Training and transfer effects of N-back training for brain-injured and healthy subjects

Jonas Kristoffer Lindeløv, Jonas Olsen Dall, Casper Daniel Kristensen, Marie Holt Aagesen, Stine Almgren Olsen, Therese Ruud Snuggerud, & Anna Sikorska

View supplementary material

Published online: 16 Feb 2016.

Submit your article to this journal

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

Training and transfer effects of *N*-back training for brain-injured and healthy subjects

Jonas Kristoffer Lindeløv, Jonas Olsen Dall, Casper Daniel Kristensen, Marie Holt Aagesen, Stine Almgren Olsen, Therese Ruud Snuggerud, and Anna Sikorska

Department of Communication and Psychology, Aalborg University, Aalborg, Denmark

ABSTRACT
Working memory impairments are prevalent among patients with acquired brain injury (ABI). Computerised training targeting working memory has been researched extensively using samples from healthy populations but this field remains isolated from similar research in ABI patients. We report the results of an actively controlled randomised controlled trial in which 17 patients and 18 healthy subjects completed training on an *N*-back task. The healthy group had superior improvements on both training tasks (SMD = 6.1 and 3.3) whereas the ABI group improved much less (SMD = 0.5 and 1.1). Neither group demonstrated transfer to untrained tasks. We conclude that computerised training facilitates improvement of specific skills rather than high-level cognition in healthy and ABI subjects alike. The acquisition of these specific skills seems to be impaired by brain injury. The most effective use of computer-based cognitive training may be to make the task resemble the targeted behaviour(s) closely in order to exploit the stimulus-specificity of learning.

Introduction

Randomised controlled studies on computer-based cognitive rehabilitation of brain injured patients goes back to at least Sturm, Dahmen, Hartje, and Willmes (1983) and more than 50 randomised controlled trials (RCTs) have been published so far (Chen, Thomas, Glueckauf, & Bracy, 1997; Lindeløv, in press). There is a much larger and growing literature on computerised cognitive training in healthy subjects (Melby-Lervåg & Hulme, 2012; Morrison & Chein, 2010; Shipstead, Redick, & Engle, 2012). However, these two fields have hitherto proceeded in parallel with no cross-talk or direct comparisons. Computerised cognitive neurorehabilitation could potentially expand its evidence base considerably if there are points of convergence with healthy subjects. After all, all subjects are humans and all would benefit from improved information processing capacities.

Unfortunately, enhancement of domain-general cognitive functions such as working memory and attention has proven difficult in brain injured (Lindeløv, in press; Park & Ingles, 2001) as well as healthy subjects (Melby-Lervåg & Hulme, 2012; Shipstead et al., 2012) where the sum of evidence shows a dominance of domain-specific effects of rehabilitation efforts. In other words, there is relatively little transfer to untrained material and contexts. Still, there is an ongoing search for interventions that could promote far transfer, one of which is variations of computer-based cognitive rehabilitation. Efforts have been made to identify "active ingredients" that promote far transfer. *Adaptiveness* has generally been found to contribute positively in healthy subjects (Jaeggi et al., 2010; Klingberg, 2010; Morrison & Chein, 2010) with non-adaptive conditions now being used as active control groups (see, e.g., Holmes, Gathercole, & Dunning, 2009) although adaptiveness in and of itself is not sufficient as demonstrated by several null results (Chooi & Thompson, 2012; Jaeggi, Buschkuehl, Jonides, & Shah, 2011; Redick et al., 2012). Other factors, such as training intensity, duration, and spacing of sessions, could be influential but the results are ambiguous (Morrison & Chein, 2010).

There is a general distinction between training effects and transfer effects. This distinction is also known as domain-specific vs. domain-general and near-transfer vs. far-transfer. For example, working memory is said to be domain-general because it operates cross-modally on a wide range of stimuli and contexts (Baddeley, 2007; Cowan, 1988; Kane et al., 2004). An improvement of working memory would then by definition lead to an improvement on all behaviours that rely on working memory. If that improvement was brought about by training only a subset of behaviours, such an effect would be far transfer. In contrast, domain-specific processes apply to a narrow range of stimuli and contexts and the challenge for all studies is to provide convincing evidence that observed improvements are not mere training effects. For example, Westerberg et al. (2007) and Lundqvist, Grundström, Samuelsson, and Rönnberg (2010) administered CogMed training to ABI patients and observed improvements on digit span and a spatial span task but both were part of the training programme. Therefore, these results could be attributed to mere training effects even though they were interpreted as transfer to working memory. Others explicitly assessed the testing–training similarity and found narrow transfer effects (Sturm & Willmes, 1991). In a rehabilitation setting, the primacy of training effects over transfer effects supports a preference for compensatory interventions over remediatory interventions.

In the present study we administered an *N*-back training procedure to a healthy sample and an ABI sample. The *N*-back has been used in numerous studies in healthy subjects with initial large positive transfer (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Jaeggi et al., 2010) but later studies have yielded null results (Chooi & Thompson, 2012; Jaeggi et al., 2011; Redick et al., 2012). An *N*-back-like intervention has only been administered to ABI patients in one small study by Cicerone (2002), which yielded very large positive effects on untrained tasks in the order of $d > 2$. The Cicerone study was heavily therapist-directed and included efforts to transfer to the patient's everyday life.

