Title

(Un)Great Expectations: The Role of Placebo Effects in Cognitive Training

Permalink

https://escholarship.org/uc/item/0f30s0s5

Journal

Journal of Applied Research in Memory and Cognition, 7(4)

ISSN

2211-3681

Authors

Tsai, Nancy
Buschkuehl, Martin
Kamarsu, Snigdha
et al.

Publication Date

2018-12-01

DOI

10.1016/j.jarmac.2018.06.001

Peer reviewed

Published in final edited form as:
J Appl Res Mem Cogn. 2018 December ; 7(4): 564–573. doi:10.1016/j.jarmac.2018.06.001.

(Un)Great Expectations: The Role of Placebo Effects in Cognitive Training

Nancy Tsai[1], **Martin Buschkuehl**[2], **Snigdha Kamarsu**[1], **Priti Shah**[3], **John Jonides**[3], **Susanne M. Jaeggi**[1]

[1]University of California, Irvine, CA

[2]MIND Research Institute, Irvine, CA

[3]University of Michigan, Ann Arbor, MI

Abstract

A growing body of literature demonstrating the malleability of critical higher-order cognitive functions by means of targeted interventions has incited widespread scientific interest, most notably in the form of cognitive training programs. The results are mixed and a point of contention: It has been argued that gains observed in cognitive training are mainly due to placebo effects. To address this, we examined the effect of participant expectations on one type of cognitive training that has been central to the controversy, namely *n*-back training, by inducing beliefs about expected outcomes. Participants receiving *n*-back training showed improvements in non-trained *n*-back performance regardless of expectations, and furthermore, expectations for positive outcomes did not result in any significant gains in an active control group. Thus, there was no detectable expectancy effect in either direction as a function of the cognitive intervention used, suggesting that training-related improvements are unlikely due solely to a placebo effect.

General Audience Summary

The brain training industry is an estimated billion-dollar industry but mounting concerns revolve around the genuine efficacy of training programs. Invested consumers have been left with little scientific guidance, however, and no study to date has tested whether positive training findings are indeed solely due to the placebo effect, as some have suggested. In the current study, we directly

---

Corresponding authors: Nancy Tsai and Susanne M. Jaeggi, School of Education, University of California, Irvine, 322 Education Building, Irvine, CA 92697, Phone: 949-824-5896, ntsai1@uci.edu and smjaeggi@uci.edu.
Author information:
Martin Buschkuehl, MIND Research Institute
Snigdha Kamarsu, School of Education, University of California, Irvine
Priti Shah, Department of Psychology, University of Michigan
John Jonides, Department of Psychology, University of Michigan

Author contributions
NT, SMJ, MB, PS, and JJ contributed to the study design. Testing and data collection were managed by NT and SK. NT performed the data analysis and interpretation under the supervision of SMJ. NT drafted the manuscript and SMJ, MB, PS, and JJ provided critical revisions. We thank Alexander Etz for his assistance with interpreting the Bayesian models. All authors approved the final version of the manuscript for submission.

Research Disclosure Statement
The total number of excluded observations and the reasons for making these exclusions has been reported in the Method section. All independent variables or manipulations have been reported in the Method section. All dependent variables or measures that were analyzed for this article's target research question have been reported in the Methods section.

address this issue and ask whether participant expectations for particular outcomes account for training outcomes. We provide new evidence that positive expectations are neither necessary nor sufficient in garnering favorable outcomes for those receiving a sterile training program. But for those in the experimental training program, we find that those who trained improve, even in the presence of negative expectations.

Keywords

working memory; brain training; cognitive plasticity; Hawthorne effect; expectancy

---

The idea that cognitive interventions can improve learning, slow the natural process of aging, or have positive life impact is the seductive promise underlying the expanding brain fitness industry. But the rise of commercial interests has sparked the attention of the scientific community, prompting critical evaluation of these claims and the formation of two camps of researchers that have reached contrary “consensus” regarding the effectiveness of brain training. While they differ in their interpretations of the literature, they unanimously call for the judicious interpretation of findings ("A Consensus on the Brain Training Industry from the Scientific Community," 2014; "Cognitive Training Data Response Letter," 2014). Of particular interest is the prospect of improving working memory (WM), a basic cognitive function associated with academic and professional outcomes (Miyake & Shah, 1999; Pickering, 2006). However, the results of WM training studies vary from reports of impressive transfer effects to non-generalized effects beyond the trained task itself, and empirical interpretation is further complicated by a number of conflicting meta-analyses (Au, Buschkuehl, Duncan, & Jaeggi, 2016; Au et al., 2015; Melby-Lervag & Hulme, 2016; Schwaighofer, Fischer, & Buhner, 2015; Soveri, Antfolk, Karlsson, Salo, & Laine, 2017; Weicker, Villringer, & Thone-Otto, 2015).

A key criticism is that positive findings might be driven by a positive expectancy or a placebo effect: positive outcomes of treatment attributed to the *belief* in treatment efficacy but not to the *genuine* efficacy of treatment (Ueberwasser, 1787). To control for such expectancy effects, cognitive training studies typically include a sterile intervention with characteristics that make it potentially believable as an intervention. The assumption is that such *active* controls mitigate the likelihood that placebo effects account for positive findings. However, not all studies employ active controls. The irregular use of active controls prompts suspicion that positive training outcomes across studies may be due to studies that use only passive controls in which participants receive no intervention and have no contact with the experimenters during the pretest-to-posttest period (Au et al, 2015; Redick, Shipstead, Wiemers, Melby-Lervag, & Hulme, 2015). Specifically, it has been argued that positive training outcomes might be due to placebo effects, assuming that active controls induce positive expectations whereas passive controls induce negative or no expectations. Researchers have since attempted to examine training outcome as a function of control groups, but the results vary across outcome measures and studies (Karbach & Verhaeghen, 2014) suggesting that it is unlikely that transfer effects depend on the type of control used (see also Au et al., 2016).

It is possible that even with active controls, placebo effects arise from differential improvement expectations for treatment and control groups. Some argue that in order to evaluate the *true* efficacy of a cognitive intervention, expectations across groups must be matched (Boot, Simons, Stothart, & Stutts, 2013). However, one meta-analysis demonstrated that more intense placebos (i.e., placebos with greater believability) were not consistently associated with larger effects compared to less intense placebos (Fässler, Meissner, Kleijnen, Hróbjartsson, & Linde, 2015), suggesting that participant expectancies might not be as influential as suspected.

