed no main effect of intelligence, no main effect of time and no interaction. Therefore, for parietal activation, no differences between the participants of different intelligence levels were found, and there was no progress over time and no difference in the progress between the groups. Thus, while the intelligence groups showed different levels of cortical activation in the frontal regions, we did not observe any differences in parietal areas. In the control group, again, no differences between the intelligence groups or any progress over time occurred for parietal activation.

2.3. Discussion of study 1

In study 1, we sought to confirm findings that the neural efficiency hypothesis holds true for moderately difficult tasks but not for simple ones. We chose WM tasks because they can be expected to be particularly strongly related to intelligence. In fact, we show that more and less intelligent participants did not differ with respect to neural activation when they were processing tasks that only slightly stressed WM. For tasks that moderately strained WM, however, we confirmed the neural efficiency hypothesis. Specifically, the more intelligent participants showed less frontal brain activation than the less intelligent participants, even though there were no performance differences between groups. It is worth noting that our entire sample was above average on the IQ scale. Therefore, we compared highly intelligent individuals against moderately high intelligent individuals. Support for the neural efficiency hypothesis in this sample could be considered as strong evidence of its validity.

In line with the review by Neubauer and Fink (2009) we found frontal but not parietal differences between the higher and lower intelligent participants.

Contrary to our expectation, we found neither training effects nor interactions with intelligence for any of the tasks on the neural level. This stands in contrast to the results of Neubauer et al. (2004), who found a reduction in cortical activation even after a short training, and to the results of Haier

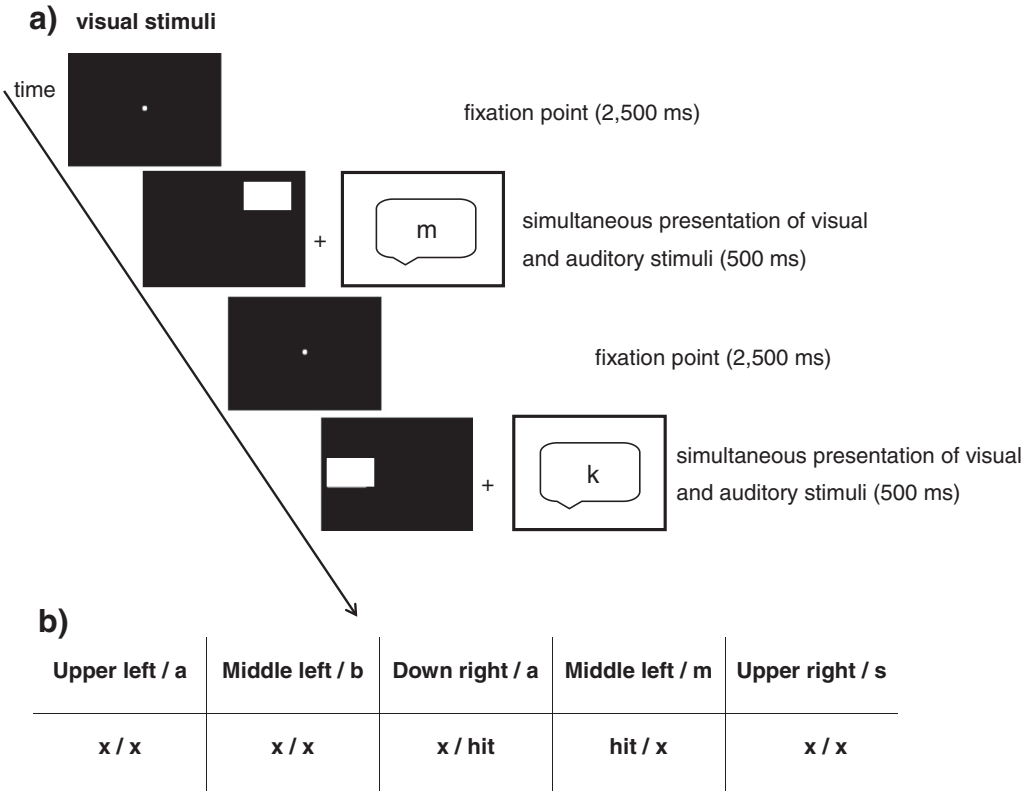


Fig. 4. Schematic view of the n-back training tasks a) dual n-back task b) possible sequence of the dual n-back task (upper line) with hits (lower line) in the version of a 2-back task.