From this, we derived three research questions: (1) Can the Cicerone findings be attributed to just doing an adaptive *N*-back task? (2) Does the *N*-back task promote transfer in healthy subjects? (3) To what extent can evidence be generalised from healthy subjects to ABI patients and vice versa—at least for the *N*-back task?

Methods

Sample

We recruited 39 ABI inpatients at Hammel Neurorehabilitation Centre and University Research Clinic. The choice of inpatients was motivated by the fact that cognitive impairments reduce the intensity and duration of other inpatient rehabilitation efforts. Therefore, an early cognitive improvement has the potential to support neurorehabilitation in general. Patients engaged with the training, in addition to treatment as usual, at a time of their own choosing, usually in the late afternoon. Seventeen patients completed the training (see Figure 1). Additional inclusion criteria were (1) no aphasia, deafness, blindness or other disabilities that would prevent testing and training, (2) the patient should be able perform reasonably ($d' > 1$) at N-back level 1 and Visual Search (VS) level 2 at the time of recruitment, and (3) that the training did not interfere with the standard treatment as judged by the patient's primary therapists.

We recruited 39 healthy participants who trained in their free time using their own personal computer. They were predominantly psychology students recruited using flyers and posts on virtual forums. Eighteen completed the training (see Figure 2). See Table 1 for descriptions of the recruited and final sample.

This study was approved by the local ethics committee, all participants signed informed consent and participation was voluntary. Participants were not reimbursed. Patients were informed that the training was designed to facilitate their cognitive recovery and healthy subjects were informed that it was designed to boost their intelligence.

Design and randomisation

This is a parallel group design with a 2 (ABI/healthy) × 2 (N-back/VS) design resulting in four treatment arms. Participants entered the study continuously and were pseudo-randomly allocated to N-back and VS so that pre-test scores on Raven's Advanced Progressive Matrices (see Outcome Measures) and ages were balanced. This allocation took place independently for patients and healthy subjects and allocation of the first four
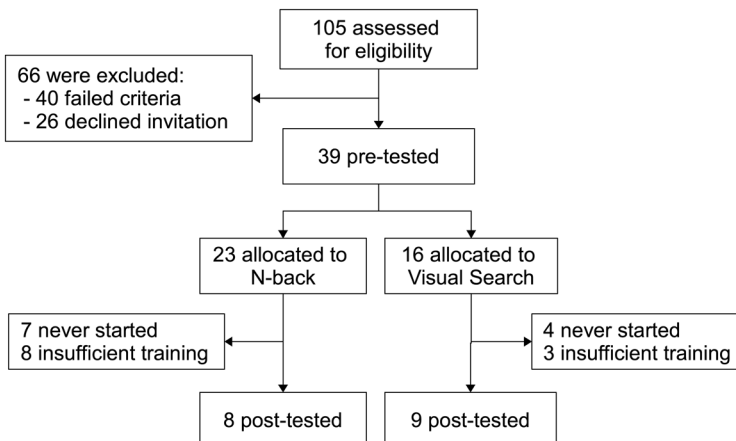

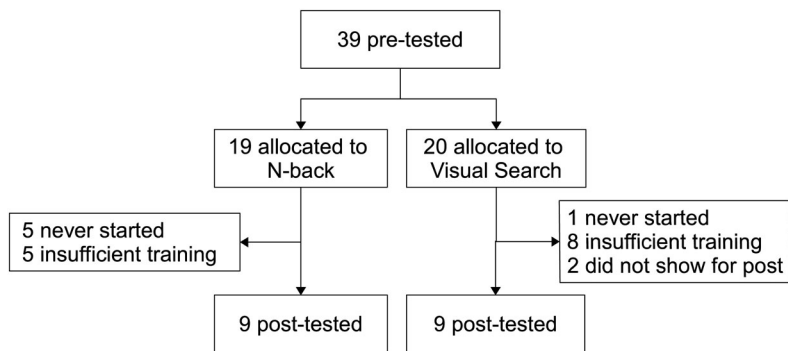Figure 1. Study design and flowchart for the patient group.

Figure 2. Study design and flowchart for the healthy group.

participants within each of these groups was truly random. When 20 training sessions were completed, participants were scheduled for a post-test.

N-back training

The single N-back task consisted of a series of stimuli presented at 3-second intervals. The participants were instructed to press a key when the presented stimulus was identical to the stimulus N back in the sequence. There were 25% targets per block and at most two consecutive targets. In order to prevent the formation of stimulus-specific strategies there were a total of 137 different stimuli as shown in Figure 1: three types of audio and four types of visual stimuli. A random selection of eight stimuli from a randomly selected stimulus type was chosen for each block (See Figure 3).

Visual search training

Participants were instructed to press a key if a target symbol was present in an $N \times N$ array of symbols. The target symbol changed from block to block but there were just six different symbols. The VS task is unrelated to working memory (Kane, Poole, Tuholski, & Engle, 2006) and served as an active control condition. It has served this purpose in other training studies (Harrison et al., 2013; Redick et al., 2012). During training, levels

Table 1. Sample descriptives at baseline for each group.