Still, few have actually assessed the impact of expectancies. To our knowledge, only two studies to date have attempted to directly test for expectancy effects in relation to cognitive training outcomes. In a study by Boot and colleagues (2013), participants watched a video of either an action video game ("intervention") or a non-action game (active control) then learned about tasks used as outcome measures in a behavioral study before indicating their expectations for hypothetical improvement on these tasks. These viewers' expectations of improvement for the two games were compared to the outcomes of a different, training study using these same video games. Differential expectations for improvement correlated with cognitive performance outcomes in the training study; however, participant expectations were neither explicitly assessed in the actual behavioral study nor did the study include an actual intervention. In an application to WM training, Foroughi and colleagues (2016) examined the outcomes of two self-selected groups differing in their expectations of training: those recruited by "brain training & cognitive enhancement" fliers (placebo group) outperformed those recruited by generic fliers with no mention of brain-training (control group) on fluid intelligence measures following a single session of WM practice. Because their study used only a single, brief training session, it is not possible to determine whether or not placebo effects account for outcomes of actual training studies. Though it is quite possible that placebo effects dominate initial outcomes but not longer-term outcomes (see Katz, Jaeggi, Buschkuehl, Shah, & Jonides, 2018).

Thus, the question remains: do expectancy differences drive cognitive training outcomes? To address this issue, our study examined expectancy effects in both directions - positive and negative - for both training and control groups. Our hypotheses reflect the sparse and conflicting findings thus far: If expectations account for cognitive training outcomes (Boot et al., 2013; Foroughi et al., 2016), participants in the control positive-expectancy condition would outperform those in the training negative-expectancy condition. Alternatively, if training gains arise regardless of expectancy, participants in the trained groups would outperform the controls.

Method

Participants

A group of 138 participants from the University of California, Irvine took part in the present study. They volunteered to participate in our study advertised as a "Brain Training Study" and received course credit. Thirteen participants withdrew before completing the study, primarily due to time constraints. Most of them (n=7) never trained at all and only one participant dropped out with more than 40% of training completed. Data from three participants were not included due to procedural errors. The final sample of participants who completed the entire study consisted of 122 participants (mean age=21.0 years, SD=3.4; 85 women).

Design

We used a between-subjects pretest-posttest intervention design and assigned participants to one of four conditions: training with positive expectancy; training with negative expectancy; active control with positive expectancy; and active control with negative expectancy. For training, we chose a task that is one of the most commonly used tasks in cognitive intervention research, namely, the *n*-back task, which has also shown promise in terms of producing generalizing effects (Weicker et al., 2016; Au et al., 2015). As the active control, we used a task that required participants to learn vocabulary and general knowledge facts.

To induce differential expectations, participants were taught about “uni-modal” and “cross-modal transfer” at the beginning of the study. Specifically, participants were told that training would either improve performance in both trained and untrained domains (positive beliefs for cross-modal transfer) or in only the domain in which they trained (negative beliefs for cross-modal transfer, i.e. uni-modal transfer). Participants were randomly assigned to the four conditions (note that in order to ensure that group sizes turned out as similar as possible, the last few participants were assigned to complete the groups with the lowest sample sizes).

Although there is no previous research examining expectancy effects on *n*-back or other WM training outcomes, one study using the same training task and the same number of training sessions (Buschkuehl et al., 2014) included 21 to 28 participants per condition and achieved a medium effect size in the same cross-modal outcome measure. Assuming a similar effect size in the current work, we recruited 29 to 32 participants per condition in order to achieve a statistical power of .8. The protocol was approved by the local institutional review board and was carried out in accordance with the provisions of the Declaration of Helsinki.

Materials

Training tasks.

Visual n-back training task.—The training task was a computerized adaptive version of the visual *n*-back task used previously (Au, Katz, et al., 2016; Buschkuehl, Hernandez-Garcia, Jaeggi, Bernard, & Jonides, 2014). During training, a series of blue squares was displayed on a computer screen, each in one of eight spatial locations (presentation time: 500 ms, interstimulus interval: 2,500 ms). Participants were required to indicate whether or not the current stimulus appeared at the same location as the one presented *n* positions back in the sequence. Using a standard computer keyboard, participants responded with the letter “a” for targets and “l” for non-targets. Participants completed 15 rounds in each training session, each round consisting of 20+*n* trials of which 6 were targets and 14+*n* were non-targets. The training was adaptive so that after each round, the level of n was varied as a function of performance: If participants made fewer than three errors, the level of *n* increased in the next round by one, but decreased by one if they made more than five errors; in all other cases, *n* remained the same.

Knowledge training task.—We used a knowledge training task similar to the one described in prior *n*-back training studies (Jaeggi, Buschkuehl, Shah, & Jonides, 2014). Participants solved GRE-type general knowledge, vocabulary questions, and trivia questions selected from a pool of approximately 5,000 questions. With every presentation, participants chose one of four answer alternatives presented below the question followed by the correct answer sometimes supplemented with additional facts related to the question. Like the experimental task, the knowledge training task was adaptive in that the difficulty of the questions was varied as a function of participants' performance, and incorrectly answered questions were presented again in the next training session.

For both training tasks, each training session lasted approximately 25 minutes.

Outcome measures.

Visual n-back task.—Participants were tested on a visual *n*-back task and an auditory *n*-back task before and after the intervention period. The visual *n*-back task was identical to the *n*-back training task but was not adaptive. The stimulus material was the same as for the *n*-back training described above, however, the *n*-back levels were fixed in that participants completed 2-, 3-, and 4-back tasks in that order, 3 rounds per level, and each round contained 20+*n* trials (6 targets).

Auditory n-back task.—The auditory *n*-back task required participants to process a continuous stream of spoken letters presented through headphones (Au, Katz, et al., 2016; Buschkuehl et al., 2014). Difficulty varied sequentially from 2- through 4-back for both tasks, with three blocks at each level, consisting of 20+*n* trials each (6 targets).

The primary dependent variable of interest for both *n*-back tasks was the hit rate minus false alarm rate averaged across the three difficulty levels (P_R; (Jaeggi et al., 2010; Snodgrass & Corwin, 1988)).

