Journal of Experimental Psychology: General
2013, Vol. 142, No. 2, 359–379

0096-3445/13/$12.00 DOI: 10.1037/a0029082

No Evidence of Intelligence Improvement After Working Memory Training: A Randomized, Placebo-Controlled Study

Thomas S. Redick, Zach Shipstead, Tyler L. Harrison, and Kenny L. Hicks
Georgia Institute of Technology

David E. Fried and David Z. Hambrick
Michigan State University

Michael J. Kane
University of North Carolina Greensboro

Randall W. Engle
Georgia Institute of Technology

Numerous recent studies seem to provide evidence for the general intellectual benefits of working memory training. In reviews of the training literature, Shipstead, Redick, and Engle (2010, 2012) argued that the field should treat recent results with a critical eye. Many published working memory training studies suffer from design limitations (no-contact control groups, single measures of cognitive constructs), mixed results (transfer of training gains to some tasks but not others, inconsistent transfer to the same tasks across studies), and lack of theoretical grounding (identifying the mechanisms responsible for observed transfer). The current study compared young adults who received 20 sessions of practice on an adaptive dual *n*-back program (working memory training group) or an adaptive visual search program (active placebo-control group) with a no-contact control group that received no practice. In addition, all subjects completed pretest, midtest, and posttest sessions comprising multiple measures of fluid intelligence, multitasking, working memory capacity, crystallized intelligence, and perceptual speed. Despite improvements on both the dual *n*-back and visual search tasks with practice, and despite a high level of statistical power, there was no positive transfer to any of the cognitive ability tests. We discuss these results in the context of previous working memory training research and address issues for future working memory training studies.

Keywords: training, working memory, attention, intelligence, multitasking

Supplemental materials: http://dx.doi.org/10.1037/a0029082.supp

The idea that a brief, inexpensive intervention can improve one's cognitive abilities is appealing and supported by some research investigations. Prominent examples of successful, scientifically validated interventions include reducing stereotype threat in African American students (Cohen, Garcia, Apfel, & Master, 2006) and treating neuropsychological impairments in psychiatric patients (Neuropsychological Educational Approach to Remediation; Medalia & Freilich, 2008). Although research on cognitive interventions is not new (e.g., Thorndike & Woodworth, 1901), the advent of inexpensive and portable computerized devices has made such programs easily accessible, as witnessed by a recent proliferation of commercial cognitive training programs (e.g., Brain Age, BrainTwister, Cogmed, JungleMemory, Lumosity, Mindsparke Brain Fitness Pro, Posit Science Brain Fitness, Posit Science InSight, WMPro). As a representative commercial example, Lumosity's website (http://www.lumosity.com/how-we-help) claims: "Based on extensive research, Lumosity improves memory, attention, processing speed, and problem-solving skills so you can feel more confident in your abilities."

What evidence is available that brief cognitive training programs actually lead to transfer, or positive gains, on nontrained fluid intelligence tests? In his landmark *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*, Carroll (1993) reviewed previous educational and research interventions to improve intelligence (pp. 669–674). Carroll's summary of the literature indicated very limited success in fundamentally and permanently changing one's general intellectual abilities. More recently, on the

This article was published Online First June 18, 2012.

Thomas S. Redick, Zach Shipstead, Tyler L. Harrison, and Kenny L. Hicks, School of Psychology, Georgia Institute of Technology; David E. Fried and David Z. Hambrick, Department of Psychology, Michigan State University; Michael J. Kane, Department of Psychology, University of North Carolina Greensboro; Randall W. Engle, School of Psychology, Georgia Institute of Technology.

The research was conducted with support by Office of Naval Research Grant N000140910129. We gratefully acknowledge Susanne Jaeggi for providing the materials for the adaptive dual *n*-back task, along with providing data and answers to several questions about previous studies. We thank Nash Unsworth and Gene Brewer for providing helpful comments on earlier versions of the article. We thank Forough Azimi, Jameil Bailey, Kati Frampton, Rakshya Khatri, Christine Kim, Princy Kuriakose, Dakota Lindsay, Shannon Lukey, Robyn Marshall, Kevin Strika, Maureen van der Water, Tamara Ware, and Allison Vercellone for their assistance in data collection. We thank Marlyssa Fillmore for assistance in preparing Table 2.

Correspondence concerning this article should be addressed to Thomas S. Redick, who is now at Division of Science, Indiana University Purdue University Columbus, 4601 Central Avenue, Columbus, IN 47203. E-mail: tsredick@iupuc.edu

basis of limited (but rapidly growing) research to date, some researchers (e.g., Klingberg, 2010; Perrig, Hollenstein, & Oelhafen, 2009; Sternberg, 2008) are optimistic about the efficacy of computerized working memory (WM) training in increasing fluid aspects of intelligence (i.e., those related to reasoning and problem solving). In contrast, other reviews (Conway & Getz, 2010; Morrison & Chein, 2011; Shipstead, Redick, & Engle, 2010, 2012) of the recent WM training literature have concluded that many of the training programs listed above have limited efficacy in improving intelligence and reasoning abilities. Shipstead et al. (2010) did note that the adaptive dual *n*-back training program (used in Brain Fitness Pro, BrainTwister, and Lumosity) held promise relative to other WM training programs. The commercial uses of the adaptive dual *n*-back task followed a report by Jaeggi, Buschkuehl, Jonides, and Perrig (2008) of an improvement in intelligence test scores in healthy, young adults, after dual *n*-back practice. This study has been widely cited both in the psychological literature (cited 142 times as of April 26, 2012, according to ISI Web of Science) and in the mainstream media (Highfield, 2008; Shellenbarger, 2011; Wang & Aamodt, 2009), due, in part, to the authors' conclusion that "the finding that cognitive training can improve [fluid intelligence] is a landmark result because this form of intelligence has been claimed to be largely immutable" (Jaeggi et al., 2008, p. 6832). Indeed, the results led Sternberg (2008) to proclaim that "fluid intelligence is trainable to a significant and meaningful degree" (p. 6791). Because of the potential importance of Jaeggi et al. and subsequent research by Jaeggi, Studer-Leuthi, et al. (2010), we will begin by critically evaluating the evidence suggesting that adaptive dual *n*-back practice improves intelligence.

After reviewing this work, we will present the results of a new study that sought to address limitations of previous WM training studies. Shipstead et al. (2010) noted two particular design problems prevalent in the WM training literature: (a) the use of no-contact control groups and (b) inadequate measurement of cognitive constructs by using single tasks. First, if the only comparison is between experimental (WM training) and control (no-contact) groups, there are a number of alternative explanations that can account for any observed differences on intelligence assessments after training (Campbell & Stanley, 1963). In addition, because no task is "process pure," in the sense that it captures only the construct of interest without measurement error, the use of a single task to represent an ability such as intelligence leaves open the possibility that nonintelligence components of test performance have been improved via training. The purpose of the current study was to address these and other issues in a comprehensive and systematic fashion in order to answer the following question: Does repeated practice on an adaptive dual *n*-back task transfer to, and actually cause, improvements in intelligence, multitasking, and WM capacity?

Adaptive Dual *n*-Back Training

Based on numerous studies indicating a strong relationship between WM capacity and higher order cognition (for review, see Kane, Conway, Hambrick, & Engle, 2007), the logic of many training programs is that increasing WM capacity should lead to improvements in tests measuring related constructs, including selective attention (Klingberg et al., 2005), inhibition (Thorell, Lindqvist, Bergman, Bohlin, & Klingberg, 2009), updating (Dahlin, Nyberg, Bäckman, & Neely, 2008), reading comprehension (Chein & Morrison, 2010), and fluid intelligence (Jaeggi et al., 2008). Note, however, that this logic is applicable only if the processes that are improved via WM training are the same processes shared between WM capacity and the targeted construct. For example, latent-variable studies indicate that WM capacity and fluid intelligence share approximately 50% of their variance (Kane, Hambrick, & Conway, 2005; Oberauer, Schulze, Wilhelm, & Süβ, 2005). Of course, this also means that 50% of the variance in each construct is not shared, which leaves substantial room for WM improvements that would not have any effect on fluid intelligence (and for fluid intelligence increases that do not have WM improvement as their cause).

However, there is evidence that the processes involved in successful dual *n*-back task performance do overlap with processes needed to solve reasoning and fluid intelligence test items. In the dual *n*-back task (Jaeggi et al., 2007), subjects respond to the identity of aurally presented letters and the location of visually presented squares, with letters and squares presented simultaneously. Subjects decide whether the current stimuli (letter and/or square) match the ones presented *n*-back, with *n* varying between 1 and 4 across experimental blocks for all subjects. Dual *n*-back accuracy correlated positively with measures of fluid intelligence test performance in three studies (Jaeggi, Buschkuehl, Perrig, & Meier, 2010, Study 3; Jaeggi, Studer-Leuthi, et al., 2010, Study 1; Redick, Shipstead, et al., 2012). Interestingly, dual *n*-back correlations with fluid intelligence tests were greater than with other WM tasks. In Jaeggi, Buschkuehl, et al. (2010, Studies 1 and 2), dual *n*-back and Reading Span correlations were not statistically different from 0. In Jaeggi, Studer-Leuthi, et al. (Study 1) and Redick, Shipstead, et al. (2012), dual *n*-back correlations with Operation Span and other complex span measures of WM were significant but smaller than the dual *n*-back correlations with fluid intelligence tests within the same subjects. Although it is unclear why dual *n*-back accuracy is weakly related to performance on other WM measures, of most importance here is that dual *n*-back accuracy is positively related to performance on fluid intelligence tests.

Jaeggi et al. (2008) published the first report concluding that adaptive dual *n*-back training improved intelligence. In the training version of the task, n changed as a function of performance across the experimental session (starting with $n = 1$). Subjects performed 20 blocks of $n + 20$ trials in each session, for approximately 30 min of daily practice. Dual *n*-back performance increased as a function of "dosage," or the number of sessions completed. Critically, Jaeggi et al. also reported that trained subjects exhibited significantly larger gains on an intelligence test compared to no-contact control subjects who did not perform the dual *n*-back between the pre- and posttest sessions (see Figure 1A).