	Group	Included	Finished	N-back	VS
Males/Females	ABI	31/8	13/4	6/2	7/2
	Healthy	18/21	8/10	3/6	5/4
Age in years	ABI	53.3 (10.4)	56.1 (6.3)	56.1 (5.6)	56.1 (7)
	Healthy	27.4 (10.3)	29.3 (11.3)	29.2 (11.1)	29.4 (11.9)
Days since injury	ABI	54 [28–94]	57 [33–95]	63 [35–89]	55 [33–95]
FIM cognitive[a] (0–35)	ABI	24 [22–28]	26 [23–28]		
FIM motor[a] (0–91)	ABI	82 [60–89]	82 [67–89]		

Mean (SD) were normally distributed, and frequencies [square brackets] were not.
Included column are intention-to-treat participants. Finished column are participants who eventually finished. N-back and VS columns subdivides the finished columns to the two treatments. VS = Visual search; ABI = acquired brain injury group. FIM = Functional Independence Measure.
[a]FIM scores at baseline were only available for 18 patients and should therefore be regarded as a rough indication of the patient group's functional level rather than a sample descriptive. FIM scores are not shown for individual treatments since only four of the patients in each group who finished had a baseline FIM score.
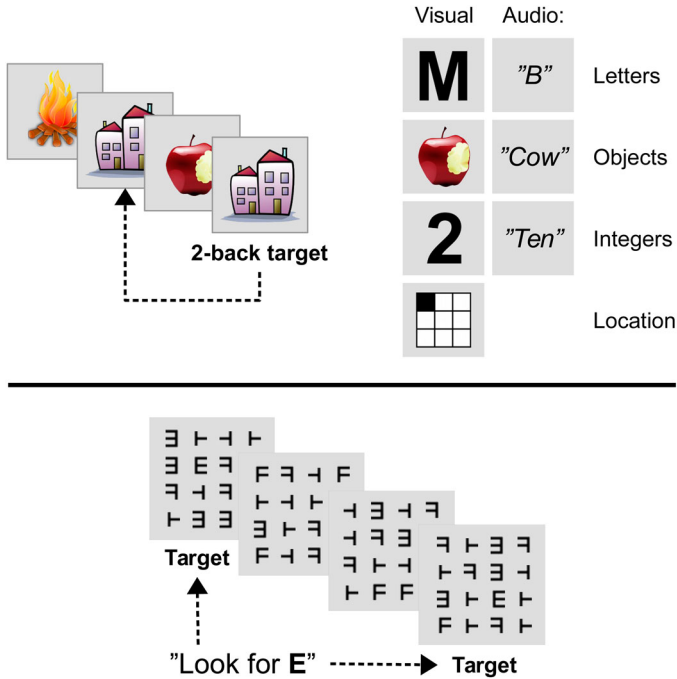
Figure 3. Top left: Four trials of the *N*-back task illustrated at 2-back level with image stimuli. Top right: seven different stimulus types. Bottom: Four trials of the Visual Search task at level 4 (4 × 4 grid) and with "E" as a target. For each block any of the six different shapes would be picked randomly as target. Participants were instructed to press a key on target trials.

increased from *N* = 1,2,3,4 ... etc. but we use the number of items to be searched (*N* = 1,4,9,16, ...) as a measure of difficulty level throughout in this paper.

Both training tasks

Tasks were kept similar in all other respects to maximise the purity of the contrast. Participants trained 12 blocks of 20 + *N* trials on an adaptive *N*-back task for 20 days. Less than 10 blocks was considered an incomplete training day. The interstimulus interval was 3 seconds. Thus the full intervention consisted of 4.4 hours of constant training, breaks not included. Participants trained unsupervised in a laptop web browser.

Visual correct/incorrect feedback was given on response (hit/false alarm) and on misses in the end of every trial. To increase motivation, participants were awarded points after each block and given progressively more attractive titles. Feedback sounds were played on level upgrade or downgrade. In addition, participants would see a graph of their own progress.

The difficulty level was adjusted after each block based on *d*-prime from Signal Detection Theory, i.e., the participant's ability to discriminate target trials from non-target trials. The level was downgraded if *d*-prime was below 1.2 and upgraded if *d*-prime was above 1.8.

All participants were given graphic instructions for the tasks and were encouraged by telephone to start training if they had two successive non-training weekdays. Patients

initially trained three blocks under the guidance of a research assistant. All participants could call for technical help or to get an update on the task instructions.

Outcome measures

We investigate the hypothesis that *N*-back training improves working memory and perhaps fluid intelligence while VS improves processing speed. The two intervention groups serve as each other's controls in this respect since the search task is unrelated to working memory (Kane et al., 2006) and signal detection in the *N*-back task is relatively unrelated to processing speed (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002).

Participants were tested on the following measures before and after training. Although each is related to many cognitive abilities, they are grouped according to the cognitive labels usually assigned to them:

(1) *Fluid intelligence:* Equal and unequal items from the Raven's Advanced Progressive Matrices (RAPM) were administered at pre-test and post-test in counterbalanced order (Raven, Raven, & Court, 1962). Participants were given 10 minutes to solve the 18 items in each set. RAPM has excellent construct validity with respect to fluid intelligence as determined by latent variable analysis (Engle, Tuholski, Laughlin, & Conway, 1999).
(2) *Working memory:* The Wechsler Adult Intelligence Scale–IV (WAIS-IV) Working Memory Index (WMI), calculated from forwards/backwards/ordered digit span, mental arithmetic and letter-number sequencing (Wechsler, 2008). The latter is optional and was skipped for some fatigued patients.
(3) *Working memory:* A computerised Operation Span (Unsworth, Heitz, Schrock, & Engle, 2005) with $3 \times$ span 2–4. Span 2–5 is standard for this test but span 5 was omitted in this experiment since all four pilot ABI patients experienced great distress from these trials to a degree where they gave up in advance during recall. Each participant was scored using the partial credit unit scoring method (Conway et al., 2005) which is the average proportion of items recalled in the correct location in each trial.
(4) *Processing speed:* The WAIS-IV Processing Speed Index (PSI), calculated from symbol search and digit-symbol coding. PSI and WMI have high internal consistencies and re-test reliabilities (Iverson, 2001).
(5) *Processing speed with inhibition:* 180 trials on a computerised Stroop task of which 20% were incongruent. Participants responded verbally to maximise interference (Liotti, Woldorff, Perez, & Mayberg, 2000) while pressing a key to register reaction time. The Stroop effect is classically regarded as a measure of inhibition task but the raw reaction time and to some extent also the Stroop effect itself aligns well with a processing speed construct (Salthouse & Meinz, 1995; Verhaeghen & De Meersman, 1998). Both are of interest with respect to the visual search training.