Self-Report measures

Baseline Assessments (Pre-Test).—The following questionnaires were administered at pretest, following study overview but prior to the administration of the cognitive tests. A subset of questions from the *Intrinsic Motivation Inventory* (McAuley, Duncan, & Tammen, 1989) was used to assess how valuable participants would hypothesize the study program to be. Seven statements such as "*I think that doing this activity is useful* were presented and participants were asked to indicate their level of agreement or disagreement on a 7-point Likert scale. We used the *Need for Cognition* (Cacioppo & Petty, 1982b) questionnaire to assess how much participants enjoy cognitively challenging tasks. The questionnaire consists of 18 statements such as "*I really enjoy a task that involves coming up with new solutions to problems*" and participants were asked to indicate their level of agreement or disagreement on a 9-point Likert scale. *Theories of Cognitive Abilities* (Dweck, 2000) was used to assess the degree to which participants think of intelligence as malleable or fixed.

The questionnaire consists of eight statements, such as "*You have a certain amount of cognitive ability and you can't really do much to change it*," and participants indicate their agreement or disagreement on a 6-point Likert scale.

Assessments During Training.—The following questions were administered after each training session. The question *"Do you think this training is improving your memory in any way?"* was used to assess expectancy during training and answered on a 5-point Likert scale (1: not at all, 5: very much). *"I tried very hard on this activity"* was used to assess their effort during training and answered on a 7-point Likert scale (1: not true at all, 7: very true). And *"I am satisfied with my performance on this task"* was used to assess participants' satisfaction with their performance on the task and answered on a 7-point Likert scale (1: not at all, 7: very true).

Post-Test Assessments.—The following questions were administered after training and post-test completion. The questions, "Do you think that this training has improved your memory in any way? Do you think that this training has improved your ability to concentrate or pay attention in your daily life? Do you think that this training has improved any other (mental) abilities?" were used to assess perceived changes due to training (averaged as "Expected Outcomes") and participants were asked to indicate their level of change on a 5-point Likert scale (1: not at all; 5: very much). The question, "Do you think this training has helped you perform better at this task?" (0: no; 1: yes) was used to assess their expectancy after the posttest, and "I put a lot of effort into this" was used to assess perceived value of the n-back tasks completed during post-test, and participants provided their responses on a 7-point Likert scale (1: not true at all; 7: very true). Self-report measures by group are presented in Table 2.

Procedure

At the onset of the study, participants read the study information sheet, then completed a baseline assessment session as part of the pretest consisting of questionnaires and the visual and auditory *n*-back task. Participants were assigned to *n*-back training or control training with either positive or negative expectations for cross-modal transfer (see Supplementary Material; Figure S1). All participants followed similar procedures: Pretest cognitive tests, expectancy induction, seven days of cognitive training (visual *n*-back training or active control) lasting approximately 25 minutes per session, expectancy re-induction on training days three and six, posttest cognitive tests, with questionnaires occurring throughout (see Figure 1).

Immediately following the pretest, participants were presented a recorded, narrated presentation designed to strategically induce expectancy beliefs. The presentation reviewed a fictitious study published in the "Journal of Neuroscience by a team of neuroscientists in Switzerland" by carefully reviewing fictitious findings, providing neuroscience-based evidence supporting these specific findings to make it more believable (Rhodes, Rodriguez, & Shah, 2014), summarizing the results, and explicitly presenting the hypothesis of the fictitious study as applied to the student population at UC Irvine (see Figure S2). Each condition received a specific recorded presentation that differed on two dimensions: 1) the content was either about WM (treatment) or memory (active control); 2) the fictitious study reviewed demonstrated either cross-modal or uni-modal findings. Cross-modality was defined as the transfer of skills between seemingly different tasks, such as a visual and auditory task or an *n*-back and vocabulary task, so that improvement on one cognitive task (e.g., auditory) might translate to gains in another cognitive task (e.g., visual). Participants were presented results that indicated either activation of identical brain regions (cross-modal transfer) or different brain regions (uni-modal transfer) during the auditory and visual task (see Figure S1). The expectancy induction was followed by a multiple-choice test to ensure comprehension. Participants were given up to 90 minutes to complete the baseline assessments and the induction presentation on their first day of the study, though most participants finished within an hour.

Upon completion of all baseline assessments and the expectancy induction, cognitive training commenced on the second day of the study. Participants were given careful instructions pertaining to the training task to which they were assigned and given practice trials to ensure that they understood how to complete the task. Following the instructions and practice, participants commenced seven days of training with either an adaptive visual *n*-back task or an adaptive general knowledge/vocabulary task. At the onset of training sessions 3 and 6, induction material pertaining to the results of the fictitious study was reviewed followed by a multiple-choice comprehension test (see Figure S3). The re-induction aimed to ensure that the participant’s expectations were maintained throughout the study. At the end of every training session, participants reported their exerted effort (“I tried very hard on this activity”) and how much they believed the training was improving their memory (“Do you think this training is improving your memory in any way?”).

Finally, after the seven days of training, participants completed the posttest assessments, which included the visual and auditory *n*-back task and questionnaires presented at pretest.Along with the questionnaires presented at pretest, we also included the question, “Do you think that this training has helped you perform better at this task? – (YES/NO),” in addition to other questions gauging improved cognitive function (e.g., memory, attention, or general mental abilities) following training. Lastly, participants were debriefed about the study. This session lasted approximately 30-45 minutes.

Results

Group related differences were analyzed with an analysis of variance (ANOVA) (Table 2). In addition, we calculated Bayes Factors in JASP using default priors (JASP Team, 2017) to report the probability of H1 relative to H0 (note that values larger than 1 are in favor of H1 and values smaller than 1 are in favor of H0). We first compared the four groups to determine whether there were any group differences at pretest. The four groups did not differ by age ($F(3, 118)= .63$, $p>.250$, $\eta_p^2=.15$, $BF_{10}=.09$) or by gender composition (X^2 (3, $N=122)=3.45$, $p>.250$, $BF_{10}=0.16$). Furthermore, the groups did not differ on visual and auditory *n*-back performance ($F(3, 117)= 1.99$, $p=0.119$, $\eta_p^2=.04$, $BF_{10}=0.42$ and $F(3, 117)=1.29$, $p>.250$, $\eta_p^2 =.03$, $BF_{10}=0.19$, respectively), in their self-reported Intrinsic Motivation Inventory ($F(3,113)=1.35$, $p>.250$, $\eta_p^2=03$, $BF_{10}=0.24$), Need for Cognition ($F(3,118)=1.71$, $p=.167$, $\eta_p^2=.04$, $BF_{10}=0.30$), or Theories of Cognitive Abilities ($F(3,118)=$.