Although the results presented in Figure 1A seem compelling, the figure represents data collapsed across four studies (see Figure 1B), in which different groups receiving either 8, 12, 17, or 19 sessions of dual *n*-back performance were compared to four separate groups of control subjects. This is not necessarily a problem, especially if the only difference among the studies (other than the subjects) was the number of dual *n*-back sessions completed. However, the four studies in Jaeggi et al. (2008) differed in other important ways, and these lead to numerous interpretative challenges of the combined Figure 1A:

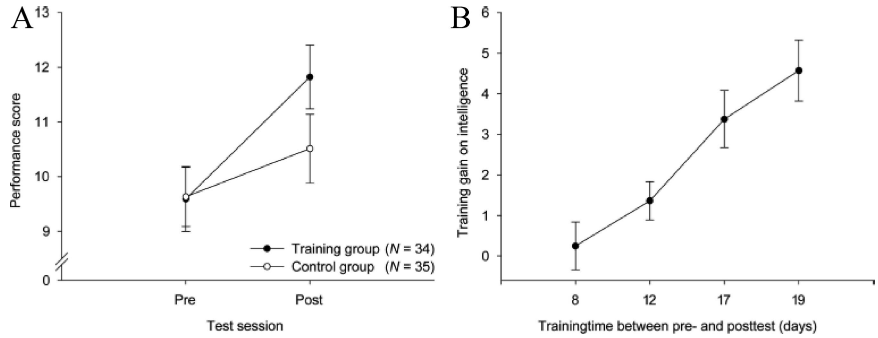

Figure 1. Matrix reasoning test performance as a function of group and session (A) and Raven Advanced Progressive Matrices or Bochumer Matrizentest gain as a function of the number of dual *n*-back sessions completed (B). From "Improving Fluid Intelligence With Training on Working Memory," by S. M. Jaeggi, M. Buschkuehl, J. Jonides, & W. J. Perrig, 2008, *Proceedings of the National Academy of Sciences of the United States of America, 105,* p. 6831. Copyright 2008 by the National Academy of Sciences of the United States of America.

1. Data were collapsed across different transfer tests administered under different time limits.

To assess transfer to intelligence, the eight-session groups completed the Raven Advanced Progressive Matrices (RAPM), and the 12-, 17-, and 19-session groups performed the Bochumer Matrizentest (BOMAT). Although both tests assess matrix reasoning (presenting 3×3 vs. 3×5 matrices, respectively), they are not comparable in length (18 vs. 29 items, respectively). In addition, the 19-session groups were given 20 min to complete BOMAT, whereas the 12- and 17-session groups received only 10 min (S. M. Jaeggi, personal communication, May 25, 2011). As shown in Figure 2, the use of the short time limit in the 12- and 17-session studies produced substantially lower scores than the 19-session study. We argue that it is inappropriate to simply average across the number of problems solved from the four tests to create Figure 1A.

2. There were procedural differences across the four studies.

Although the dual *n*-back groups differed systematically in the number of practice sessions performed between pre- and posttest, other procedural changes justify keeping the four studies separate. First, the eight-session study also included an active-control group that completed simple and choice reaction time tasks (Jaeggi, 2005).[1] Second, the 17-session study also included electroencephalograph recordings during performance of the dual *n*-back tasks in the pre- and posttest sessions and an extra nonpractice session between pre- and posttest for both groups. Finally, in addition to either RAPM or BOMAT, Jaeggi et al. (2008) reported that Reading Span (no transfer) and Digit Span (positive transfer), were administered during the pre- and posttest sessions. Subjects in the 19-session study also performed additional transfer tasks during multiday pre- and posttest sessions, exhibiting positive transfer to Stroop and delayed free-recall tasks; negative transfer to digit–symbol substitution test; and no transfer to visuospatial span, task-switching, immediate free-recall, and semantic priming tasks (Jaeggi, 2005). Note also that in the 19-session study, positive transfer was observed for Reading Span. The procedural variations mentioned here serve as additional reasons not to collapse the intelligence transfer results across the individual studies.

3. Patterns of transfer differed across the four studies.

As noted in Jaeggi et al. (2008), the analysis of covariance (ANCOVA) results, with posttest score as the dependent variable and pretest score as the covariate, were not significant for the eight- and 12-session dual *n*-back training groups. In fact, the 17-session study is the only one that is visually similar to Figure 1 collapsed across the four individual studies, with matched intelligence scores between control and training groups before training and substantial differences in intelligence scores between groups after training (see Figure 2). Although Jaeggi et al. interpreted the results across the four studies as consistent with a dose-dependent relationship (see Figure 1B), it is also correct to state that whereas two of the studies found evidence for dual *n*-back transfer to matrix reasoning, two of the studies did not.

The individual studies of Jaeggi et al. (2008) are also based on very small samples (e.g., $n = 7$, 8, or 11 in each group of the four studies). In a follow-up study, Jaeggi, Studer-Leuthi, et al. (2010, Study 2) assigned subjects to a dual *n*-back ($n = 25$), visuospatial single *n*-back ($n = 21$), or no-contact control ($n = 43$) group. Both *n*-back groups performed adaptive versions of the tasks for 20 sessions. All subjects completed both RAPM (11-min time limit) and BOMAT (16-min time limit) in counterbalanced order, among other measures, at pre- and posttest. In summarizing the intelligence transfer results, both the single *n*-back and dual *n*-back groups showed more improvement on RAPM (see Figure 3A) and BOMAT (see Figure 3B) than the control group, although the effect of dual *n*-back training appeared stronger in RAPM (dual *n*-back: $d = 0.98$; control: $d = .010$) than in BOMAT (dual *n*-back: $d = 0.49$; control: $d = 0.29$); indeed, the dual *n*-back BOMAT gain did not statistically differ from the control gain (S. M. Jaeggi, personal communication, September 15, 2010).

As for the dual *n*-back versus no-contact control group comparisons across the Jaeggi et al. (2008) and Jaeggi, Studer-Leuthi, et

[1] Jaeggi (2005) refers to a dissertation that presents the full results of the eight- and 19-session studies included in Jaeggi et al. (2008). This dissertation was downloaded from http://www.zb.unibe.ch/download/eldiss/05jaeggi_s.pdf and has been archived on WebCite (http://webcitation.org/662gjXMf3).

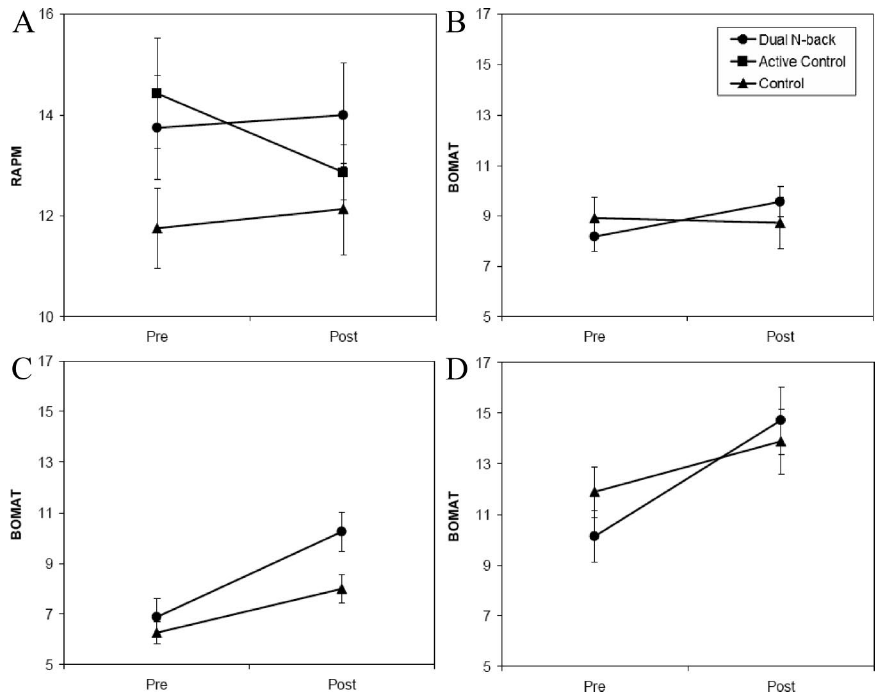

Figure 2. Raven Advanced Progressive Matrices (RAPM) performance for the eight-session study (A) and Bochumer Matrizentest (BOMAT) performance for the 12-session (B), 17-session (C), and 19-session (D) studies. Data provided by S. M. Jaeggi (personal communication, March 31, 2012). Error bars represent ±1 standard error of the mean.

al. (2010) studies, only the 17-session BOMAT results in Jaeggi et al. and the RAPM results in Jaeggi, Studer-Leuthi, et al. show clear evidence for transfer to fluid intelligence after adaptive dual *n*-back training. When one considers also that the comparison to a no-contact control group maximizes the likelihood of observing an effect of training that is influenced by placebo, Hawthorne, and related motivational and expectancy-based effects (French, 1953; Shipstead et al., 2010), the evidence for intelligence transfer after WM training is less compelling.

Jaeggi, Studer-Leuthi, et al. (2010) addressed a number of limitations of Jaeggi et al. (2008), including (a) using larger samples, (b) matching all training subjects on number of practice sessions, (c) administering both BOMAT and RAPM to all subjects as transfer measures, and (d) counterbalancing the order of BOMAT and RAPM versions across pre- and posttest sessions. However, the promising results of Jaeggi, Studer-Leuthi, et al. are limited by the use of a no-contact control group and only matrix-reasoning tests to measure intelligence. The present study sought to address these and other limitations in recent WM training studies.

Current Study

We followed several recommendations (Buschkuehl & Jaeggi, 2010; Shipstead et al., 2010; Sternberg, 2008) to critically

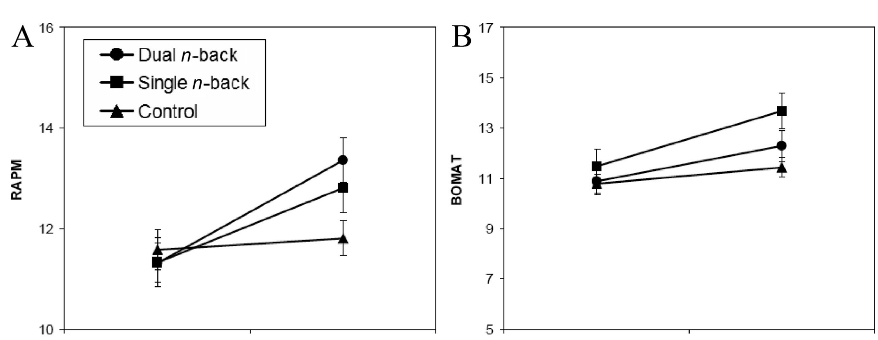

Figure 3. Raven Advanced Progressive Matrices (RAPM; A) and Bochumer Matrizentest (BOMAT; B) performance for the three groups in Jaeggi, Studer-Leuthi, et al. (2010). Error bars represent ±1 standard error of the mean.

examine the effectiveness of adaptive dual *n*-back training. As Sternberg (2008) argued, the Jaeggi et al. (2008) results are so potentially important that replicating them across different laboratories and samples is necessary. The idea that relatively brief dual *n*-back training can increase an individual's intelligence has important implications for applied contexts outside of the laboratory, such as educational practices and remediation for low-IQ individuals. So, one basic goal of the current study was to replicate the dual *n*-back training results by showing transfer to fluid intelligence.