The Operation Span Test and the Stroop Test were computerised using PsychoPy v. 1.79 (Peirce, 2007) and administered in a separate test session.

Statistical models and inferences

Outcome data were modelled as mixed models with main effects of time, treatment and group and their interactions in R 3.1.2 using the lmer (Bates, Maechler, Bolker, & Walker,

2014) 4.1.1 and BayesFactor (Morey & Rouder, 2015) 0.9.10 packages. There was a random intercept per participant to account for correlations between repeated measures. Inference was based on model selection between a full model and a null model. The null model was the full model less the fixed effect in question, e.g., the three-way interaction.

Training data were modelled using a power function ($N = k + ax^b$) for task level (N) as a function of time (block number) with random intercepts (k) per participant and random a and b parameters for intervention and group respectively to reflect differences in gain.

P-values from the chi-square statistic of a likelihood ratio test (LRT) (Barr, Levy, Scheepers, & Tily, 2013) were reported to comply with current publishing practices. It has long been known that p does not quantify evidence for or against the null and therefore has poor inferential value (Berger & Sellke, 1987; Sellke, Bayarri, & Berger, 2001; Wetzels et al., 2011). Bayes factors (BF) with a relatively uninformative Cauchy $(0,(\sqrt{2}/2))$ prior on each covariate were used to quantify the relative evidence for each model (Rouder & Morey, 2012) with the exception of the power function where a more uninformative unit information prior (Wagenmakers, 2007) was used for computation convenience. A BF is the odds ratio between two models. For example, BF = 5 means that these data shift the odds 5:1 in favour of the full model and simultaneously shift the odds $5^{-1} = 0.2$ for the null.

Results

Training tasks

Participants' progression on the training is shown in Figure 4 and effect sizes in Table 2. The participants mostly trained on consecutive days: 72% of the training sessions were on consecutive days and 93% within three days. Block number was used as the time unit with a total of 200–240 blocks for completers (20 days × 12 blocks per day). The power fits for individual participants are superposed on the data in Figure 4. This model was by far preferred to an intercept-only model for all four group × treatment cells (p_{LRT} << 0.001, BF_{BIC} >> 1000) providing evidence that there was improvement on the training task in all conditions, i.e., the power model is much more than 1000 times more likely to have generated the data than the intercept-only model. This is of crucial theoretical importance since it establishes the training as a potential source of transfer to the outcome measures. It was also preferred to a linear model with the same random-effects structure (p_{LRT} < < 0.001, BF_{BIC} >> 1000).

It is apparent from Figure 4 that the healthy group had a much greater numerical improvement on the training tasks than the ABI group. This was confirmed by a

Table 2. Pre- and post-test means, standard deviations and standardised differences for task level N.

Group	N-back			VS		
	Pre	Post	SMD	Pre	Post	SMD
ABI	2.1 (1)	2.6 (0.8)	0.45	13.9 (8.1)	22.5 (11.9)	1.06
Healthy	2.7 (0.8)	7.5 (3.5)	6.11	20.6 (8.7)	49.7 (24.3)	3.34

"Pre" was computed from the first 12 blocks starting from the first block where the participant did not increase a level. "Post" was computed from the last 12 blocks. Given the large sample of blocks, even very small effects would trivially would fall under the full model (p_{LRT} < < 0.001 and BF_{BIC} >> 1000) so we consider effect sizes is the most informative statistic here. VS = Visual search; SMD = standardised mean difference.
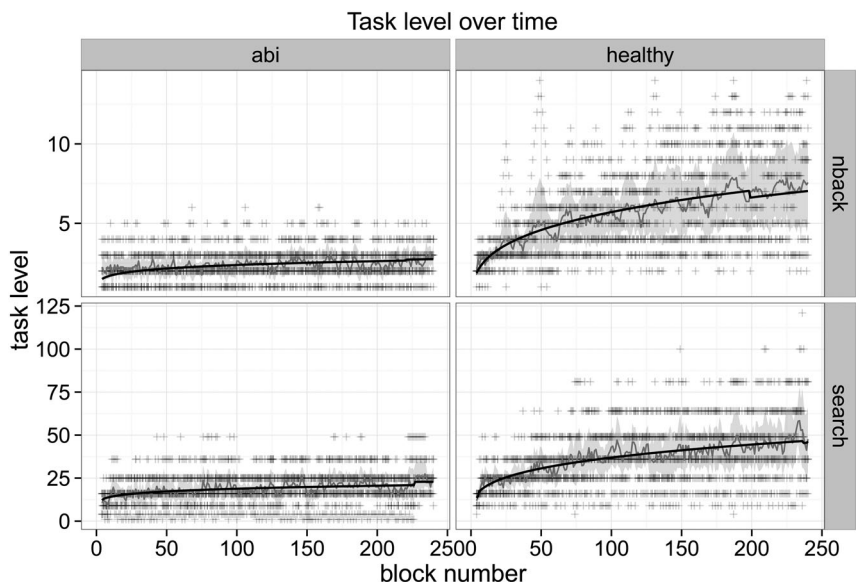
Figure 4. Level of training task as a function of block number (out of 240 in total) for completers in each group by task. The ABI group was consistently at a lower level and improved little on the training tasks whereas the healthy group improved on both tasks. The thick black line is the average of the predictions from a fitted power function. The thin grey line and the grey area are the means and 95% bootstrapped confidence intervals for the mean level at each block number. The transparent "+" symbols are the data from individual participants which the above represents. Two healthy completers improved to $N = 27$ and $N = 45$ on the N-back task and were not plotted. The break in the power function around block 200 is caused by participants who stopped training there. Plots for all participants are available in supplementary Figure S1.