48, p>.250, η_p^2=.01, BF_{10}=0.07). Finally, the groups did not differ on their perceived value of the training prior to training (F(3, 113)=1.35, p>.250, η_p^2=.03, BF_{10}=0.21). The descriptive data are reported in Table 1 and Table 2.

To examine performance on the trained visual n-back task, a two-way ANCOVA with pretest performance as the covariate, and treatment (n-back training vs. control) and expectancy (positive vs. negative) as between-subject variables revealed a highly significant main effect of treatment (F(1, 116)= 52.25; p<.001, η_p^2=.31, $BF_{inclusion}$=7.618e^7), indicating that the n-back training groups (collapsed across positive and negative expectancy; M=.74, SD=.23 and M=.81, SD=.20, respectively) outperformed the control groups (M=.56, SD=. 19 and M=.63, SD=.23, respectively), suggesting that the improvement at post-test was due to n-back training irrespective of any expectancy effects (Figure 2). Specifically, there was no significant main effect of expectancy (F(1, 116)= 1.42, p=.236, η_p^2=.012, $BF_{inclusion}$=0.34) nor was there a significant treatment by expectancy interaction (F(1,116)= .02, p>.250, η_p^2<.01, $BF_{inclusion}$=0.30)[1].

Next, we examined the performance on the untrained auditory n-back task. A two-way ANCOVA with auditory n-back pretest performance as the covariate, and treatment (n-back training vs. control) and expectancy (positive vs. negative) as between-subject variables revealed a highly significant main effect of treatment in favor of the n-back training groups (F(1,117)= 18.33; p<.001, η_p^2=.14, $BF_{inclusion}$=345.80) showing that those in the n-back training groups (collapsed across positive and negative expectancy; M=.63, SD=.21 and M=.66, SD=. 19, respectively) outperformed the control groups (M=. 58, SD=. 17 and M=. 56, SD=. 19, respectively). Furthermore, there was no significant main effect of expectancy (F(1,117)=1.22, p>.25, η_p^2=.01, $BF_{inclusion}$=0.28) nor was there a significant treatment by expectancy interaction (F(1,117)= 0.38, p>.25, η_p^2<.01, $BF_{inclusion}$=0.29)[2]. To specifically compare the performance of the control group who expected to improve in the untrained domain (control positive) with the performance of the training group who did not expect to improve (training negative) (highlighted in Figure 3), we conducted an additional ANCOVA which revealed that the n-back training negative expectancy group still outperformed the control positive group (F(1, 57)=5,94, p=.02, η_p^2=.09, $BF_{inclusion}$=4.073), indicating that the treatment group improved on the auditory n-back despite expecting no improvement after having trained on the visual n-back. Together, these results further corroborate our visual n-back finding that it is the effect of treatment rather than expectations that drive n-back outcomes (Figure 3).

---

[1] Note: we also conducted a 2 (session; pre vs. post) × 2 (treatment; experimental vs. control) × 2 (expectancy; positive vs. negative) repeated measures ANOVA that resulted in similar patterns, namely, we observed a main effect of session (F(1, 117)=94.88, p<.001, η_p^2=.45, BF_{10}=3.87e^{10}) but neither a main effect of treatment (F(1, 117)=3.03, p=.08, η_p^2=.02, BF_{10}=.78) nor a main effect of expectancy (F(1, 117)=2.98, p=.08, η_p^2=.02, BF_{10}=.82). However, there was a highly significant session × treatment interaction (F(1, 117)=58.44, p<.001, η_p^2=.33, BF_{10}=1.16e^9), but there was neither a session × expectancy (F(1, 117)=.72, p>.25, η_p^2<.01, BF_{10}=.20) nor a session × treatment × expectancy interaction (F(1, 117)=0.05, p>.25, η_p^2<.01, BF_{10}=.26).

[2] Note: A 2 (session; pre vs. post) × 2 (treatment; experimental vs. control) × 2 (expectancy; positive vs. negative) repeated measures ANOVA resulted in similar patterns, that is, a main effect of session (F(1, 118)=21.78, p<.001, η_p^2=.15, BF_{10}=839.48) but neither a main effect of treatment (F(1, 118)=0.18, p>.25, η_p^2<.01, BF_{10}=.29) nor a main effect of expectancy (F(1, 118)=0.72, p>.25, η_p^2<.01, BF_{10}=.37). In addition, there was a highly significant session × treatment interaction (F(1, 118)=20.55, p<.001, η_p^2=.15, BF_{10}=1558.81), but there was neither a session × expectancy (F(1, 118)=2.10, p=.15, η_p^2=.15, BF_{10}=.50) nor a session × treatment × expectancy interaction (F(1, 118)=0.26, p>.25, η_p^2=.002, BF_{10}=.28).

During the training period, there were no reported group differences in exerted effort ("I tried very hard on this activity") across the seven sessions ($F(3,118)= 1.72$, $p=.167$, $\eta_p^2=.04$, $BF_{10}=0.30$). Furthermore, when asked whether participants were satisfied with their performance on the task immediately following the task, and whether they thought the training was improving their memory in any way, participants in all four groups answered similarly ($F(3, 118)= 1.74$, $p=.163$, $\eta_p^2=.04$, $BF_{10}=0.31$ and $F(3,118)=1.16$, $p>.250$, $\eta_p^2=.02$, $BF_{10}=0.16$, respectively) (see Table 2).

Immediately following the posttest, when asked whether they thought "this training improved their memory in any way" again there were no group differences ($F(3,115)= 0.76$, $p>. 250$, $\eta_p^2=.02$, $BF_{10}=0.10$). However, when asked more specifically whether participants thought "this training has helped you perform better at *this* task" a one-way ANOVA revealed group differences ($F(3,117)=5.54$, $p=.001$, $\eta_p^2=.12$, $BF_{10}=22.55$). A post-hoc Tukey test revealed that treatment positive and treatment negative groups (M=.87, SD=.33 and M=.96, SD=. 18, respectively) were both more likely to report "yes" to the question compared to the control positive group (M=.58, SD=.50), differing significantly at $p<.05$ ($BF_{10} = 2.83$). Additionally, while there were no reported differences in effort throughout the study, a difference in effort ("I put a lot of effort into this") was reported at posttest ($F(3,107)=3.14$, $p=.028$, $\eta_p^2=.08$, $BF_{10}=1.62$), mirroring the differences in the response to the aforementioned question ("Do you think this training has helped you perform better at *this* task?) (see Table 2).