We also followed Sternberg's (2008) recommendation to evaluate the efficacy of dual *n*-back training on "behaviors that extend beyond the realm of psychometric testing" (p. 6791). In previous work (Redick, Shipstead, et al., 2012), we found that dual *n*-back accuracy was positively correlated with performance on different measures of multitasking ability. In addition, other studies have shown that WM tasks are strong predictors of multitasking performance (Bühner, König, Pick, & Krumm, 2006; Colom, Martínez-Molina, Shih, & Santacreu, 2010; Hambrick, Oswald, Darowski, Rench, & Brou, 2010). We predicted that successful dual *n*-back training could also increase multitasking performance, especially because the dual *n*-back task, itself, requires a form of multitasking.

In addition, we included a diverse sample of young adults to ensure that the results of Jaeggi et al. (2008) and Jaeggi, Studer-Leuthi, et al. (2010) are not specific to those of above-average intelligence. As noted by Sternberg (2008), the University of Bern students in Jaeggi et al. represent a selective sample. Although the majority of the current subjects were college students, we sampled across three universities with varying academic profiles in an attempt to include subjects with a reasonably wide range of intelligence and WM abilities.

To draw conclusions about latent abilities, like intelligence, instead of particular tasks, like RAPM, we administered 17 transfer measures assessing fluid intelligence, multitasking, WM capacity, crystallized intelligence, and perceptual speed, with multiple verbal and nonverbal measures of each construct. Because multiple causes contribute to variance in performance on any single test, it is important to use numerous measures to rule out explanations based on task-specific abilities or processes. By creating factors for each of the constructs, we examined the efficacy of dual *n*-back training at the construct level. We also wanted to measure fluid intelligence with tasks other than matrix reasoning to ensure that any observed transfer was not due to the use of visuospatial materials in both the training and transfer tasks (Sternberg, 2008).

Our rationale for including multiple cognitive-ability constructs, beyond fluid intelligence, in the transfer sessions was not that we expected transfer to occur for all tasks or constructs. Rather, if fluid intelligence is actually improved via WM training, then tests showing the highest loadings on a general factor of intelligence (*g*) should also show the most transfer, and tests with the lowest *g* loadings would show the least; this is because fluid intelligence tests are more strongly associated with *g* than are other aspects of intelligence (Jensen, 1998; see Colom, Ángeles Quiroga, et al., 2010, for similar logic). Therefore, after dual *n*-back training, the largest cognitive-ability improvements should be observed on the fluid intelligence tests (which have very high *g* loadings), and the smallest improvements should be observed on the perceptual speed tests (which tend to have lower *g* loadings; Marshalek, Lohman, & Snow, 1983).[2]

The use of an active-control group is critical to elucidate the mechanisms responsible for transfer after any cognitive training (Buschkuehl & Jaeggi, 2010; Shipstead et al., 2010, 2012; Sternberg, 2008). Although there are no firm guidelines about what makes for a good active-control condition, we agree with Sternberg (2008) that the task should be as adaptive and challenging as the dual *n*-back, but not thought or previously shown to depend heavily upon WM. In this way, subjects' motivations, beliefs, expectations, and efforts would match between dual *n*-back and active-control groups, but their WM capacities (after training) would not. Therefore, we included an adaptive visual search training group in the current study, which we aimed to be as difficult and engaging as the dual *n*-back task and thus to serve as a placebo control. We chose visual search, in particular, because extensive studies with over 500 subjects have shown that individual differences in WM capacity are not related to performance in a variety of visual search tasks (Kane, Poole, Tuholski, & Engle, 2006). Because visual search performance is not likely to be determined by WM capacity, visual search training is unlikely to improve WM capacity. By comparing a visual search group that did not train WM to a dual *n*-back group that arguably did train WM, we can separate the potential transfer effects due to improving WM from those associated with placebo-type effects (Shipstead et al., 2010; Sternberg, 2008).

We also administered transfer sessions at three occasions (pre-, mid-, and posttest) to investigate the assertion by Jaeggi et al. (2008) that the amount of dual *n*-back training dosage determines intelligence transfer after training (see Figure 1B). Assessing this dose–response relationship within subjects, instead of between subjects as in Jaeggi et al., allowed us to potentially trace the growth of improvements in cognitive abilities as a function of amount of training. Based on Jaeggi et al., fluid intelligence gains for the dual *n*-back group during the midtest session (after 10 training sessions) might be expected to be small or nonexistent, but large by the posttest session (after 20 training sessions).

Predictions

We evaluated four possible transfer outcomes for the current study. The first (see Figure 4A) is that the processes trained via

[2] Our plan in designing the study was to use a correlated-vectors approach (Jensen, 1998) to interpret the pattern of expected transfer, similar to the logic used in Colom, Ángeles Quiroga, et al. (2010). We conducted this analysis, but because we observed no transfer for any tasks, it seemed unnecessary and redundant. We examined the pretest data for all subjects (N = 123 with data on all 17 pretest measures) and extracted one factor (principal axis factoring) from all the tests. The relative ordering of the tests indicated that, as predicted, the bulk of the fluid intelligence (G*f*) and multitasks (Multi) had high loadings on the general factor, and the perceptual speed (PS) tasks had the lowest loadings: (1) Raven Standard (G*f*), .71; (2) Raven Advanced (G*f*), .67; (3) Cattell's (G*f*), .66; (4) Letter Sets (G*f*), .62; (5) Control Tower (Multi), .62; (6) Number Series (G*f*), .60; (7) SynWin (Multi), .56; (8) Running Letter Span (WM), .52; (9) General Knowledge (crystallized intelligence [G*c*]), .52; (10) ATClab (Multi), .52; (11) Paper Folding (G*f*), .48; (12) Symmetry Span (WM), .46; (13) Inferences (G*f*), .46; (14) Analogies (G*f*), .45; (15) Vocabulary (G*c*), .29; (16) Number Comparison (PS), .21; (17) Letter Comparison (PS), .16.

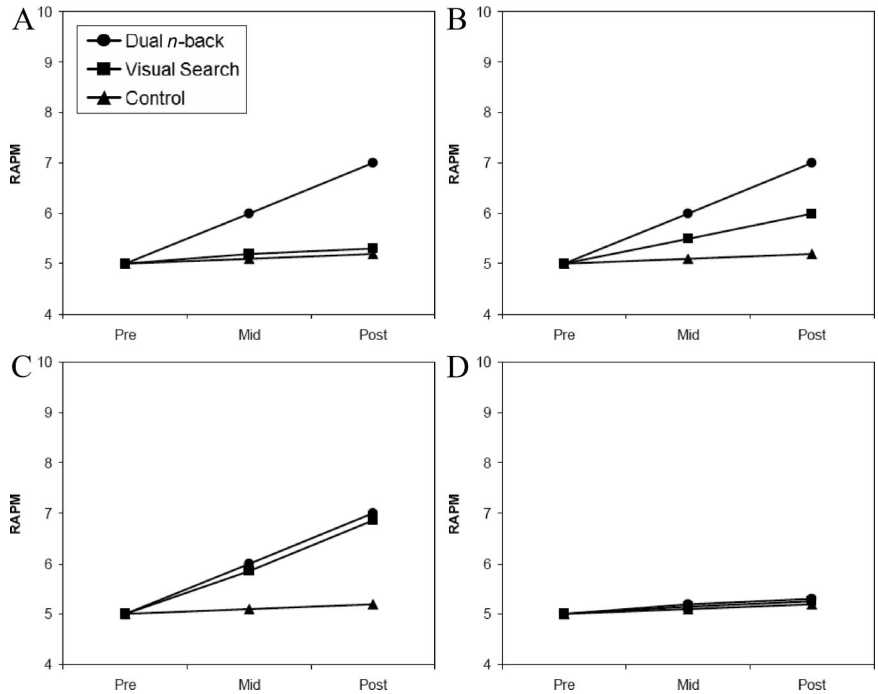

Figure 4. Four possible outcomes of current study. RAPM = Raven Advanced Progressive Matrices.

dual *n*-back practice cause specific improvements in fluid intelligence, whereas visual search practice does not. The second possibility (see Figure 4B) is that visual search training produces improvement relative to the no-contact control group, but dual *n*-back training yields even greater improvement; this would indicate that the visual search practice produced placebo effects and/or that visual search training increased fluid intelligence somewhat. In any case, the Figure 4B results would still indicate that dual *n*-back training improves intelligence benefits over and above placebo effects. A third possibility (see Figure 4C) is that the dual *n*-back and visual search training would increase fluid intelligence equivalently relative to the no-contact control groups. Equivalent transfer gains for the dual *n*-back and visual search training groups would indicate that the processes trained in the dual *n*-back are not specifically responsible for improving intelligence (leaving room for the possibility that cognitive training gains are entirely driven by placebo responses). Finally, the fourth possibility (see Figure 4D) is the null hypothesis—none of the groups show differential improvement on the intelligence tests; this result would be consistent with the eight-session study of Jaeggi et al. (2008), but here with 20 practice sessions.

Although the predictions above have used RAPM as the example dependent measure, they could extend to other fluid intelligence tests, and the multitasking and WM capacity measures, given that these measures have strong relationships to *g* (e.g., Hambrick et al., 2010). If we observed the same amount of transfer to crystallized intelligence and perceptual speed, which are predicted to have weaker relationships with *g*, then we could conclude that the cognitive processes trained in the adaptive dual *n*-back or visual search tasks are not specifically related to increases in fluid intelligence.