et al. (1992a) and Haier et al. (1992b), who found very strong behavioral training effects, as well as a reduction in cortical activation, after a four to eight weeks training with Tetris.

However, we found clear learning effects on the behavioral level in terms of increased solution rates and reduced solution times. It seems that after three weeks of daily training, the burden on WM had not been changed sufficiently to affect our EEG measures. Thus, there may be a difference between our WM tasks and tasks with a clear demand on pattern recognition and chunking as applied by Neubauer et al. (2004) and Haier et al. (1992b).

3. Study 2

3.1. Method of study 2

3.1.1. Participants

A total of 29 healthy students of science- and humanities-related fields from three Swiss universities completed the study, including all training sessions as well as pre-and post-testing (for study 2: $n = 29$, $M_{age} = 23.7$, $SD = 2.7$, 17 males, 12 females). From the initial 30 participants, one participant had to be excluded due to non-adherence to the training paradigms and transfer sessions at the institute, resulting in the 29 participants for subsequent analysis. Again, all participants were right-handed and without any medical or psychological diseases. These individuals were also paid for their participation in the study.

3.1.2. Procedure and EEG

All of the participants of study 2 were assigned to a dual task training in which they trained in a dual n-back task 5 days a week during a 3-week period for half an hour daily on their home computer. The procedure, as well as the EEG assessment and analyses, were parallel to study 1.

3.1.3. Material

3.1.3.1. Training paradigms. The participants trained with an adaptive dual n-back task with high WM load and a large number of interference trials (similar to Jaeggi, Buschkuehl, Jonides, & Perrig, 2008). The participants were presented with a sequence of simultaneous visual and auditory stimuli (see Fig. 4). The visual stimuli consisted of squares arranged in a square configuration, resulting in eight different locations. The auditory stimuli were presented through headphones and consisted of one of eight consonants. The stimuli were presented at a rate of 3000 ms, with a stimulus length of 500 ms and an interstimulus interval of 2500 ms. Every block began with a central fixation cross. The participants had to indicate whenever one or both of the presented stimuli matched the one that was presented n positions before. Each block consisted of 20 plus n trials. Of these, five trials were no hits, five trials were targets in the auditory modality, five trials were targets in the visual modality and five trials were targets for both modalities simultaneously. This task is characterized by a high demand for interference resolution in that lures were

Table 5
Means and standard deviations for the n-back level, ERD-frontal, and ERD-parietal for pre- and post-testing.

	Behavioral measure: n-back level		ERD frontal		ERD parietal	
	Pre	Post	Pre	Post	Pre	Post
Low	1.61 (0.30)	3.02 (0.64)	−12.65 (12.88)	−7.31 (18.94)	−11.95 (14.91)	−6.22 (23.74)
High	1.87 (0.31)	3.34 (1.36)	−16.71 (7.54)	−12.26 (11.24)	−20.05 (12.67)	−13.39 (16.32)

presented, i.e., items with a high familiarity due to recent presentation but not actually hits. Of the non-target trials, three quarters were interference trials, i.e., a letter was presented matched the letter shown two, four, or five trials before. The participants had to indicate when the current stimulus matched the stimulus from n steps earlier in the sequence. The responses were required for targets only. The visual targets had to be indicated by a button press with their left index finger, auditory targets by a button press of their right index finger and double targets by two button presses with each hand. According to the participants' performance, the value of n could change after each block but always remained the same for visual and auditory stimuli. The level of n increased by 1 if the participant made fewer than three mistakes per modality and decreased by 1 if more than five mistakes were made; for 3, 4 or 5 mistakes per modality, the level of n stayed the same. Each training session consisted of 24 blocks, resulting in a total daily training time of ca. 30 min.