In sum, our data suggest that expectations as manipulated in this study do not noticeably account for *n*-back training outcomes. By inducing differential expectations for transfer, we were able to disentangle the impacts of expectations on a near-transfer measure. Our data indicate that the transfer gains were exclusive to those who trained on an *n*-back task but were notably absent in those who trained on a non-working-memory task, in spite of expectancies. To corroborate our frequentist-derived interpretation, we used Bayesian methods to facilitate statistical interpretation by comparing relative probabilities of two competing hypotheses, the null and the alternative. In fact, the Bayes Factor for treatment effects (*n*-back training) across groups was $BF_{10}>100$, suggesting these data are >100 more likely to be observed under the alternative hypothesis relative to the null hypothesis; in other words, there is decisive evidence for H_1 (effect of treatment). On the other hand, the Bayes Factor for expectancy across groups was $BF_{10}<1/3$, suggesting these data are 3 times more likely to be observed under the null hypothesis relative to the alternative hypothesis; that is, there is moderate to substantial evidence for H_0 (null effect of expectancy) (Kass & Raftery, 1995). Simply put, we found a positive effect of training despite the presence of negative expectation, and furthermore, we did not observe any significant improvement in the control group, despite having positive expectations. These findings challenge the notion that beliefs alone account for the positive gains reported in the cognitive training literature.

Discussion

As the first study to test the interaction of expectancy and cognitive training, we employed seven days of *n*-back training to test how expectancies influence training outcome, a direct contrast to the single practice session used by Foroughi, Monfort, Paczynski, McKnight, and

Greenwood (2016). Our findings demonstrate that *n*-back training improved outcomes on both the trained (visual) and untrained (auditory) *n*-back tasks, even amongst participants who held negative expectancies for the untrained task. Critically, active controls showed no reliable gains on either task, despite having positive expectations. These results indicate that improvements following *n*-back training are not simply driven by positive expectancies.

It is important to consider alternative explanations—Studies examining individual differences highlight a potential force underlying intervention outcomes: individuals who hold implicit beliefs regarding the malleability of intelligence (Dweck, 2000) or vary in their need for cognitive challenge (Cacioppo & Petty, 1982) may differ in their engagement during cognitive training (Jaeggi et al., 2014). Accounting for these possibilities, we assessed participants' mindsets at baseline, but did not observe any group differences, and furthermore, participants across conditions reported similar levels of interest for the task at the study onset. There were also no group differences in effort or beliefs regarding the training efficacy throughout the training days; thus, any performance differences are unlikely driven by those factors.

However, immediately following the posttest, individuals in the control group were less likely to report that training improved their posttest performance and reported exerting less effort during posttest. But these differences occurred only *after* the posttest, a stark divergence from reporting similar levels of effort up to that point, which may reflect self-preservation and recognition of unexpected poor performance within the control group, despite prior expectations.

While our results suggest that improvements in *n*-back performance are likely due to the treatment itself, future work is needed to clarify the relationship between cognitive training and expectancies. For example, a double-blind procedure may have offered a more rigorous test of treatment effects but was an unavailable option in the current study. Second, it is important to consider the possibility that our expectancy manipulation was insufficient to alter fundamental beliefs about training outcome. To account for this, participants were given comprehension tests to ensure understanding of cross-modality and expectations were covertly assessed through questionnaires, none of which revealed group differences during and after training. Still, a more targeted assessment of beliefs regarding specific outcomes may be of value, although such questions risk revealing the nature of the study and undermine attempts at maintaining expectancy. Third, in addition to including an active control condition, a no-contact control group could shed light on the different degrees of expectancy participants might hold during a research study. Lastly, our primary interest was to assess the specific effects of expectancy. Therefore, we strategically focused on expectancy effects for the most proximal outcome measures and refrained from including far transfer measures, though we acknowledge that the current discussions in the training literature focus on the issue of far transfer (Simons et al., 2016). Overall, the inclusion of a wider range of outcome measures could uncover whether and why expectancy might play a differential role depending on the type of outcome measure. Nonetheless, it seems unlikely that expectancy effects would emerge in far transfer measures given the absence of any discernible effects in near transfer measures as illustrated by meta-analytic work reporting the diminishing effect size as a function of far vs. near transfer (Soveri et al., 2017).

Our data cannot speak to whether our observed changes in *n*-back performance reflect a genuine change in WM-related skills or merely increased *n*-back familiarity, but this was not the aim of the study. Relatedly, although one might conclude that the observed improvements in a non-trained *n*-back variant are trivial, it is important to note that any reliable transfer beyond the specific trained task variant has been difficult to obtain, as prominently discussed in the perceptual learning literature (e.g., Seitz, 2017).

Importantly, our findings directly address the issue of expectancy, a concern that has plagued the cognitive training discourse. While our study was not intended to settle the debate surrounding the efficacy of training or transfer, the finding that positive expectations alone do not drive positive training outcomes could reframe the standing issue that has often centered around the implementation of active and passive controls. Specifically, the debate around control conditions is predicated on the *assumption* that active controls entail positive expectancy whereas passive controls entail negative (or no) expectancy, but we propose that researchers exercise caution in accepting this assumption. While the literature has shown negligible differences between active and passive controls (Clemenson & Stark, 2015; Bürki, Ludwig, Chicherio, & de Ribaupierre, 2014; Chooi & Thompson, 2012; Redick et al., 2013; Thompson et al., 2013), expectancies may also vary by individual beliefs rather than by assigned condition (see Katz et al., 2018). Therefore, we urge considering other factors related to expectancy that may mediate outcomes (e.g., incremental mindsets, need for cognition, engagement). Though prior work reported training effects over and above such dimensions (Jaeggi et al., 2014), such differences beyond the variable of interest may still confound the relationship between expectancy and outcomes. As such, we emphasize the importance of assessing participants' expectancies before, during, and after the intervention along with other beliefs, irrespective of the type of condition or control employed. In an effort to advance this area of inquiry, we suggest that researchers continue to ask: what kind of interventions and what features of those interventions might work and why? For whom does training work and under what conditions? In regard to the latter question, it is worth pointing out that a participant's motivation to engage in an intervention may depend on the expectation for outcomes. However, even though such a factor can produce beneficial effects independent of the intervention effect (Motter et al., 2016), our data does not confirm this to be a general phenomenon.