Method

Subjects

Subjects between 18 and 30 years of age completed practice and transfer sessions at Georgia Institute of Technology or Michigan State University; the sample included students from those universities, students from Georgia State University, and a small number of nonstudents. In addition to the 75 subjects who completed all sessions, an additional 55 subjects completed at least the pretest session. Thirty-six of these additional subjects began participation near the end of a semester. Most of these subjects wanted to continue participating after the semester break, but we determined that they could not continue in the study with a 3-week absence during the training period. The other 19 subjects (10 dual *n*-back, five visual search, and four control) began the study but did not complete all sessions. Out of the 75 subjects who completed all sessions, two control subjects were excluded from data analysis because they received the same transfer test items at pretest and posttest. Demographic information for the final sample is provided in Table 1.

Training Tasks

Adaptive dual *n*-back. Dual *n*-back training subjects performed 20 sessions of the adaptive task described previously. Subjects made button responses to visual (location of squares) and auditory (identity of letters) stimuli using their left and right index fingers, respectively. There were eight possible visual and auditory stimuli used. Simultaneous visual–auditory stimuli were presented for 500 ms, followed by a fixation screen for 2,500 ms in which

Table 1
Demographic Information

Group	*N*	Gender		Age (years)		No. of subjects			
		Male	Female	*M*	*SD*	GT	GSU	MSU	Other
Dual *n*-back	24	10	14	21.1	2.7	9	7	7	1
Visual search	29	12	17	20.7	2.5	9	11	8	1
Control	20	10	10	21.2	2.5	7	7	5	1

Note. GT = Georgia Tech student; GSU = Georgia State University student; MSU = Michigan State University student; other = not currently attending one of these three colleges.

subjects could respond. On each trial, subjects had four response choices: (a) visual match/left key, (b) auditory match/right key, (c) visual and auditory matches/both keys, and (d) no match/no response. Each block presented $n + 20$ trials, and subjects completed 20 blocks in each of the 20 sessions. Each block presented four visual targets, four auditory targets, two visual and auditory targets, and $14 + n$ nontargets.

Subjects' performance determined the level of *n* in the subsequent block. If the subject's visual and auditory accuracy was greater than or equal to 90% for the block, then *n* increased by 1; if accuracy was less than or equal to 70%, then *n* decreased by 1. Any other combination of visual and auditory performance meant no change in *n*. The first dual *n*-back session provided subjects with detailed instructions and examples before initiating one block each of 2-back, 3-back, and 4-back trials as practice before proceeding to the adaptive task. For Sessions 1–3, the adaptive task started at $n = 1$. For Sessions 4–20, the adaptive task started at $n = 2$. The dependent variable was the mean level of *n* reached during each session. In keeping with Jaeggi et al. (2008) and Jaeggi, Studer-Leuthi, et al. (2010), we used the mean *n* achieved during the session excluding Blocks 1–3, under the assumption that Blocks 1–3 largely represented practice for most subjects.

Adaptive visual search. Visual search training subjects performed 20 sessions of an adaptive task. On each trial, subjects reported whether a target *F* presented somewhere in the array was facing left or right with a left or right keypress, respectively; a leftward or rightward *F* always appeared among distractor stimuli (left- and right-facing *E*s, left- and right-tilted *T*s). After a 500-ms fixation, the visual search array appeared for 500 ms, followed by a 2,500-ms mask during which subjects could respond. Each block presented 24 trials, with equal numbers of left-facing and right-facing *F*s. Subjects completed 20 blocks in each of the 20 sessions.

Subjects' performance determined both the number and type of distractors in the subsequent trial block. Figure 5 provides examples of the search arrays from different levels of task adaptation. Target *F*s could appear anywhere in the array. On levels with homogeneous-distractor trials (e.g., Level 1, Level 3), the distractor identity changed across trials but was fixed within a trial (e.g., all right-facing *E*s on one trial, all left-tilted *T*s on another trial). The levels were ordered such that an array of a given size always started with homogeneous distractors, the subsequent level was the same size array but with heterogeneous distractors (e.g., mix of right-facing *E*s, left-facing *E*s, left-tilted *T*s, and right-tilted *T*s as distractors on a given trial), and the next level was an increase in array by adding two more columns and rows of homogeneous distractors. If subjects' accuracy was greater than or equal to 87.5% for the block, then the level increased by 1; if accuracy was less than or equal to 75%, then the level decreased by 1. Any other accuracy led to no change in the level on the next block. The first visual search session provided subjects with detailed instructions and examples before presenting one block each of 2×2 homogeneous (Level 1), 2×2 heterogeneous (Level 2), and 4×4 homogeneous (Level 3) trials as practice before proceeding to the adaptive task. For all sessions, the adaptive visual search task started at Level 1. The dependent variable was the mean level reached during each session.

Transfer Tasks

Computerized, alternate forms of the fluid and crystallized intelligence tests were created by taking the original items from each test and creating three test versions. The logic is similar to that used for previous studies that have divided tests into even and odd items to use for pre- and posttest (Chein & Morrison, 2010; Colom, Ángeles Quiroga, et al., 2010; Jaeggi et al., 2008; Jaeggi,

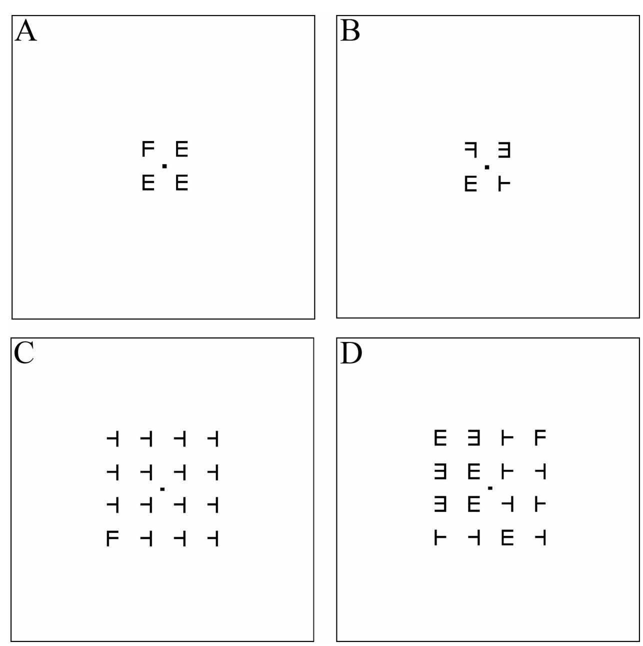

Figure 5. Example stimuli from different levels of the adaptive visual search task: 2×2 homogeneous (Level 1; A), 2×2 heterogeneous (Level 2; B), 4×4 homogeneous (Level 3; C), 4×4 heterogeneous (Level 4; D).

Studer-Leuthi, et al., 2010). However, because we had pre-, mid-, and posttest sessions, we divided the fluid and crystallized intelligence tests into thirds to create three alternate versions of each measure (for similar strategy with RAPM, see Basak, Boot, Voss, & Kramer, 2008). Because the fluid and crystallized intelligence measures were all power tests, with item difficulty approximately increasing with item number, we chose a "snake" procedure to distribute the items across versions so that the tests would have equal difficulty (see supplemental materials, Table S1). In addition, time limits were only used for RAPM and Raven Standard Progressive Matrices (RSPM); all other tests were given with no time limit. For the multitasks and perceptual speed tasks, parallel versions were created by generating unique versions from a pool of items. For the WM capacity tasks, we did not explicitly create alternate versions; because the items and order of items are generated randomly each time these programs are started, each test administration is always an alternate version.

Tests were presented in one of three orders (A, B, C) across sessions (see supplemental materials, Table S2). Test orders were generated so that tests of the same construct did not occur consecutively within a session and that each test appeared toward the beginning, middle, and end of a session across the three sessions. Test version order was counterbalanced across subjects via a Latin square procedure (ABC, BCA, CAB).

RAPM (fluid intelligence, spatial; Raven, Raven, & Court, 1998). Items present abstract shapes and patterns in a 3 × 3 matrix. The shape in the bottom right location is missing, and subjects must select from the eight possible choices the item that best completes the overall pattern both vertically and horizontally. Subjects had 10 min to complete the test. The number of correct responses out of 12 is used as the dependent variable.

RSPM (fluid intelligence, spatial; Raven et al., 1998). The test is similar to the advanced version, but the individual must select from either six or eight possible choices the item that best completes the matrix. Subjects had 15 min to complete the test. The number of correct responses out of 20 is used as the dependent variable.

Cattell Culture-Fair Test (fluid intelligence, spatial; Cattell, 1973). The Cattell test is composed of four subtests (Series Completion, Odd Elements, Matrix Completion, Dot Task) of spatial reasoning tasks. The number of correct responses out of 19 is used as the dependent variable.

Paper Folding (fluid intelligence, spatial; Ekstrom, French, Harman, & Dermen, 1976). Items present a square piece of paper on the left of the problem. The markings indicate that the paper has been folded a certain number of times, with a hole or holes then punched through the paper. Subjects decide which one of the five response choices depicts what the piece of paper would look like when completely unfolded. The number of correct responses out of six is used as the dependent variable.

Letter Sets (fluid intelligence, verbal; Ekstrom et al., 1976). Each item presents five sets of four letters, and subjects induce a rule that applies to the composition and ordering of four of the five letter sets, and then indicate the set that violates the rule. The number of correct responses out of 10 is used as the dependent variable.

Number Series (fluid intelligence, numeric; Thurstone, 1938). Each item presents a series of numbers, and subjects identify the response option that continues the sequence. The number of correct responses out of five is used as the dependent variable.

Inferences (fluid intelligence, verbal; Ekstrom et al., 1976). Each item presents a one-to-three sentence passage, and subjects choose the response option that is a logical necessity following only from the information provided. The number of correct responses out of six is used as the dependent variable.

Analogies (fluid intelligence, verbal; Berger, Gupta, Berger, & Skinner, 1990). Each item presented an analogy in the format of *A is to B as C is to D*, with either the *C* or the *C is to D* missing. Subjects chose which of the five response options best completed the analogy. The number of correct responses out of eight is used as the dependent variable.