substantial group-specific testing of the random effects of group on a and b ($p_{LRT} <<$ 0.001, $BF_{BIC} >> 1000$).

Outcome measures

See Table 3 for descriptives and inferences on outcome measures. A difference in effect between the VS training and the N-back training was not supported by the data on any outcome measure, neither in the ABI group nor the healthy group. All BFs favoured the model without the treatment × time interaction with a BF of about 2:1 against the full model. This BF is suggestive but too weak to say anything definitive about the direction of the effect as it still puts around 1/3 posterior probability on the full interaction model. Furthermore, the controlled standardised mean differences ($SMD_c = SMD_{nback} - SMD_{search}$) were in the zero-to-small range (around 0.0–0.3) with no signs of a differential effect between single tests or cognitive domains.

Interestingly, the data suggest that the ABI and healthy group did not differ in their gains on any outcome as revealed by inferences on the group × time × test interaction term (see Table 3). BFs again favoured the null more than the alternative, and no interactions were statistically noticeable ($p > 5\%$).

As expected, an analysis of all 39 ABI patients and 39 healthy subjects at baseline revealed that the patients performed worse than the healthy subjects on all outcome measures (all ps below .01 and all BFs above 3) except for Operation Span where,

Table 3. Descriptives, effect sizes, and inferential statistics on all outcome measures.

Test	Group	N-back				VS				time × treat			ti × tr × grp	
		Pre	SD	Post	SD	Pre	SD	Post	SD	SMD	p	BF$_g$	p	BF$_g$
Digit span	ABI	7.25	4.53	8.38	4.17	7.78	3.19	8.22	2.99	0.18	0.41	1.92*		
	Healthy	10.44	3.43	10.67	2.83	10.50	2.39	9.88	2.30	0.29	0.32	1.74*	0.89	2.34*
Arithmetic	ABI	8.75	3.69	9.88	4.02	9.88	3.27	9.75	2.76	0.37	0.14	1.22*		
	Healthy	9.56	2.70	10.00	2.50	9.75	2.87	10.25	2.87	−0.02	0.92	2.49*	0.2	1.21*
Num-let	ABI	10.67	5.09	12.00	4.10	9.38	2.33	9.88	2.10	0.23	0.58	2.26*		
	Healthy	8.56	2.46	10.00	3.16	9.50	3.02	9.62	3.11	0.49	0.09	1.07	0.76	2.05*
WMI index	ABI	88.00	22.06	92.75	23.29	91.56	16.19	89.67	14.75	0.36	0.14	0.17*		
	Healthy	96.78	15.75	101.44	15.39	99.22	18.34	101.22	14.77	0.16	0.43	1.94*	0.47	2.08*
Search	ABI	7.12	2.03	8.25	1.83	7.00	2.40	7.56	2.60	0.26	0.36	1.89*		
	Healthy	11.56	2.55	13.00	4.30	11.88	3.14	13.62	3.70	−0.11	0.81	2.41*	0.53	1.92*
Coding	ABI	6.88	3.23	8.12	2.53	7.00	2.62	8.25	2.43	0	1	2.42*		
	Healthy	11.89	3.06	13.89	2.98	10.62	3.46	12.75	3.85	−0.04	0.88	2.16*	0.91	2.94*
PSI	ABI	84.12	13.86	92.00	10.10	83.00	14.00	87.00	14.90	0.29	0.23	1.66*		
	Healthy	109.56	15.47	121.00	15.27	107.00	13.87	119.00	18.73	−0.04	0.88	2.34*	0.38	1.93*
RAPM	ABI	3.75	2.66	3.25	2.71	4.11	1.90	4.11	1.36	−0.23	0.59	2.33*		
	Healthy	8.89	2.80	10.67	2.50	8.78	4.76	9.89	3.79	0.18	0.66	2.37*	0.51	2.29*
OSPAN	ABI	74.69	21.94	74.07	15.16	71.30	9.15	70.52	19.80	0.01	0.99	2.36*		
	Healthy	59.26	16.85	64.40	20.30	65.02	10.35	69.75	12.42	0.03	0.94	2.38*	0.98	2.28*
Log(Stroop)	ABI	0.42	0.13	0.35	0.11	0.42	0.03	0.39	0.05	−0.52	0.44	1.68*		
	Healthy	0.31	0.12	0.28	0.12	0.26	0.07	0.25	0.12	−0.17	0.7	2.4*	0.84	1.88*
Log(str RT)	ABI	6.89	0.04	6.94	0.09	6.97	0.11	6.95	0.04	0.7	0.33	1.42*		
	Healthy	6.52	0.27	6.41	0.29	6.66	0.19	6.54	0.11	0.03	0.93	2.6*	0.65	1.87*