Conclusion

The present findings challenge the notion that placebo effects are the only contributors for the gains in cognitive training, but we also argue that a failure to consider the role of expectancy on outcomes would be negligent. With any intervention, there is a rich set of biological, psychological, and social factors that make training programs effective. Recognizing that one of those factors can enhance the desired outcome does not render the fundamental component(s) ineffective. If positive beliefs towards training lead to optimal results, is that undesirable? For example, the belief that running improves cardiovascular health might enhance and sustain practice, which in turn may lead to improved cardiovascular health. It would not be argued that a positive regard for running while not actually engaging in the sport is sufficient in improving cardiovascular fitness. Analogously, our findings indicate that expectancies are neither sufficient nor necessary to explain positive cognitive training outcomes, but whether, how, and to what degree expectancies influence factors that *do* facilitate treatment efficacy is still largely unexplored. Therefore, we propose that future training work should control for and assess participants' expectancies and engagement during the continued examination of mechanisms involved in cognitive training.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding and Conflict of Interest Statement

SMJ is supported by NIH/NIA grant no. 1K02AG054665. MB is employed at the MIND Research Institute, whose interest is related to this work and SMJ has an indirect financial interest in the MIND Research Institute. No other authors declare any conflicts of interests.

References

References

Au J, Buschkuehl M, Duncan GJ, & Jaeggi SM (2016). There is no convincing evidence that working memory is NOT effective: A reply to Melby-Lervag and Hulme (2015). . Psychological Bulletin & Review, 23(1), 331–337.

Au J, Katz B, Buschkuehl M, Bunarjo K, Senger T, Zabel C, . . . Jonides J. (2016). Enhancing working memory training with transcranial direct current stimulation. Journal of cognitive neuroscience, 28(9), 1419–1432. doi:10.1162/jocn_a_00979 [PubMed: 27167403]

Au J, Sheehan E, Tsai N, Duncan G, Buschkuehl M, & Jaeggi SM (2015). Improving fluid intelligence with training on working memory: a meta-analysis. Psychological Bulletin & Review, 22(2), 366–377.

Boot WR, Simons DJ, Stothart C, & Stutts C (2013). The Pervasive Problem With Placebos in Psychology: Why Active Control Groups Are Not Sufficient To Rule Out Placebo Effects. Perspectives on Psychological Science, 8(4), 445–454. [PubMed: 26173122]

Bürki CN, Ludwig C, Chicherio C, & de Ribaupierre A (2014). Individual differences in cognitive plasticity: an investigation of training curves in younger and older adults. Psychological Research, 78(6), 821–835. [PubMed: 24652343]

Buschkuehl M, Hernandez-Garcia L, Jaeggi SM, Bernard JA, & Jonides J (2014). Neural effects of short-term training on working memory. Cognitive, Affective, & Behavioral Neuroscience, 14(1), 147–160.

Cacioppo JT, & Petty RE (1982a). The need for cognition. Journal of Personality and Social Psychology, 42(1), 147–160.

Cacioppo JT, & Petty RE (1982b). The need for cognition. Journal of Personality and Social Psychology, 42(1), 116.

Chooi WT, & Thompson LA (2012). Working memory training does not improve intelligence in healthy young adults. Intelligence, 40(6), 531–542.

Clemenson GD, & Stark CE (2015). Virtual environmental enrichment through video games improves hippocampal-associated memory. Journal of Neuroscience, 35(49), 16116–16125. [PubMed: 26658864]

Cognitive Training Data Response Letter. (2014, 4 17, 2017). Retrieved from http://www.cognitivetrainingdata.org/the-controversy-does-brain-training-work/response-letter/

A Consensus on the Brain Training Industry from the Scientific Community. (2014). Retrieved from http://longevity3.stanford.edu/blog/2014/10/15/the-consensus-on-the-brain-training-industry-from-the-scientific-community-2/

Dweck CS (2000). Self-theories: Their role in motivation, personality, and development.: Psychology Press.

Fässler M, Meissner K, Kleijnen J, Hróbjartsson A, & Linde K (2015). A systematic review found no consistent difference in effect before more and less intensive placebo interventions. . Journal of clinical epidemiology, 68(4), 442–451. [PubMed: 25639981]

Foroughi CK, Monfort SS, Paczynski M, McKnight PE, & Greenwood PM (2016). Placebo effects in cognitive training. Proceedings of the National Academy of Sciences.

Jaeggi SM, Buschkuehl M, Shah P, & Jonides J (2014). The role of individual differences in cognitive training and transfer. . Memory & Cognition, 42(3), 464–480. [PubMed: 24081919]

Jaeggi SM, Studer-Luethi B, Buschkuehl M, Su YF, Jonides J, & Perrig W (2010). The relationship between n-back performance and matrix reasoning--implications for training and transfer. Intelligence, 38(6), 625–635.

Karbach J, & Verhaeghen P (2014). Making working memory work a meta-analysis of executive-control and working memory training in older adults. Psychological Science, 25(11), 2027–2037. [PubMed: 25298292]

Katz B, Jaeggi SM, Buschkuehl M, Shah P, & Jonides J (2018). The effect of monetary compensation on cognitive training outcomes. Learning & Motivation, 63, 77–90.

McAuley E, Duncan T, & Tammen V (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. Research quarterly for exercise and sport, 60(1), 48–58. [PubMed: 2489825]

Melby-Lervag M, & Hulme C (2016). There is no convincing evidence that working memory training is effective: A reply to Au et al.(2014) and Karbach and Verhaeghen(2014). . Psychological Bulletin & Review, 23(1), 324–330.

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

Redick TS, Shipstead Z, Harrison TL, Hicks KL, Fried DE, Hambrick DZ, . . . Engle RW (2013). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. . Journal of Experimental Psychology: General, 142(2), 359. [PubMed: 22708717]

Redick TS, Shipstead Z, Wiemers US, Melby-Lervag M, & Hulme C (2015). What's working in working memory training? An edcational perspective. Educational Psychology Review, 27(4), 617–633. [PubMed: 26640352]

Rhodes RE, Rodriguez F, & Shah P (2014). Explaining the alluring influence of neuroscience information on scientific reasoning. Journal of Experimental psychology: Learning, Memory, and Cognition, 40(5), 1432.

Schwaighofer M, Fischer F, & Buhner M (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. Educational Psychologist, 50(2), 138–166.