SynWin (multitasking; Elsmore, 1994). A visual display with four simultaneous subtasks is presented, with each subtask in its own quadrant (see supplemental materials, Figure S1): (a) Probe-recognition: Five letters are presented briefly at the beginning of the task. Throughout the rest of the task, a probe letter is presented every 10 s, and the subject makes a yes/no decision whether the probe was in the memory set. Points were earned for correct answers, and points were subtracted for incorrect answers. (b) Arithmetic: Subjects must add two three-digit numbers, and a new problem is shown after an answer is submitted. Points were earned for correct answers, and points were subtracted for incorrect answers. (c) Visual monitoring: A fuel gauge continuously drops gradually from 100 to 0 and must be reset to 0 by clicking on the gauge. Points were earned for responding before the gauge reaches 0, and points were subtracted if the gauge reached 0. (d) Auditory monitoring: High- and low-frequency tones occur every 10 s, and subjects click on the quadrant when the rarely occurring high-frequency tone is presented. Points were earned for hits, and points were subtracted for misses or false alarms. Subjects were supposed to complete two 5-min blocks of the task, but because of experimenter errors, several subjects were not administered the second block of the task during each session. Therefore, SynWin performance was based on the first block in each session, which all subjects completed. The subject's score is determined by a formula that combines the points earned across all four subtasks, and this composite score is used as the dependent variable.

Control Tower (multitasking; Redick, Shipstead, et al., 2012). This multitask contains a primary comparison task with distractor tasks that interrupt primary task performance (see supplemental materials, Figure S2). The primary task is to search through side-by-side arrays of numbers, letters, and symbols. Certain elements of the left array are highlighted, and the appropriate items in the corresponding row of the right array must be clicked by the subject. For numbers, subjects click on the matching numbers in the right array. For letters, subjects click on the letter that precedes it alphabetically in the right array. For symbols, subjects click on the relevant symbols in the right array by referring to a consistently mapped symbol codebook. During this primary task, several distractor tasks occur that interrupt performance. For the radar task, subjects click on the radar when a blip occurs inside a specific area of the radar. For the airplane task, requests for landing are presented via headphones and the subject decides if one of three runways is clear for landing. For the color task, subjects press one of three buttons depending on the color that flashes. For the problem-solving task, subjects solve auditory questions by clicking the correct answer among the three response options provided. Subjects completed one 10-min block of the task. The subject's score on the primary task is determined by the number of correct comparisons (numbers, letters, symbols) minus incorrect comparisons, which is used as the dependent variable.

ATClab (multitasking; modified from Fothergill, Loft, & Neal, 2009). Each trial presents a display of four to 10 planes that move dynamically along various flight paths, traveling at varying rates of speed (see supplemental materials, Figure S3). Subjects are given a maximum of 1 min to make two to four yes/no decisions about whether two or three specific planes clustered together are in conflict given their current positions, flight path, and speed. For example, in Figure S3, Planes 3–5 will be in conflict because Plane 5's flight path intersects too closely with Planes 3 and 4 given the speed of each plane. The proportion of correct conflict decisions (45 decisions across 15 trials) is used as the dependent variable.

Symmetry Span (WM capacity, spatial; modified from Redick, Broadway, et al., 2012). Subjects made a vertical symmetry judgment about a black-and-white figure via mouse click and then were presented with a red square location within a 4×4 matrix that is to be remembered. The task is the same as the one described by Redick, Broadway, et al. (2012), with the exception that longer list lengths were used to try to avoid performance at ceiling at posttest. After three to six symmetry-square elements, subjects recalled the red squares in the order in which they were presented, by clicking on a blank 4×4 matrix. The total number of squares out of 54 recalled in the correct order across 12 trials (three trials each of 3, 4, 5, and 6 elements) is used as the dependent variable.

Running Letter Span (WM capacity, verbal; Broadway & Engle, 2010). Subjects recalled the final n letters that were presented sequentially every 500 ms, where n equaled 3–7 across trials. Subjects saw $n + 0$, $n + 1$, and $n + 2$ letters in each trial and clicked on the letters in a fixed response grid. The total number of letters out of 75 recalled in the correct serial order across 15 trials is used as the dependent variable.

Vocabulary (crystallized intelligence, verbal; Zachary, 1986). Each item presented a word and one of the four response options was a synonym. The total number correct out of 13 is used as the dependent variable.

General Knowledge (crystallized intelligence, verbal; Ekstrom et al., 1976). Items presented trivia questions about literature, world history, geography, and other topics, each with four response options. The total number correct out of 10 is used as the dependent variable.

Letter Comparison (perceptual speed, verbal; computerized version of Salthouse & Babcock, 1991). Subjects decided whether sets of three, six, or nine consonants on either side of a line were the same or different. If the sets were the same, subjects clicked *SAME*; if sets differed, subjects clicked *DIFFERENT*. Subjects had 30 s to complete as many comparisons as possible, with the total correct across two 30-s administrations used as the dependent variable.

Number Comparison (perceptual speed, numeric; computerized version of Salthouse & Babcock, 1991). This task was identical to Letter Comparison, but numbers were used instead of letters.

Procedure

Experimenters did not inform subjects that they were participating in a training study, nor did they give an indication that subjects should expect any aspect of performance to improve (in contrast to coaching methods used in commercial programs such as Cogmed). If subjects inquired about the study's purpose, they were told that the researchers were investigating the effects of practice on memory and attention tasks (a generic description applicable to both training and control subjects because there were a minimum of three test sessions for all subjects). During recruitment, we informed potential subjects that they should be available multiple times over the course of 4–5 weeks to complete the study.

For the pretest, midtest, and posttest sessions, subjects were compensated \$40 per completed session; subjects completing all three transfer sessions received a 10% bonus (\$12). On average, subjects took 2 hr and 20 min to complete the pretest session, and about 1 hr and 40 min to complete the mid- and posttest sessions.[3] Pretest sessions included collecting demographic information and longer instructions for the multitasks (including demonstration video).

Subjects in the two training groups completed an additional 20 practice sessions, each of which took between 30 and 40 min. Subjects could not complete more than one experimental session per day, with a maximum of seven sessions per week. Dual n-back and visual search subjects completed all 20 practice sessions (and the midtest session) in an average of 46 ($SD = 13.7$) and 41 ($SD = 10.7$) days, respectively; the time to complete the training did not differ for the two groups, $t(51) = 1.59$, $p = .12$. Dual n-back and visual search group subjects performed midtest sessions after 10 practice sessions.[4] Control subjects performed mid- and posttest sessions at approximately the same interval of days as the training groups, and the intervals did not differ for the three groups: pre- to midtest, $F(2, 70) = 1.23$, $p = .30$; pre- to posttest, $F(2, 70) = 1.58$, $p = .21$. Compensation for each practice session was \$10, with a 10% bonus at the end of the study for completing all practice sessions (\$20).

We assigned subjects to the dual n-back, visual search, and no-contact control groups as follows. We first assigned subjects to groups such that the pretest performance on the three multitasks was not significantly different between groups. Our attempt to match groups suffered, however, when subjects dropped out and were replaced by another on our standby list. Moreover, to ensure that we had adequate statistical power, we tested an extra wave of subjects where we randomly assigned them to one of the two training groups without consideration of their pretest data. Despite these complications, one-way analyses of variance (ANOVAs) on all 17 pretest measures indicated no significant differences among the no-contact, visual search, and dual n-back groups at pretest (all

[3] Although these are the averages of all the subjects in the final sample, it is possible that individuals in a particular group may have spent more or less time working on the 17 transfer tasks during the pre-, mid-, and posttest sessions. However, the duration of the sessions did not differ among the groups: pretest, $F(2, 70) = 0.39$, $p = .68$; midtest, $F(2, 70) = 0.26$, $p = .77$; posttest, $F(2, 70) = 1.46$, $p = .24$.

[4] Due to experimenter error, one dual n-back subject completed the midtest session after eight practice sessions.

ps $>$ .14). Of particular interest, the three groups performed similarly at pretest on the two WM measures: Symmetry Span, $F(2, 70) = 0.13$, $p = .88$, and Running Letter Span, $F(2, 70) = 0.11$, $p = .90$, despite no individual performing the same order of items and memoranda as any other subject in any group. To further assess whether the groups were different at pretest, we conducted Bonferroni-corrected post hoc comparisons, and all between-group comparisons were nonsignificant (all ps $>$ .17). Note that the inherent limitation in this pretest comparison is that we are relying upon the failure to reject the null hypothesis, an important consideration not just for the current research but for any training study arguing for a lack of pretest group differences, especially if the sample size is small.

All subjects completed a survey after the last task in the posttest session. Questions focused on the amount of perceived improvement in several categories, strategies used during the practice sessions, and self-reported engagement during the experiment. Most questions used a 4-point rating scale, although open-ended answers were allowed for two questions, too.

Design and Analyses

We evaluated dual n-back and visual search performance via repeated-measures ANOVAs with Practice 20 as the within-subjects factor. We evaluated transfer performance on the ability tasks via factorial ANOVAs with Group 3 as the between-subjects factor and Session 3 as the within-subjects factor. Significant Group $\times$ Session interactions were decomposed with simple effects analyses focusing on the effects of Group and Session independently. Partial eta-squared is reported as index of effect size. Because of the number of analyses that were conducted, an alpha of .01 was used for all transfer analyses (two-tailed).

Results

Practice Effects

As shown in Figure 6, both training groups improved with practice. For the dual n-back group (see Figure 6A), the practice effect was significant, $F(19, 437) = 18.77$, $p < .01$, $\eta_p^2 = .45$. We also observed substantial individual differences in dual n-back performance and dual n-back improvement across the 24 subjects: One subject reached $n = 10$, whereas another subject maxed out at only $n = 4$. As a crude index, we also examined individual differences in improvement by comparing the subjects' maximum n of Session 20 to their maximum n of Session 1. Twenty-two of the 24 subjects achieved a higher n in Session 20 compared to Session 1, but whereas one subject's maximum n improved by 6, the next highest improvement by any other subject was 3.

Practice also significantly improved performance for the visual search group, $F(19, 475) = 17.30$, $p < .01$, $\eta_p^2 = .41$ (see Figure 6B), from an approximate mean level of 5 (6×6 homogeneous) to 7 (8×8 homogeneous). Again, we observed substantial individual differences in visual search performance and visual search improvement across the 29 subjects. For example, whereas one subject reached Level 12, indicating accurate discrimination of a left- or right-facing F among 123 distractors from a masked array, another subject reached asymptote at Level 6 (35 distractors). Twenty-one subjects achieved a higher level in Session 20 compared to Session 1, with seven subjects improving by four levels.