N is the number of completers for this group × treatment cell. SMD$_c$ is the controlled standardised effect size, computed using pre-test standard deviation. P-values are from the likelihood-ratio test of the critical interaction and BF$_g$ = Bayes Factor with g-priors on regression coefficients (Rouder & Morey, 2012).
*BF$_g$ is in favour if the null (1/BF$_g$). Briefly, "time × treat" answers the question: "Is the task associated with different gains for this group?" and "time × treat × group" answers the question "Is the controlled N-back gain different between groups?" Stroop reaction times are reported in logarithmic units since they were approximately log-normal.
Means, standard deviations and univariate inferential results are for all outcome measures.
WMI = Working Memory Index; PSI = Processing Speed Index; RAPM = Raven's Advanced Progressive Matrices; OSPAN = Operation Span.

surprisingly, patients performed better than the healthy subjects (mean difference = 10.9%, CI = 0.6–21.2%, p_{ttest} = 0.038, BF = 1,9). The strength of the evidence for the latter is, however, anecdotal—especially in light of the other results.

Discussion

We observed an improvement on the training task but no transfer to untrained tasks. Therefore, with respect to our first research question, the positive findings of Cicerone's (2002) N-back intervention on ABI patients do not seem to be attributable to the N-back task itself. In light of the present results, the Cicerone results are more likely to have arisen from the therapist-directed activities tailored to each individual patient. Such activities were absent in the present study.

With respect to our second research question about the N-back literature on healthy adults, these results fail to replicate some positive findings by Jaeggi et al. (2008, 2010) but are in line with several null findings (Chooi & Thompson, 2012; Jaeggi et al., 2011; Redick et al., 2012). The former studies used passive control groups while the latter used active control groups. It is possible that the Jaeggi et al. results were caused by nonspecific factors in the training, such as expectation and motivation, which cannot be attributed to the task itself. If that is the case, these studies do not constitute evidence for a transfer effect.

The remainder of the discussion pertains to our third research question: To what extent can evidence be generalised from healthy subjects to ABI patients and vice versa?

Convergence: Computerised training yields specific effects

N-back and VS training did not differentially improve performance on neuropsychological tests which are thought to reflect working memory, processing speed or fluid intelligence. Since previous research has shown that N-back performance reflects working memory and that VS does not (Kane et al., 2006) and we observe no selective effect of N-back improvement on working memory measures, we conclude that participants developed specific strategies to solve the N-back task during the course of training. These strategies were so specific that N-back training on digits did not transfer to digit span (part of WAIS Working Memory Index); N-back training on letters did not transfer to Operation Span letter recall, as was also found by Jaeggi et al. (2010) and Redick et al. (2012); VS training did not transfer to Symbol Search (part of WAIS processing speed index); and N-back training on locations in a 3 × 3 grid did not transfer to the RAPM 3 × 3 grid, as opposed to Jaeggi et al. (2008, 2010). These findings resonate with other computer-based rehabilitation studies on ABI patients. For example, Åkerlund, Esbjörnsson, Sunnerhagen, and Björkdahl (2013) found a delayed positive effect on one trained outcome measure on working memory but not three other untrained outcome measures as compared to the control group. Lundqvist et al. (2010) similarly demonstrated improvements on measures which resembled the training, although a missing analysis of interaction with the control group renders it unassessed to what extent these improvements were superior to the improvements in the control group. Possibly the most explicit assessment of near- versus far-transfer effects was carried out by Sturm and Willmes (1991) who ordered the outcome measures according to their hypothesised similarity with the training. Here too, a specific-effects dominance was observed with no transfer to dissimilar outcome measures.

It is clear that these data pose problems for the naïve view that abstract cognitive abilities were acquired during the training and even the view that something intermediate was trained, such as recalling sequential items or scanning visual grids. Instead, we believe that computerised training follows the well-known learning principle that the efficiency of the decoding/use of a skill is proportional to the similarity of the decoding context to the encoding/learning context of this skill (Perkins & Salomon, 1989; Tulving & Thomson, 1973). The repetitive nature of computerised training represents a highly stable context and thus the N-back and VS skills become "locked" to this context in a way that does not generalise to a neuropsychological test setting or even the computerised tests.

This is evidence that the development of specific strategies that do not transfer to untrained tasks might be a possible point of convergence between healthy subjects and ABI patients. Although specific improvement may seem unflattering compared to generalised improvement, it could actually be thought of as a very efficient information processing strategy where cheap local strategies are preferred to slow and costly high-level cognition (Clark, 2013; Friston, 2010). As such, the tendency and ability to develop specific strategies could be regarded as a property of a healthy cognitive system.

Divergence: The formation of specific skills

The healthy group improved 2.5–5.5 SMD more on the training tasks than the ABI group. We interpret the training data to reflect an impaired ability to form specific strategies in the ABI group. One explanation for this observation is that well-functioning domain-general cognition is necessary for the effective formation of specific strategies. However, five out of 17 ABI patients had a baseline Working Memory Index score over 100 and two patients had a Processing Speed Index score over 100. But neither of these improved nearly as much on the training task as the average person in the healthy group, thus discrediting this hypothesis.