3.2. Results and discussion of study 2

The two intelligence groups were again formed by a median split of the RAPM raw scores at pre-testing (lower intelligence-group: $n = 12$, $M = 10.00$, $SD = 1.41$; higher intelligence-group: $n = 17$, $M = 14.88$ $SD = 1.54$; resulting in an effect size of $d = 3.31$). This sample consisted of university students who were performing at a comparably high level as the participants of study 1.

In a first step, we tested whether more or less intelligent students differed in the initial achievement on the n-back task. This was indeed found to be the case. A *t*-test revealed that the n-back level for the lower intelligence group (n-back level: $M = 1.61$, $SD = 0.33$) was lower than for the higher intelligent group (n-back level: $M = 1.87$, $SD = 0.31$), ($t(27) = -2.19$, $p = .037$, Cohen's $d = 0.88$). The impact of intelligence on the initial n-back level was also reflected in a significant correlation of $r = .39$, $p = .037$.

Next, we performed t-tests to determine if there were any initial differences in frontal and parietal cortical activation between the intelligence groups. At pre-test for both areas, the results were non-significant (ERD-frontal $t(26) = 1.06$, n.s. ($p = .30$) and ERD-parietal $t(26) = 1.54$, n.s. ($p = .14$)). In addition, correlations between intelligence level and cortical activation were non-significant and low (RAPM raw score and ERD frontal at pre-testing: $r = -.20$, n.s. ($p = .30$); RAPM raw score and ERD parietal at pre-testing: $r = -.29$, n.s. ($p = .14$)). This result is consistent with those of Neubauer and Fink (2009) and Dunst, Benedek, Jauk, and Neubauer (2014), who showed that more intelligent students invested an equal amount of cortical resources but attained a higher achievement level. Consequently, our data support the neural efficiency hypothesis, such that higher intelligence should manifest either as a lower level of cortical activation or as better performance while displaying an equal amount of cortical activation as less intelligent individuals.

Analyses of the post-test time point revealed that the intelligence groups did not differ regarding their behavioral measure or concerning cortical activation (n-back level $t(27) = 0.11$, n.s. ($p = .44$); ERD-frontal $t(26) = 0.87$, n.s. ($p = .39$); *t*-test ERD-parietal $t(26) = 0.95$, n.s. ($p = .35$)). The correlations of intelligence with n-back level ($r = .32$, n.s., $p = .10$) and intelligence with cortical activation (RAPM raw score with ERD frontal at post-testing: $r = -.20$, n.s. ($p = .30$) and RAPM raw score with ERD parietal at post-testing: $r = -.29$, n.s. ($p = .14$)) were not significant.

We lastly analyzed the impact of the training using ANOVAs for the intelligence group (i.e., lower vs. higher intelligence) as a between-subject variable and time (pre- vs. post-testing) as a within-subject variable. Concerning the n-back level, the ANOVA showed a significant main effect of time but no group differences or interactions. Thus, all of the participants were able to improve their n-back level from pre- to post-testing irrespective of their intelligence level (see Table 6 and for descriptive values Table 5). With respect to frontal cortical activation, none of the values derived from the ANOVA reached significance. Therefore, that was no progression over time, no differences between the intelligence groups, and no interactions (see Table 6). Analyses of parietal cortical activations again did not show any significant values. Thus, for parietal activation as

Table 6
Training results for study 2: Reporting main effects and interactions for an ANOVA with the between-subject factor of intelligence group (lower vs. higher intelligence) and the within–subject factor of time (pre– and post-testing) for the n-back level, ERD-frontal, and ERD-parietal.