Transfer Data

Descriptive statistics for each of the transfer tasks are presented in Table 2. Because the results clearly indicate no transfer effects for either the dual n-back or visual search group relative to the control group, for brevity, the significance testing results for the transfer data are provided in Table 3. Out of 17 ANOVAs, there were no significant Group $\times$ Session interactions.[5]

Given our emphasis on using multiple indicators of a construct instead of single tasks, we also examined transfer performance as a function of ability composites for the tasks representing fluid intelligence (separate spatial and nonspatial factors), multitasking, WM capacity, crystallized intelligence, and perceptual speed. We calculated composites in a manner analogous to that reported in Jaeggi, Buschkuehl, Jonides, and Shah (2011). Specifically, for each task, two standardized gain scores were calculated for each group for the intervals of pre- to midtest and pre- to posttest. The standardized gain scores were created by taking the gain for each subject and dividing by the pretest standard deviation for the entire sample, collapsing across the groups. To create the composite gain scores, we averaged the standardized gain scores across the relevant constructs. The statistical results are presented in Table 4. Overall, the composite analyses confirm the lack of transfer observed in the individual task analyses. Note that the marginal results for the gain from pre- to midtest for the multitasking composite represent greater improvement for the no-contact control group relative to the other two groups (see Table 2).

To facilitate comparisons with previous research, we reanalyzed all transfer data excluding the visual search group. The results matched the full analyses—no transfer for the dual n-back group relative to the no-contact control group. In addition, we reanalyzed all transfer data excluding the midtest session, comparing only the pre- and posttest sessions, and, again, we found no transfer for the dual n-back group relative to the no-contact or active-control groups. Finally, we conducted ANCOVAs for each transfer task, using the pretest score as the

[5] With an alpha of .05, the Group $\times$ Session interaction for RSPM was significant (see Table 2). However, as shown in Table 2, the interaction did not reflect improvements for the dual n-back or visual search group relative to controls; simple main-effects analyses (one-way ANOVAs) indicated that the three groups did not significantly differ from one another at pre-, mid-, or posttest (all ps $>$ .26). Repeated-measures ANOVAs calculated separately for each group revealed that the interaction was due to a significant effect of Session only for the no-contact control group, $F(2, 38) = 5.02$, $p = .01$, $\eta_p^2 = .21$; for the dual n-back and visual search groups, Fs < 1.

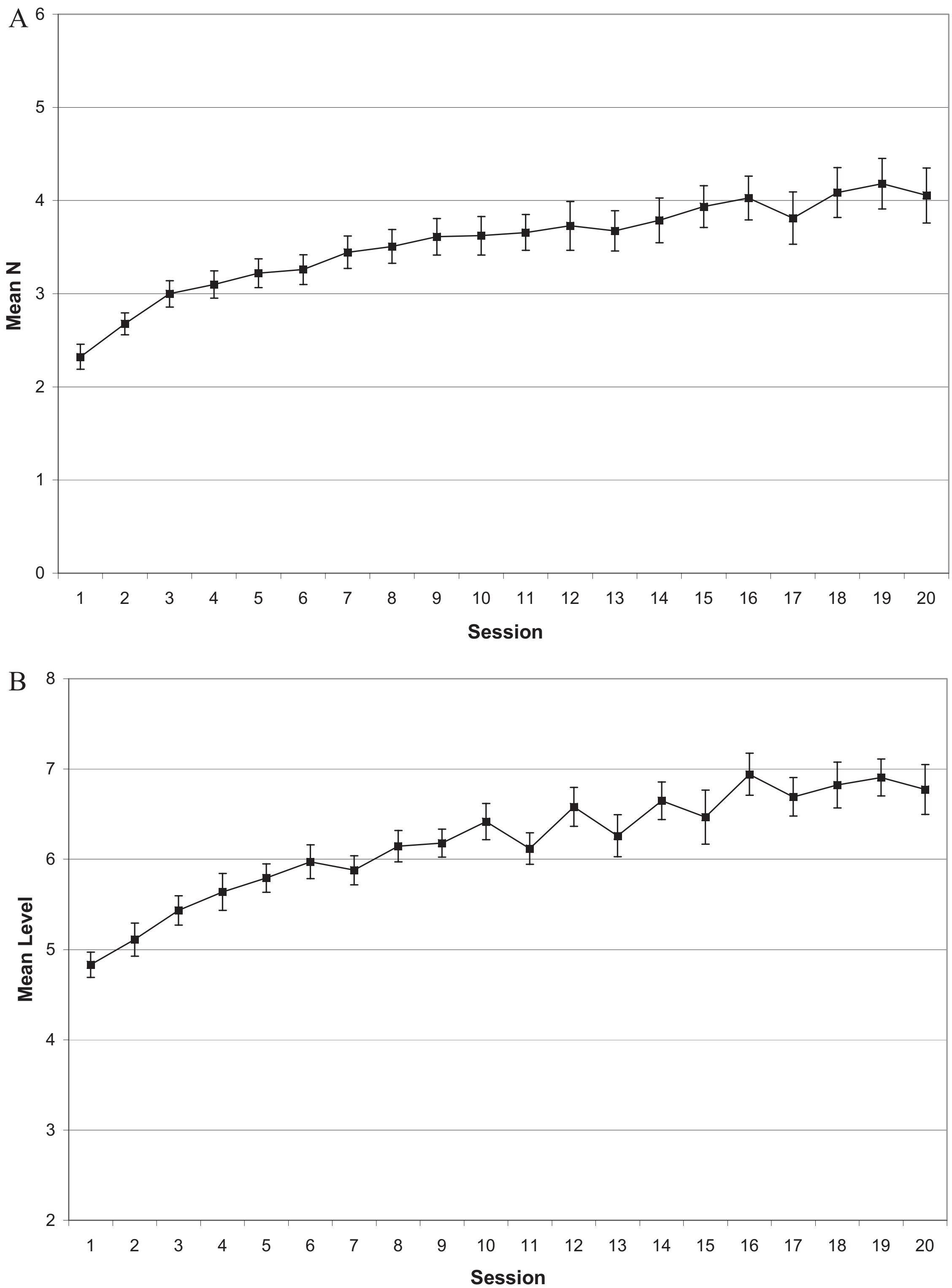

Figure 6. Practice data for the dual *n*-back (A) and visual search (B) tasks. Error bars represent ±1 standard error of the mean.

Table 2
Mean Performance for the Transfer Tasks as a Function of Training Group and Session

Task	Control			Visual search			Dual *n*-back		
	Pre	Mid	Post	Pre	Mid	Post	Pre	Mid	Post
RAPM	6.65 (2.18)	6.45 (2.50)	6.00 (3.00)	6.52 (3.04)	6.07 (2.87)	6.24 (3.34)	7.04 (2.48)	6.17 (2.28)	6.25 (3.08)
RSPM[a]	17.15 (2.39)	15.85 (2.50)	16.85 (2.35)	16.66 (2.53)	16.34 (2.30)	16.45 (2.47)	16.30 (2.67)	16.74 (2.54)	16.09 (2.61)
Cattell	11.95 (2.63)	11.75 (2.05)	11.45 (2.65)	10.72 (2.79)	11.07 (2.07)	11.24 (2.25)	12.00 (2.38)	11.71 (2.29)	11.38 (2.45)
Paper Folding	4.05 (1.70)	4.50 (1.43)	4.00 (1.26)	4.41 (1.38)	4.00 (1.60)	4.52 (1.34)	3.79 (1.47)	4.46 (1.69)	4.33 (1.34)
Letter Sets	6.85 (1.90)	6.75 (2.29)	6.80 (2.22)	7.79 (1.84)	6.90 (2.16)	6.83 (2.19)	7.08 (2.45)	7.17 (1.52)	7.04 (2.14)
Number Series	4.20 (0.83)	3.75 (0.85)	3.70 (1.22)	3.59 (1.32)	3.76 (1.35)	3.52 (1.24)	3.96 (0.96)	3.92 (1.18)	3.75 (1.19)
Inferences	4.35 (1.31)	4.30 (1.78)	4.45 (1.54)	4.41 (1.40)	4.03 (1.84)	4.24 (1.60)	3.67 (1.97)	4.04 (1.60)	4.04 (1.65)
Analogies	4.90 (1.59)	4.65 (1.42)	3.90 (1.71)	4.83 (1.65)	4.45 (1.64)	4.38 (1.66)	4.46 (1.62)	4.46 (1.53)	3.79 (1.50)
SynWin	352.40 (626.95)	682.50 (190.05)	701.50 (214.60)	461.14 (252.22)	625.76 (205.12)	729.14 (193.33)	480.28 (218.08)	581.88 (231.66)	655.08 (201.34)
Control Tower[b]	29.62 (10.45)	32.47 (11.22)	34.05 (11.02)	29.90 (11.38)	29.63 (13.87)	37.43 (15.63)	29.20 (8.10)	31.41 (11.54)	34.26 (11.29)
ATClab[c]	0.72 (0.09)	0.73 (0.12)	0.72 (0.12)	0.73 (0.14)	0.71 (0.12)	0.75 (0.12)	0.74 (0.13)	0.75 (0.09)	0.75 (0.12)
Symmetry Span	25.60 (9.30)	30.25 (9.34)	28.90 (12.14)	24.55 (11.15)	27.28 (12.53)	26.76 (10.99)	25.88 (8.75)	32.29 (9.92)	31.54 (11.80)
Running Span	38.50 (8.65)	40.90 (8.81)	43.00 (9.06)	39.52 (13.03)	39.52 (12.42)	42.34 (12.49)	37.96 (12.69)	40.13 (10.86)	42.21 (11.94)
Vocabulary	10.10 (1.21)	10.70 (1.08)	10.35 (1.79)	10.00 (1.79)	10.38 (1.43)	9.79 (1.74)	10.33 (1.13)	10.04 (1.40)	10.50 (1.41)
Knowledge	6.75 (1.83)	6.30 (2.18)	6.20 (1.61)	5.90 (1.99)	6.17 (1.91)	6.10 (1.78)	6.25 (2.03)	6.04 (1.23)	6.29 (1.81)
Letter Comparison	18.75 (3.77)	20.65 (3.50)	20.85 (3.72)	19.93 (4.08)	20.45 (5.68)	20.45 (5.38)	19.04 (4.84)	19.92 (4.03)	21.38 (3.61)
Number Comparison	28.90 (5.21)	31.15 (4.61)	31.00 (4.24)	29.14 (6.11)	29.52 (5.92)	29.93 (7.02)	28.83 (5.54)	28.58 (4.95)	29.00 (5.43)

Note. Standard deviations are shown in parentheses. Pre = pretest; mid = midtest; post = posttest; RAPM = Raven Advanced Progressive Matrices; RSPM = Raven Standard Progressive Matrices.
[a] $N = 23$ for dual *n*-back group due to experimenter error during midtest session. [b] $N = 19$ for dual *n*-back group due to computer problem during posttest session. [c] $N = 28$ for visual search group and $N = 19$ for control group due to computer problem during posttest session.