The group differences in aetiologies could be confounded by the age difference. However, Dahlin, Nyberg, Bäckman, and Neely (2008) gave computerised training to healthy young and elderly adults and found no difference in gain on the training task. A meta-analysis on a 26 computer-based parallel-groups RCT on ABI patients similarly found no effect of age on improvement (Lindeløv, in press). Thus both within-study and between-study evidence discredits age as the sole explanation for the observed discrepancy between the ABI and the healthy group.

This leaves us in a limbo with no single candidate explanation for the observed difference between groups. There is a vast literature on the topic of specific learning impairments following acquired brain injury (Schmitter-Edgecombe, 2006; Vanderploeg, Crowell, & Curtiss, 2001). However, it has almost exclusively investigated verbal and motor learning within single sessions. Four weeks of training on the N-back task and the VS tasks do not readily subsume under these categories so the present study may contribute new evidence. It is up to future studies to narrow in on the mechanisms driving this effect. To support this effort, we encourage authors to report and interpret training data explicitly when doing this type of study.

Limitations

The final sample size per condition is small, even though a total of 78 participants started the training. The present study should not be considered a basis for clinical

guidelines but rather as preliminary evidence which puts a few ideas about the mechanisms underlying computer-based training on the table. Although small, it is not smaller than the two most cited studies on the topic (Sturm, Willmes, Orgass, & Hartje, 1997; Westerberg et al., 2007).

The sample size was influenced by a large dropout rate of 50% which was not biased with respect to age, gender, baseline scores or (for patients) Functional Independence Measure score. The dropout rate suggests that low adherence should be expected for fully self-initiated training without monetary reward for healthy subjects and ABI patients alike. Lack of motivation (too hard/boring) was the primary reason for dropout. Thus the final sample might be biased towards higher motivation and consequently the effect sizes reported here might be somewhat positively biased.

As with any experiment, the results could be attributed to the specificities of the treatments. In particular, 20 minutes of unsupervised training per day is relatively short. One study, which had more than 500 1-hour therapist-assisted training sessions, showed no generalised improvement (Middleton, Lambert, & Seggar, 1991), demonstrating that more training does not necessarily promote transfer.

The optimal evidence in support of the conclusions above would have been obtained if the ABI group and the healthy group had been matched on all nuisance parameters such as age, education and socioeconomic status.

Future directions

We suggest that future research on computerised cognitive rehabilitation may progress along two different routes: First, prevention of specific learning. This is not easy. Simply training on a large array of different tasks may not be sufficient as demonstrated by several null findings from those who used this strategy (Chen et al., 1997; Middleton et al., 1991). A true context-breaking intervention would constantly present novel problems, shift between devices, change colours, be trained at different locations, etc. We expect that this approach is too chaotic to be feasible with ABI patients. Second, alternatively, exploit the context-specific effects and make the training task as similar to the transfer target as possible, i.e., practise reading television subtitles, doing mental arithmetic on shopping costs, etc. For example, Yip and Man (2013) successfully improved real-life shopping performance after training in a matching virtual reality environment. This is a much less ambitious target than high-level cognition but may also be more realistic.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Åkerlund, E., Esbjörnsson, E., Sunnerhagen, K. S., & Björkdahl, A. (2013). Can computerized working memory training improve impaired working memory, cognition and psychological health? *Brain Injury*, 27(13-14), 1649–1657. http://doi.org/10.3109/02699052.2013.830195

Baddeley, A. (2007). *Working memory, thought, and action* (1st ed.). Oxford: Oxford University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. http://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. Retrieved from http://CRAN.R-project.org/package = lme4

Berger, J. O., & Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of P values and evidence. Journal of the American Statistical Association, 82(397), 112. http://doi.org/10.2307/2289131

Chen, S. H. A., Thomas, J. D., Glueckauf, R. L., & Bracy, O. L. (1997). The effectiveness of computer-assisted cognitive rehabilitation for persons with traumatic brain injury. Brain Injury, 11(3), 197–210.

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

Cicerone, K. D. (2002). Remediation of "working attention" in mild traumatic brain injury. Brain Injury, 16(3), 185–195. http://doi.org/10.1080/02699050110103959

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36(03), 233–253. http://doi.org/10.1017/S0140525X12000477

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

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.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. Psychological Bulletin, 104(2), 163–191. http://doi.org/10.1037/0033-2909.104.2.163

Dahlin, E., Nyberg, L., Bäckman, L., & Neely, A. S. (2008). Plasticity of executive functioning in young and older adults: Immediate training gains, transfer, and long-term maintenance. Psychology and Aging, 23(4), 720–730.

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. http://doi.org/10.1037/0096-3445.128.3.309

Friston, K. (2010). The free-energy principle: A unified brain theory? Nature Reviews Neuroscience, 11(2), 127–138. http://doi.org/10.1038/nrn2787

Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. Psychological Science, 24(12), 2409–2419. http://doi.org/10.1177/0956797613492984

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. Developmental Science, 12(4), F9–F15. http://doi.org/10.1111/j.1467-7687.2009.00848.x

Iverson, G. L. (2001). Interpreting change on the WAIS-III/WMS-III in clinical samples. Archives of Clinical Neuropsychology, 16(2), 183–191.

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.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. Proceedings of the National Academy of Sciences, 108(25), 10081–10086. http://doi.org/10.1073/pnas.1103228108

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y. F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning–implications for training and transfer. Intelligence, 38, 625–635.