Seitz AR (2017). Generalizable Learning: Practice Makes Perfect--But at what? Current Biology, 27, 225–227.

Simons DJ, Boot WR, Charness N, Gathercole SE, Chabris CF, Hambrick DZ, & Stine-Morrow EAL (2016). Do "Brain-Training" Programs Work? Psychological Science, 17(3), 103–186.

Snodgrass JG, & Corwin J (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. Journal of Experimental Psychology: General, 117(1), 34. [PubMed: 2966230]

Soveri A, Antfolk J, Karlsson L, Salo B, & Laine M (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. . Psychological Bulletin & Review, 1–20.

Thompson TW, Waskom ML, Garel KLA, Cardenas-Iniguez C, Reynolds GO, Winter R, . . . Gabrieli J (2013). Failure of working memory training to enhance cognition or intelligence. PloS one, 8(5), e63614. [PubMed: 23717453]

Ueberwasser F (1787). Anweisungen zum regelmabigen Studium der Empirischen Psychologie fur die Candidaten der Philosophie zu Munster: Friedrich Christian Theibing.

Weicker J, Villringer A, & Thone-Otto A (2015). Can Impaired Working Memory Functioning Be Improved By Training? A Meta-Analysis With a Special Focus on Brain Injured Patients. Neuropsychology, 30(2), 190–212.

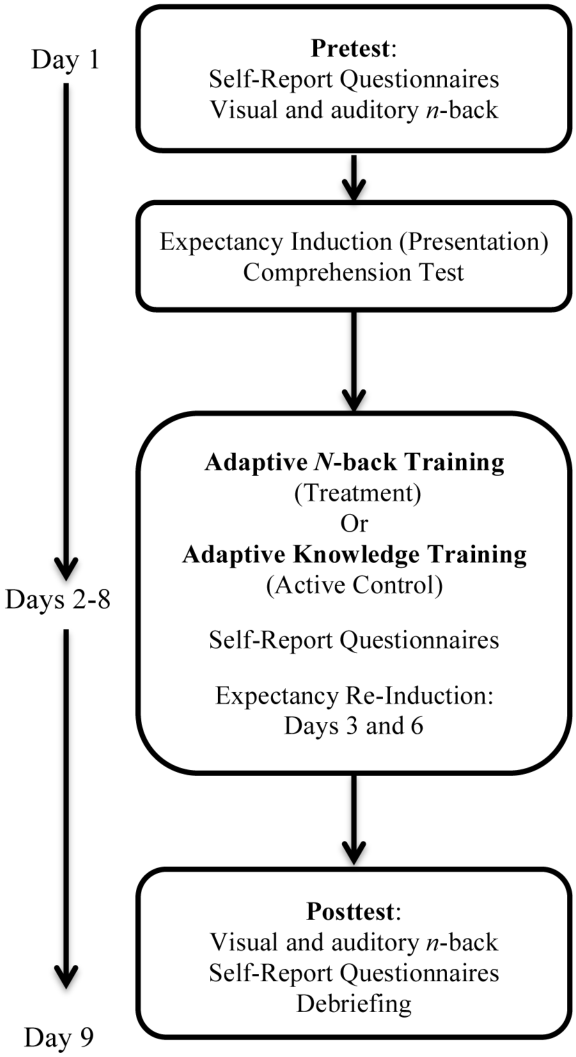

Figure 1.
Visualization of Study Procedures.

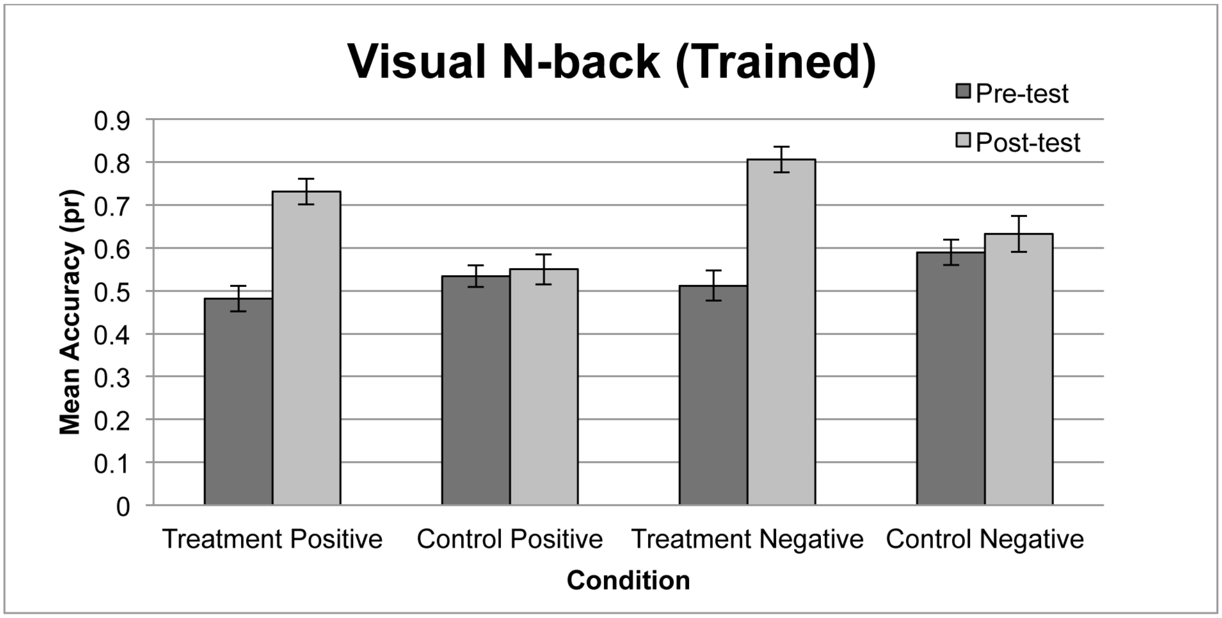

Figure 2.

Performance in the visual *n*-back task, tested before and after the intervention as a function of condition. The same task also served as the training vehicle. The overall *n*-back performance score is the average accuracy (pr; i.e., proportion of hits minus proportion of false alarms) across all three difficulty levels (2-, 3-, and 4-back). Error bars indicate standard errors of the mean.