Measure	Main effect time	Main effect IQ-group	Interaction time * IQ-group
Behavioral measure: n-back Level	$F(1, 27) = 48.61$ $p < .001$ $\eta^2_p = .64$	$F(1,27) = 0.86$ n.s.($p = .22$) $\eta^2_p = .06$	$F(1,27) = 0.04$ n.s.($p = .85$) $\eta^2_p = .001$
ERD-frontal	$F(1, 26) = 3.14$ n.s. ($p = .09$) $\eta^2_p = .11$	$F(1, 26) = 1.28$ n.s. ($p = .27$) $\eta^2_p = .05$	$F(1, 26) = 0.03$ n.s. ($p = .87$) $\eta^2_p = .001$
ERD-parietal	$F(1, 26) = 2.81$ n.s. ($p = .11$) $\eta^2_p = .10$	$F(1, 26) = 2.03$ n.s. ($p = .17$) $\eta^2_p < .07$	$F(1, 26) = 0.02$ n.s. ($p = .90$) $\eta^2_p = .001$

well, there was no progression over time, no differences between the intelligence groups, and no interactions (see Table 6).

Before the n-back task training began, there was a substantial correlation between performance in this task and intelligence, indicating that individuals with higher intelligence started on a higher n-back level. Over the course of the training, all of the participants improved their n-back level by advancing to a more difficult task version, with no evident impact of intelligence on learning gains. Moreover, no differences between the intelligence groups were apparent at post-testing.

Concerning cortical activation no differences were observed between the intelligence groups neither before nor after the training. This result is in line with the reformulation of the neural efficiency hypothesis by Neubauer and Fink (2009) and the specifications by Dunst et al. (2014), who found that with the same (high) subjective task difficulty – in the case of the n-back task at maximum – no intelligence difference was found in cortical activation. While we observed achievement gains in the n-back level between pre- and post-testing, there was no change in cortical activation and no differences between the intelligence groups. Better performance without a change in cortical activation could also be interpreted as a relative gain of efficiency.

4. General discussion

We measured cortical activation while solving WM tasks to test the scope of the neural efficiency hypothesis. More specifically, we aimed at replicating the specification of the neural efficiency hypothesis, i.e., that whether frontal cortical activation is moderated by intelligence depends on the difficulty level of the tasks. In addition, we sought to determine how cortical activation changes as a result of training. For this purpose, we conducted two studies. In study 1, we compared participants who underwent a three-week moderate WM training period (interference group) with an active control group. In study 2, we investigated cortical activation in a group that trained with a high WM load.

As the two training groups in study 1 were fully parallelized regarding stimulus presentation, timing, and task demands, and differed solely in the amount of WM load, differences between these groups can readily be attributed to the amount of WM-load during training. In line with our expectations, our results revealed that the neural efficiency effect appeared only in the interference group but not in the control group. It has to be emphasized that the frontal activation difference between the intelligence groups in the face-recognition task was of considerable effect size of $\eta^2_p = .24$, which corresponds to $d = 1.12$. To compare this effect with those of similar studies, we calculated Cohen's d based on the reported statistical information. The Tetris-training study of Haier et al. (1992a) described correlations between RAPM and frontal activation around .6, which corresponds to $d = 1.25$. In the training study of Neubauer et al. (2004) non-significant correlations between ERD-frontal and results in a reasoning test were found at pre-testing whereas significant correlations of .41 emerged in the post-test ($d = .89$).

In addition to these training studies, there is a variety of studies calculating ANOVAs with the between subject factor IQ-groups formed through applying a median split (as was done in the current study) showing no main effects of IQ-group but interactions in which the IQ-group is involved. For example, Grabner et al. (2003) found no main effect of IQ-group but a significant interaction between IQ-group and task ($d = .96$). Also, Doppelmayr, Klimesch, Hödlmoser, Sauseng, and Gruber (2005b) found no significant main effects of IQ-group but only interactions between intelligence and task (d around 1). Grabner et al. (2006) reported a three-way interaction between hemisphere, area and IQ-group ($F(3.12, 130.91) = 3.02, p < .05$, $\eta^2 = .07$), which equates $d = .55$. Taken together, the size of the neural efficiency effect we found for the 4-face task is relatively large and comparable to other studies. It is worth mentioning that we found direct evidence for intelligence differences in terms of a main effect of intelligence on cortical activation rather than interactions.