Table 3
Significance Testing Results for the Transfer Measures

Task	Group			Session			Group × Session		
	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Fluid intelligence (spatial)									
Raven Advanced	0.05	.95	.001	2.00	.14	.028	0.33	.86	.009
Raven Standard	0.06	.94	.002	1.53	.22	.022	2.85	.03	.076
Cattell	1.08	.35	.030	0.28	.75	.004	0.92	.45	.026
Paper Folding	0.11	.90	.003	0.75	.47	.011	1.92	.11	.052
Fluid intelligence (verbal/numeric)									
Letter Sets	0.31	.74	.009	1.14	.32	.016	1.08	.37	.030
Number Series	0.70	.50	.020	1.56	.21	.022	0.78	.54	.022
Inferences	0.70	.50	.020	0.20	.82	.003	0.67	.61	.019
Analogies	0.46	.64	.013	*6.08*	*.00*	*.080*	0.69	.60	.019
Multitasking									
SynWin[a]	0.16	.85	.005	*29.95*	*.00*	*.300*	1.81	.13	.049
Control Tower	0.26	.98	.001	*17.28*	*.00*	*.200*	1.96	.10	.054
ATClab	0.23	.78	.007	0.32	.73	.005	0.80	.53	.023
Working memory capacity									
Symmetry Span	1.02	.37	.028	*10.30*	*.00*	*.128*	0.70	.59	.020
Running Letter Span	0.24	.98	.001	*8.64*	*.00*	*.110*	0.38	.82	.011
Crystallized intelligence									
Vocabulary	0.57	.57	.016	0.68	.51	.010	1.62	.17	.044
General Knowledge	0.37	.70	.010	0.17	.85	.002	0.68	.61	.019
Perceptual speed									
Letter Comparison	0.02	.98	.001	*5.55*	*.01*	*.073*	0.98	.42	.027
Number Comparison	0.58	.57	.016	1.54	.22	.022	0.75	.56	.021

Note. Entries in italics indicate values significant at $\alpha = .01$.
[a] At the Michigan State University testing location, 21 subjects were administered the same test version of SynWin at pre-, mid-, and posttest. Data were reanalyzed with subjects who only performed unique versions of SynWin at all three transfer sessions ($N = 14$, 21, and 17 for control, visual search, and dual *n*-back, respectively). The interpretation of the significance tests was the same as listed above with the full data.

covariate. The results were qualitatively the same as the ANOVA results presented in Table 3.[6]

A post hoc power analysis (G*Power; Faul, Erdfelder, Lang, & Buchner, 2007) indicated that with our sample size, we had sufficient power to detect a significant Group (between subjects) × Session (within subjects) interaction, if it was present in our transfer data. Our power to detect a large ($f = .40$) or medium ($f = .25$) effect was greater than .99, based on our sample size and the use of the within-subjects correlation of .53, which was the average correlation among the repeated measures across all 17 transfer tasks (the default value in G*Power is $r = .50$). We also reran the power analyses using a correlation among repeated measures of .30, which was the lowest observed correlation among pre-, mid-, and posttest performance, ignoring test version (Paper Folding). With $r = .30$, the power to detect a large or medium effect was greater than .95. Note that we based our power analysis on a medium or large effect of dual *n*-back training, given the previous literature ($d = 0.68$, Jaeggi et al., 2008; $d = 0.98$ for RAPM, $d = 0.49$ for BOMAT, Jaeggi, Studer-Leuthi, et al., 2010).

Survey Data

The dual *n*-back and visual search groups did not differ in their ratings for either effort (dual *n*-back: $M = 3.13$, $SD = 0.63$; visual search: $M = 3.32$, $SD = 0.61$), $F(1, 49) = 1.21$, $p = .28$, $\eta_p^2 = .024$, or enjoyableness (dual *n*-back: $M = 2.09$, $SD = 0.79$; visual search: $M = 2.39$, $SD = 0.79$), $F(1, 49) = 1.90$, $p = .18$, $\eta_p^2 = .037$, for the repeated practice sessions. All subjects were asked about their perceived improvement as a function of study participation. When asked whether they thought that their performance

[6] The only slight difference between the ANCOVA and ANOVA results came from the Control Tower task, where the Group × Session interaction ($p = .037$) indicated a potential difference between the groups in the nature of their session effect from mid- to posttest after covarying pretest performance. Similar to the RSPM results (see Footnote 5), simple main-effects analyses (one-way ANOVAs) indicated that the three groups did not significantly differ from one another at pre-, mid-, or posttest (all $ps > .59$). Also similar to the RSPM results, inspection of Table 2 indicates that the direction of the difference between the mid- and posttest sessions likely drove the ANCOVA interaction result. Whereas the control and dual *n*-back groups showed slight numerical increases from mid- to posttest, the visual search group showed a numerically larger improvement. However, note that numerically the visual search group showed a slightly lower midtest score than the other two groups, indicating more room from improvement from mid- to posttest. Most importantly, there is no evidence here for greater Control Tower transfer gain from dual *n*-back training than from visual search training or no contact.

Table 4
Inferential Results of the Transfer Composite Standardized Gain Scores

Construct	Midtest	Posttest
Fluid intelligence (spatial)	$F(2, 69) = 0.83$, $p = .44$	$F(2, 70) = 0.39$, $p = .68$
Fluid intelligence (verbal)	$F(2, 70) = 1.51$, $p = .23$	$F(2, 70) = 0.62$, $p = .54$
Multitasking	$F(2, 70) = 3.09$, $p = .05$	$F(2, 67) = 1.44$, $p = .24$
Working memory	$F(2, 70) = 1.89$, $p = .16$	$F(2, 70) = 0.89$, $p = .41$
Crystallized intelligence	$F(2, 70) = 1.44$, $p = .24$	$F(2, 70) = 0.16$, $p = .86$
Perceptual speed	$F(2, 70) = 1.15$, $p = .32$	$F(2, 70) = 0.86$, $p = .43$

had improved by the third session, the three groups did not differ, $F(2, 69) = 1.46$, $p = .24$, $\eta_p^2 = .040$, with all but two subjects perceiving that their performance improved either "moderately so" or "very much so." Proportions of subjects who endorsed specific improvements in each of several abilities are presented in Table 5. Chi-square tests indicated a difference across the three groups in the proportion of "Yes" responses for memory ($p = .02$) and intelligence ($p = .06$), with the dual *n*-back group more likely to report memory and intelligence changes than the visual search and control groups. Numerically, the visual search group had the highest rates of endorsement to changes in attention and perception, although the chi-square tests were not significant for these categories. Dual *n*-back subjects also endorsed changes to their everyday functioning. Ten of 23 dual *n*-back subjects said the study changed the way they carried out their daily activities, compared to only five of 49 combined visual search and control subjects. When given the opportunity to elaborate on these changes, dual *n*-back subjects offered, for example, "My ability to multitask has improved," "Better short term memory when doing tasks," "I think it helps me focus better in class and while studying," and "How to memorize orders at work."

Finally, we asked the dual *n*-back and visual search groups about their strategies for the practice sessions, first as an open-ended question about what advice they would give to a friend who was just beginning the study. The most consistent response across the visual search and dual *n*-back groups involved getting sufficient rest before beginning the sessions. Other suggested dual *n*-back strategies included (a) grouping items into sets of three, (b) visualizing the letter inside the blue square, (c) forgetting old items, and (d) giving more attention to the auditory task instead of the visual task. The dual *n*-back subject that reached $n = 10$ listed both the chunking and visualization strategies. Suggested visual search strategies included (a) fixating on the central location, (b) unfocusing attention to passively encode the array, (c) moving one's head further away from the computer screen, and (d) resting during the intertrial interval and then preparing for the array when the subsequent fixation point appeared. The visual search subject who reached Level 12 listed the passive encoding strategy. After the open-ended question, subjects were given five task-specific strategies and had to rate from 1 (*almost never*) to 4 (*almost always*) the degree to which they used the strategies listed during the training task. Analyses of the open-ended and forced-response strategy questions showed that there were no systematic relationships between the self-reported strategy used and training task performance: Subjects who reached high levels of performance reported using similar strategies as those who reached low levels of performance.

Discussion

Our study yielded three main findings. First, subjects improved with practice on both the dual *n*-back and visual search tasks. Second, training group subjects showed no transfer to any of the ability measures, in keeping with the prediction outlined in Figure 4D. Third, dual *n*-back trained subjects reported subjective improvements in various aspects of cognition in the absence of any objective evidence for change.

Of importance, we observed dual *n*-back improvement with practice that was consistent with the training gains shown in previous dual *n*-back training studies with young adults. If we had not obtained such training gains, then the null transfer effects obviously would have been uninformative. We were also successful in designing an adaptive, active-control treatment (visual search training) that yielded the same amount of experimental

Table 5
Posttest Survey Data

Topic	Dual *n*-back (%)	Visual search (%)	Control (%)	$\chi^2(2)$	*p*
Attention	52	72	50	3.27	.20
Intelligence	65	41	30	5.73	.06
Language	4	3	10	1.06	.59
Memory	78	45	40	8.01	.02
Perception	35	59	45	2.98	.23
Daily activities	43	10	10	10.51	<.01

Note. Due to experimenter error, survey data were not available for one dual *n*-back subject and two control subjects. However, survey data were included for the two control subjects who received the same transfer test items at pretest and posttest. The format of the question for each topic was "Do you feel that your participation in this study has changed your ___ ?"

contact as the dual *n*-back task, as well as similar self-reported effort and enjoyment. Despite the performance improvements on the dual *n*-back and visual search tasks, no positive transfer to any of the intelligence, multitasking, WM capacity, and perceptual speed tasks was observed (although transfer was not expected for the crystallized intelligence and perceptual speed tasks). In addition, we did not find any evidence of a dose-dependent relationship between the amount of dual *n*-back training practice and fluid intelligence gains. That is, based on Jaeggi et al. (2008), fluid intelligence gains for the dual *n*-back group during the midtest session might have been expected to be small or nonexistent, but large by the posttest session. However, we did not observe this pattern for any of the fluid intelligence measures.