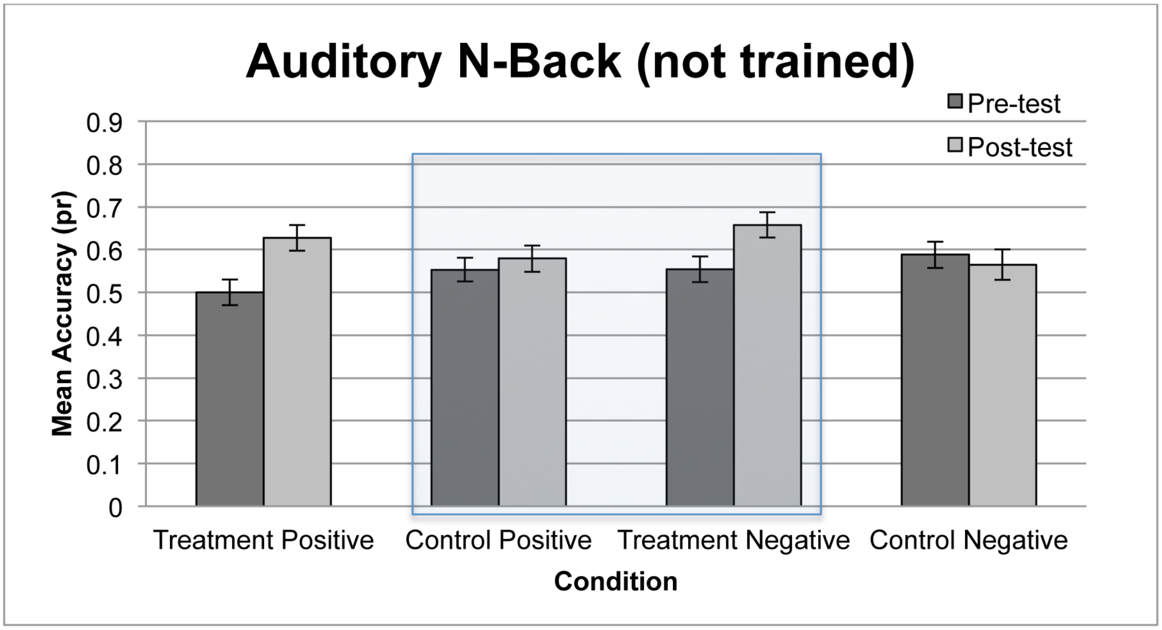

Figure 3.

Performance in the untrained auditory *n*-back task. The overall *n*-back performance score is the average accuracy (pr; i.e., proportion of hits minus proportion of false alarms) across all three difficulty levels (2-, 3-, and 4-back) as a function of condition. Error bars indicate standard errors of the mean.

Table 1.

Descriptive data for pre- and posttest outcome measures as a function of group.

			Pretest				Posttest				Pre vs. Post			
	N	No. women	Mean	*SD*	Min	Max	Mean	*SD*	Min	Max	*p*	*r*	ES	BF_{10}
Treatment Positive	32	22												
visual *n*-back			0.48	0.18	0.20	0.86	0.74	0.23	0.28	1.00	.00	.54	1.29	312304
auditory *n*-back			0.50	0.17	0.19	0.79	0.63	0.21	0.24	0.97	.01	.69	0.84	476.7
Treatment Negative	31	18												
visual *n*-back			0.51	0.22	0.05	0.93	0.81	0.20	0.23	1.00	.00	.60	1.58	$9.34e^{6}$
auditory *n*-back			0.55	0.21	0.12	0.93	0.66	0.19	0.17	0.93	.04	.83	0.93	795.1
Control Positive	29	23												
visual *n*-back			0.53	0.13	0.16	0.84	0.56	0.19	0.16	0.84	.54	.70	0.22	0.29
auditory *n*-back			0.55	0.15	0.16	0.84	0.58	0.17	0.30	0.96	.54	.62	0.21	0.30
Control Negative	30	22												
visual *n*-back			0.59	0.17	0.16	0.90	0.63	0.23	0.50	0.96	.41	.74	0.25	0.54
auditory *n*-back			0.59	0.17	0.16	0.99	0.56	0.19	0.80	0.87	.63	.73	0.22	0.28

r = test–retest reliability; ES = effect size that accounts for the correlation between the pre- and posttest measures, $\mu_2 - \mu_1 / \sqrt{(\sigma_1^2 + \sigma_2^2 - 2r_{12}\sigma_1\sigma_2)}$; BF = Bayes Factors

Table 2.

Self-Report data for pre-test, during training, and post-test as a function of group.

	Treatment Positive		Treatment Negative		Control Positive		Control Negative			
Pretest	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	***p***	**B_{10}**
Age	21.31	3.77	21.19	1.99	21.34	5.09	20.3	1.82	.59	.09
Intrinsic Motivation Inventory	5.60	0.99	5.50	1.20	5.30	1.05	5.09	1.16	.26	.24
Need for Cognition	0.21	0.89	0.44	0.98	0.72	0.86	0.31	0.97	.17	.30
Theories of Cognitive Abilities	4.23	1.06	4.53	1.10	4.46	1.12	4.52	1.16	.69	.07
After Each Training Session										
"*Do you think this training is improving your memory in any way*?" (1: not at all; 5: very much)	3.31	0.61	3.38	0.64	3.13	0.55	3.15	0.71	.33	.16
"*I tried very hard on this activity*" (1: not true at all; 7: very true)	5.35	1.03	5.74	1.05	5.04	1.37	5.43	1.34	.16	.30
"*I am satisfied with my performance on this task*" (1: not true at all; 7: very true)	4.11	0.86	4.24	1.33	3.58	1.28	3.79	1.43	.16	.31
Posttest										
Expected Outcomes (1: not at all; 5: very much)	3.21	0.77	3.32	0.70	3.02	0.88	3.23	0.72	.52	.10
"*Do you think this training has helped you perform better at this task*?" (No: 0; Yes: 1)	0.87	0.33	0.96	0.18	0.58	0.50	0.73	0.45	.001**	22.55
"*I put a lot of effort into this*" (1: not true at all; 7: very true)	5.28	1.37	5.45	1.14	4.34	1.58	5.13	1.52	.03*	1.62

Note. p-values and Bayes Factors (BF) reflect one-way analysis of variance between four groups. Expected Outcomes was composed of three questions: "Do you think that this training has improved your ability to concentrate or pay attention in your daily life?"; "Do you think that this training has improved any other (mental) abilities?"; "Do you think that this training has improved your memory in any way?"