While we found strong neural-efficiency effects for the 4-face task, neither the 3-back task nor the 4-letter task reached significance, indicating that more and less intelligent participants did not differ in their cortical activation. Given the high solution rates and the low solution times for these two tasks depicted in Table 1, it seems that these problems were not challenging enough for our sample of university students to produce differences in cortical activation. Remembering letters and distinguishing them from each other, as it was required in the 3-back task and the 4-letter task is a highly familiar activity for university students, and they reach high performance levels without great effort. Therefore, intelligence differences within our already highly selected sample do not significantly moderate neural processing. In contrast, remembering faces presented in a context free situation and distinguishing between them is a relatively unfamiliar activity which requires the development of new strategies. More intelligent individuals apparently need less neural resources for mastering this novel task.

Indirect support for the neural efficiency hypothesis was found in study 2. Specifically, intelligence affected the performance level of the n-back task but not neural activation, presumably because participants were expected to invest their entire cognitive resources in improving performance. The dual task group trained with an adaptive WM task that placed high WM load during training by including a high number of interference trials and due to the dual task characteristics. The adaptivity ensured that the task had a high WM load and that each participant trained at its limit at all times. Therefore, we compared tasks of different difficulty. Our result is in line with Dunst et al. (2014), who found that the neural efficiency hypothesis is only true when all of the participants work on tasks with the same difficulty irrespective of their intelligence level. It can be supposed that neural efficiency reflects that more intelligent participants face less of a challenge to solve the tasks and that less intelligent participants face a greater challenge. When controlling for subjective task difficulty (i.e., when applying tasks with a difficulty level according to one's intelligence level), the participants with lower versus higher intelligence no longer show differences in brain activation (see also Larson, Haier, LaCasse, & Hazen, 1995). The adaptivity of the training in the high WM load group in the current study is comparable to the situation in the study of Dunst et al. (2014). Therefore, our finding that there were no intelligence differences in cortical activation while solving this adaptive task is in agreement with the finding of Dunst et al. (2014).

The brain activation patterns observed in the interference group of Study 1 supported the neural efficiency hypothesis for frontal but not parietal areas. This result is in concordance with various studies reporting less frontal activation for more intelligent participants (Gevins & Smith, 2000; Jaušovec & Jaušovec, 2004 and Rypma et al., 2006) and with the review by Neubauer and Fink (2009). In the present study we also did not find a tendency for a simultaneous higher parietal activation, as did e.g. Rypma et al. (2006).

While most previous studies on the neural efficiency hypothesis used either intelligence test items or elementary cognitive tasks, few have shown the effect for WM tasks (e.g., Grabner, Fink, Stipacek, Neuper, & Neubauer, 2004). Our findings also should be viewed in light of our sample of university students with above-average intelligence scores. That we still found differences in cognitive activation when WM tasks of medium difficulty were processed confirms the strength of the neural efficiency hypothesis.

Contrary to our expectation, we did not observe a decrease in cortical activation after WM task training in any group, irrespective of participants' intelligence level. The amount of cortical activation after the three-week-training remained at the same level, and no training-induced changes in neural efficiency depending on participants' intelligence level were found. Learning effects, however, were obvious on the behavioral level, as evidenced by increasing solution rates and decreasing solution times. The observation of no reduction in cortical activation after the training is in contrast to two other training studies. Haier et al. (1992a), Haier et al. (1992b) found a reduction in cortical activation after intensive practice of the computer game Tetris, which fosters spatial pattern matching and fine motor skills. Neubauer et al. (2004) even found training cortical effects after a two-hour training session for elementary cognitive tasks. Given that participants in study 2 only worked on a single task during the entire training, a lack of a decrease in cortical activation can hardly be traced to insufficient practice. The tasks used in our studies still draw on WM capacity even after intensive training. Future research should develop trainable tasks that allow the detection of changes in the brain that reflect learning trajectories modulated by the intelligence of the individuals.