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. http://doi.org/10.1037/0096-3445.133.2.189

Kane, M. J., Poole, B. J., Tuholski, S. W., & Engle, R. W. (2006). Working memory capacity and the top-down control of visual search: Exploring the boundaries of "executive attention". Journal of Experimental Psychology: Learning, Memory, and Cognition, 32(4), 749–777. http://doi.org/10.1037/0278-7393.32.4.749

Klingberg, T. (2010). Training and plasticity of working memory. Trends in Cognitive Sciences, 14(7), 317–324. http://doi.org/10.1016/j.tics.2010.05.002

Lindeløv, J. (in press). Computer-based cognitive rehabilitation following acquired brain injury: A meta-analysis of transfer effects. Clinical Psychology Review.

Liotti, M., Woldorff, M. G., Perez, R.III & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia*, 38, 701–711.

Lundqvist, A., Grundström, K., Samuelsson, K., & Rönnberg, J. (2010). Computerized training of working memory in a group of patients suffering from acquired brain injury. *Brain Injury*, 24(10), 1173–1183. http://doi.org/10.3109/02699052.2010.498007

Melby-Lervåg, M., & Hulme, C. (2012). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, 49(2), 270–291. http://doi.org/10.1037/a0028228

Middleton, D. K., Lambert, M. J., & Seggar, L. B. (1991). Neuropsychological rehabilitation: Microcomputer-assisted treatment of brain-injured adults. *Perceptual and Motor Skills*, 72(2), 527–530.

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of Bayes factors for common designs*. Retrieved from http://CRAN.R-project.org/package = BayesFactor

Morrison, A. B., & Chein, J. M. (2010). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, 18(1), 46–60. http://doi.org/10.3758/s13423-010-0034-0

Park, N. W., & Ingles, J. L. (2001). Effectiveness of attention rehabilitation after an acquired brain injury: A meta-analysis. *Neuropsychology*, 15(2), 199–210.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, 162 (1-2), 8–13. http://doi.org/10.1016/j.jneumeth.2006.11.017

Perkins, D., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher*, 18(1), 16–25. http://doi.org/10.3102/0013189X018001016

Raven, J. C., Raven, J. C., & John Hugh Court. (1962). *Advanced progressive matrices*. London: HK Lewis.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., … Engle, R. W. (2012). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*. http://doi.org/10.1037/a0029082

Rouder, J. N., & Morey, R. D. (2012). Default Bayes factors for model selection in regression. *Multivariate Behavioral Research*, 47(6), 877–903. http://doi.org/10.1080/00273171.2012.734737

Salthouse, T. A., & Meinz, E. J. (1995). Aging, inhibition, working memory, and speed. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 50B(6), P297–P306. http://doi.org/10.1093/geronb/50B.6.P297

Schmitter-Edgecombe, M. (2006). Implications of basic science research for brain injury rehabilitation: A focus on intact learning mechanisms. *The Journal of Head Trauma Rehabilitation*, 21(2), 131–141.

Sellke, T., Bayarri, M. J., & Berger, J. O. (2001). Calibration of p values for testing precise null hypotheses. *The American Statistician*, 55(1), 62–71. http://doi.org/10.1198/000313001300339950

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, 138(4), 628–654. http://doi.org/10.1037/a0027473

Sturm, W., Dahmen, W., Hartje, W., & Willmes, K. (1983). Ergebnisse eines Trainingsprogramms zur Verbesserung der visuellen Auffassungsschnelligkeit und Konzentrationsfähigkeit bei Hirngeschädigten. *Archiv Für Psychiatrie Und Nervenkrankheiten*, 233(1), 9–22.

Sturm, W., & Willmes, K. (1991). Efficacy of a reaction training on various attentional and cognitive functions in stroke patients. *Neuropsychological Rehabilitation*, 1(4), 259–280. http://doi.org/10.1080/09602019108402258

Sturm, W., Willmes, K., Orgass, B., & Hartje, W. (1997). Do specific attention deficits need specific training? *Neuropsychological Rehabilitation*, 7(2), 81–103.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5), 352–373.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37(3), 498–505.

Vanderploeg, R. D., Crowell, T. A., & Curtiss, G. (2001). Verbal learning and memory deficits in traumatic brain injury: Encoding, consolidation, and retrieval. *Journal of Clinical and Experimental Neuropsychology (Neuropsychology, Development and Cognition: Section A)*, 23(2), 185–195. http://doi.org/10.1076/jcen.23.2.185.1210

Verhaeghen, P., & De Meersman, L. (1998). Aging and the stroop effect: A meta-analysis. *Psychology and Aging*, 13(1), 120–126. http://doi.org/10.1037/0882-7974.13.1.120

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14(5), 779–804. http://doi.org/10.3758/BF03194105

Wechsler, D. (2008). *Wechsler adult intelligence scale - Fourth edition*. San Antonio: Pearson.

Westerberg, H., Jacobaeus, H., Hirvikoski, T., Clevberger, P., Östensson, M.-L., Bartfai, A., & Klingberg, T. (2007). Computerized working memory training after stroke–A pilot study. *Brain Injury*, *21*(1), 21–29. http://doi.org/10.1080/02699050601148726

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*, *6*(3), 291–298. http://doi.org/10.1177/1745691611406923

Yip, B. C., & Man, D. W. (2013). Virtual reality-based prospective memory training program for people with acquired brain injury. *Neurorehabilitation*, *32*(1), 103–115.