In conclusion, the present studies corroborate the moderating role of task difficulty for the neural efficiency hypothesis. Our results extend previous findings by demonstrating this moderating effect in the context of WM demands and by revealing that in this task domain, training did not impact on the relationship between intelligence and brain activation. These insights contribute to a better understanding of the neurophysiological correlates of individual differences in intelligence.

References

Chein, J.M., & Morrison, A.B. (2010). Expanding the mind's workspace: Training and transfer effects with a complex memory span task. *Psychonomic Bulletin & Review*, *17*(2), 193–199.

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

Conway, A.R., Kane, M.J., & Engle, R.W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552.

De Smedt, B., Grabner, R.H., & Studer, B. (2009). Oscillatory EEG correlates of arithmetic strategy use in addition and subtraction. *Experimental Brain Research*, *195*(4), 635–642.

Doppelmayr, M., Klimesch, W., Hödlmoser, K., Sauseng, P., & Gruber, W. (2005a). Intelligence related upper alpha desynchronization in a semantic memory task. *Brain Research Bulletin*, *66*(2), 171–177.

Doppelmayr, M., Klimesch, W., Sauseng, P., Hödlmoser, K., Stadler, W., & Hanslmayr, S. (2005b). Intelligence related differences in EEG-bandpower. *Neuroscience Letters*, *381*(3), 309–313.

Dunst, B., Benedek, M., Jauk, E., & Neubauer, A.C. (2014). Neural efficiency depending on subjective and objective level of task difficulty — An fMRI study. *Personality and Individual Differences*, *60*, 36–37.

Engle, R.W., Tuholski, S.W., Laughlin, J.E., & Conway, A.R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology. General*, *128*(3), 309.

Gevins, A., & Smith, M.E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex*, *10*(9), 829–839.

Grabner, R.H., & De Smedt, B. (2011). Neurophysiological evidence for the validity of verbal strategy reports in mental arithmetic. *Biological Psychology*, *87*(1), 128–136.

Grabner, R.H., Fink, A., Stipacek, A., Neuper, C., & Neubauer, A.C. (2004). Intelligence and working memory systems: Evidence of neural efficiency in alpha band ERD. *Cognitive Brain Research*, *20*(2), 212–225.

Grabner, R.H., Neubauer, A.C., & Stern, E. (2006). Superior performance and neural efficiency: The impact of intelligence and expertise. *Brain Research Bulletin*, *69*(4), 422–439.

Grabner, R.H., Stern, E., & Neubauer, A.C. (2003). When intelligence loses its impact: Neural efficiency during reasoning in a familiar area. *International Journal of Psychophysiology*, *49*(2), 89–98.

Gray, J.R., Chabris, C.F., & Braver, T.S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*(3), 316–322.

Haier, R.J., Siegel, B.V., MacLachlan, A., Soderling, E., Lottenberg, S., & Buchsbaum, M.S. (1992a). Regional glucose metabolic changes after learning a complex visuospatial/motor task: A positron emission tomographic study. *Brain Research*, *570*, 134–143.

Haier, R.J., Siegel, B.V., Nuechterlein, K.H., Hazlett, E., Wu, J.C., Paek, J., et al. (1988). Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography. *Intelligence*, *12*, 199–217.

Haier, R.J., Siegel, B.V., Tang, C., Abel, L., & Buchsbaum, M.S. (1992b). Intelligence and changes in regional cerebral glucose metabolic rate following learning. *Intelligence*, *16*, 415–426.

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.

Jaušovec, N., & Jaušovec, K. (2004). Differences in induced brain activity during the performance of learning and working-memory tasks related to intelligence. *Brain and Cognition*, *54*(1), 65–74.

Jung, R.E., & Haier, R.J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behavioral and Brain Sciences*, *30*(02), 135–154.

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 Educational Psychology: General*, *133*(2), 189–217.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, *29*(2-3), 169–195.

Klimesch, W., Doppelmayr, M., Pachinger, T., & Ripper, B. (1997). Brain oscillations and human memory: EEG correlated in the upper alpha and theta band. *Neuroscience Letters*, *28*(238), 8–12.

Kyllonen, P.C., & Christal, R.E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*(4), 389–433.

Larson, G.E., Haier, R.J., LaCasse, L., & Hazen, K. (1995). Evaluation of a "mental effort" hypothesis for correlations between cortical metabolism and intelligence. *Intelligence*, *21*(3), 267–278.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A metaanalytic review. *Developmental Psychology*, *49*(2), 270–291.

Nee, D.E., Wager, T.D., & Jonides, J. (2007). Interference resolution: insights from a meta-analysis of neuroimaging tasks. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(1), 1–17.

Neubauer, A.C., & Fink, A. (2003). Fluid intelligence and neural efficiency: Effects of task complexity and sex. *Personality and Individual Differences*, *35*(4), 811–827.

Neubauer, A.C., & Fink, A. (2009). Intelligence and neural efficiency: Measures of brain activation versus measures of functional connectivity in the brain. *Intelligence*, *37*(2), 223–229.

Neubauer, A.C., Fink, A., & Grabner, R.H. (2006). Sensitivity of alpha band ERD to individual differences in cognition. In C. Neuper, & W. Klimesch (Eds.),

Progress in brain research: Event-related dynamics of brain oscillations. Vol. 159. (pp. 167–178). Elsevier.

Neubauer, A., Freudenthaler, H.H., & Pfurtscheller, G. (1995). Intelligence and spatiotemporal patterns of event-related desynchronization (ERD). *Intelligence*, *20*(3), 249–266.

Neubauer, A.C., Grabner, R.H., Freudenthaler, H.H., Beckmann, J.F., & Guthke, J. (2004). Intelligence and individual differences in becoming neurally efficient. *Acta Psychologica*, *116*(1), 55–74.

Neubauer, A.C., Sange, G., & Pfurtscheller, G. (1999). Psychometric intelligence and event-related desynchronization during performance of a letter matching task. *Handbook of Electroencephalography and Clinical Neurophysiology*, *6*, 219–231.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology*, *84*, 429–434.

Paas, F., Tuovinen, J.E., Tabbers, H., & Van Gerven, P.W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71.

Pfurtscheller, G., & Aranibar, A. (1977). Event-related cortical desynchronization detected by power measurements of scalp EEG. *Electroencephalography and Clinical Neurophysiology*, *42*, 817–826.

Raven, J.C. (1990). *Advanced progressive matrices sets 1 and 2*. Oxford: Oxford Psychologists Press.

Rypma, B., Berger, J.S., Prabhakaran, V., Martin Bly, B., Kimberg, D.Y., Biswal, B.B., et al. (2006). Neural correlates of cognitive efficiency. *NeuroImage*, *33*(3), 969–979.

Schloegl, A., Keinrath, C., Zimmermann, D., Scherer, R., Leeb, R., & Pfurtscheller, G. (2007). A fully automated correction method of EOG artifacts in EEG recordings. *Clinical Neurophysiology*, *118*(1), 98–104.

Shipstead, Z., Redick, T.S., & Engle, R.W. (2010). Does working memory training generalize? *Psychologica Belgica*, *50*(3–4), 245–276.

Shipstead, Z., Redick, T.S., & Engle, R.W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*(4), 628–654.

Smith, E.E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology*, *33*(1), 5–42.