WM Training and Transfer to Fluid Intelligence

Clearly, then, the question is, Why did not we observe fluid intelligence transfer for the dual *n*-back group? One possible answer is that we did not observe transfer simply because WM transfer effects to intelligence are actually not commonly observed. After completing the current study, we became aware of another adaptive dual *n*-back training study that is relevant for understanding our results. First, Seidler et al. (2010) assigned subjects to a dual *n*-back ($n = 29$) or knowledge trainer active-control ($n = 27$) group.[7] Subjects in the knowledge trainer group answered multiple-choice and vocabulary questions. Although the knowledge trainer control task was not adaptive based on the subject's performance, it provided a similar amount of contact with experimenters as did the dual *n*-back task. All subjects completed a minimum of 17 practice sessions, and multiple transfer measures during pre- and posttest sessions. Despite significant dual *n*-back practice improvements, there was no significant transfer for the dual *n*-back group versus the active-control group on BOMAT, RAPM, or verbal analogies, although transfer was observed on Operation Span.

A recent review (Morrison & Chein, 2011) of the broader WM training literature with young adult subjects detailed (a) four studies reporting significant transfer to reasoning and intelligence measures (Jaeggi et al., 2008; Klingberg, Forssberg, & Westerberg, 2002; Olesen, Westerberg, & Klingberg, 2004; Westerberg & Klingberg, 2007), (b) three published studies reporting no significant transfer to reasoning and intelligence measures (Chein & Morrison, 2010; Dahlin et al., 2008; Owen et al., 2010), and (c) one study reporting significant transfer to some intelligence measures but not others (Schmiedek, Lövdén, & Lindenberger, 2010). Two of the significant transfer studies in the review (Klingberg et al., 2002; Olesen et al., 2004) had training group samples of $n =$ 4 and 3, respectively. The subjects in Olesen et al. (2004) were the same as those in Westerberg and Klingberg (2007; T. Klingberg, personal communication, February 14, 2010). Note that the positive intelligence transfer observed in Jaeggi, Studer-Leuthi, et al. (2010) and the lack of transfer observed in Seidler et al. (2010) were not included in Morrison and Chein's (2001) review. In addition, Morrison and Chein's assessment of training benefits may have been unwittingly biased because of the file-drawer problem (Rosenthal, 1979), in which nonsignificant transfer results such as those described in the current research are less likely to be published.

However, a recent meta-analysis by Melby-Lervåg and Hulme (2012) indicates that even when considering published studies, few appropriately powered empirical studies have found evidence for transfer from various WM training programs to fluid intelligence. Melby-Lervåg and Hulme reported that WM training showed evidence of transfer to verbal and spatial WM tasks ($d = 0.79$ and 0.52, respectively). When examining the effect of WM training on transfer to nonverbal abilities tests in 22 comparisons across 20 studies, they found an effect of 0.19. Critically, a moderator analysis showed that there was no effect ($d = 0.00$) in the 10 comparisons that used a treated control group, and there was a medium effect ($d = 0.38$) in the 12 comparisons that used an untreated control group.

More specifically, concerning the efficacy of adaptive dual *n*-back training in young adults, there are two results reporting transfer to RAPM and/or BOMAT when compared to a no-contact control group (Jaeggi et al., 2008; Jaeggi, Studer-Leuthi, et al., 2010) and one result of no transfer to RAPM and/or BOMAT when compared to an active-control group (Seidler et al., 2010). Our data show no transfer to RAPM or other measures of fluid intelligence when compared to either a no-contact control or active-control group. On the whole, then, our lack of significant fluid intelligence transfer results may not be that surprising.

In addition, other WM training studies have used children, older adults, or special populations (e.g., stroke patients, children with attention-deficit/hyperactivity disorder) as subjects. As with the young adults in Jaeggi, Studer-Leuthi, et al. (2010), certain studies of older adults have shown preliminary evidence that WM training

[7] Although Seidler et al. (2010) is a technical report, the data subsume the transfer results reported in Experiment 2 of the recently published Anguera et al. (2012) journal article (R. D. Seidler, personal communication, December 27, 2011). Anguera et al. reported that

> participants completed two days of pre-testing as part of a larger study. Of relevance here, the test battery included a working memory assessment using an *n*-back task ($n = 3$ and 4) with abstract shapes . . . , an automated operation span task . . . , as well as the card rotation task and the digit symbol substitution task from Experiment 1, and finally, a visuomotor adaptation task. (p. 110).

Anguera et al. reported positive transfer to the *n*-back and Operation Span tasks.

Seidler et al. (2010) included these transfer tests, but also listed seven other tests that did not exhibit significant dual *n*-back transfer: BOMAT, RAPM, verbal analogies, visual arrays comparison, Attention Network Test, motor sequence learning task, and various conditions of a driving simulator task. Perhaps most importantly, the technical report provides an attempt at replication as the reason for including the fluid intelligence measures:

> Type 2 tests included Raven's matrices (Raven et al., 1990), which is a standardized test of fluid intelligence, and the BOMAT and verbal analogies tests of intelligence (Hossiep et al., 1995). We have previously shown that working memory training transfers to performance on this task (Jaeggi et al., 2008), and we included it here for the sake of replication. (p. 7)

The Seidler et al. technical report was downloaded from http://m-castl.org/files/2010-01SeidlerReport.pdf and has been archived on WebCite (http://www.webcitation.org/662goIQRW).

can transfer to untrained measures of fluid intelligence. For example, Borella, Carretti, Riboldi, and De Beni (2010) trained older adults on a WM span-like task and compared them to a control group of older adults that completed questionnaires instead. They reported positive transfer to several tasks, including the Cattell Culture-Fair Test. However, other studies have been less optimistic. For example, three other recent studies reported no transfer to different versions of the Raven Progressive Matrices after WM span training in older adults (Brehmer, Westerberg, & Bäckman, 2012; Richmond, Morrison, Chein, & Olson, 2011; Zinke, Zeintl, Eschen, Herzog, & Kliegel, 2012). Likewise, a recent review (Shipstead et al., 2012) found that the majority of published studies using developmental and patient samples have not observed transfer to fluid intelligence after WM training. The Melby-Lervåg and Hulme (2012) meta-analysis confirmed that age was not a significant moderator of transfer to nonverbal abilities after WM training. In fact, young children ($d = 0.03$) and adolescent children ($d = -0.05$) showed no evidence of transfer to nonverbal intelligence, inconsistent with the idea that younger children may be more susceptible to WM training and improvements in intelligence because of their increased brain plasticity relative to adults. Again, our findings do not appear to be an aberration; there is little evidence for transfer from WM training to fluid intelligence.

Limitations

We acknowledge a limitation in our data related to three of the fluid intelligence tasks. Specifically, for Number Series, Paper Folding, and Inferences, the mean pretest scores were close to the maximum possible score, and this was likely due to our use of shortened versions of these tests, which left five or six items per test version. Given this limitation, the reader may put less emphasis on the nonsignificant results from these three transfer tasks. However, this ceiling effect problem does not affect the interpretation of any of the other 14 transfer measures. Moreover, we reanalyzed the spatial and verbal fluid intelligence composite standardized gain scores, after removing Number Series, Paper Folding, and Inferences. None of the ANOVAs on the composites were significant (spatial fluid intelligence: midtest, $p = .54$; posttest, $p = .55$; verbal fluid intelligence: midtest, $p = .15$; posttest, $p = .78$).

In addition, it was difficult to assess the reliability of the shortened intelligence measures we used. Ideally, we would use the pretest data in order to calculate Cronbach's alpha, before either the effects of practice or the specific interventions could influence performance (as could occur on the mid- and posttest sessions). However, because subjects performed a different version (A, B, or C) at pretest, the sample sizes for each version ($N =$ 20, 24, and 29) were too small to calculate useful Cronbach's alpha information. Likewise, we could instead calculate test–retest reliability, but this would need to be done using only the no-contact control group, which again had a limited sample size ($N = 20$) for a meaningful test–retest correlation.

As alternative measures of reliability, the *g* factor loadings (see Footnote 2) indicate there was substantial systematic variance in our shortened tests. For example, RSPM and RAPM had the highest loadings (.71, .67), and Letter and Number Comparison had the lowest loadings (.21, .16). Note that the two Raven tasks might be suspected of having low reliabilities due to shortening the test, but the *g* loadings indicate that such suspicions are not warranted. In addition, we calculated the squared multiple correlations as an estimate of reliability reflecting the communalities of the 17 transfer measures. These values ranged from .23 (Vocabulary) to .69 (Number Comparison). None of the values obtained were particularly low (for similar analysis and comparison, see Engle, Tuholski, Laughlin, & Conway, 1999).

Note also that shortened measures of tasks such as RAPM have been used in the literature, without adverse effect upon reliability. For example, Arthur and Day (1994) developed a 12-item version of RAPM. Arthur and Day reported in a sample of 461 young adults that their 12-item version of RAPM had a Cronbach's alpha of .69 and a test–retest reliability of .75. In addition, Basak et al. (2008) divided the RAPM into three sections of 12 items each in order to have a pre-, mid-, and posttest administration. Basak et al. found evidence of transfer to RAPM performance (a Group $\times$ Session interaction), indicating again that our use of a shortened RAPM administration did not preclude the possibility of observing transfer.

One final note on the point of reliability: In the current context—which was an experiment (and not an individual-differences study)—what is most important is whether or not there are changes produced in the transfer measures as a function of the intervention. That is, in this experimental design, our interest is in between-subjects variability, and more specifically between-subjects variability in the pre- to posttest difference score. We know that several experimental effects in cognitive psychology have low reliability in terms of internal consistency (e.g., Stroop effect difference scores), yet we still appropriately use these measures in experimental research because we want to see whether a particular between-subjects manipulation (e.g., vocal vs. manual responses, proportions of congruent vs. incongruent trials) produces an observable change in the dependent variable (for more on this issue, see Salthouse, Siedlecki, & Krueger, 2006).

WM Training and Transfer to WM

Although general or broad transfer after repeated practice on a specific task may not be typical, the idea that transfer may occur to tasks that share "identical elements" (Thorndike & Woodworth, 1901) is reasonable. Thus, although the lack of evidence for "far-transfer" from WM training to fluid intelligence may not be surprising, especially given the broader history of intelligence training research (Carroll, 1993), we also did not find evidence for "near-transfer" to two WM span tasks. Across studies, WM training typically leads to changes in untrained WM tasks (Morrison & Chein, 2011; Shipstead et al., 2012). Subjects who train on adaptive simple or complex span measures of WM have exhibited transfer to other untrained versions of simple and complex span measures of WM (Bergman Nutley et al., 2011; Chein & Morrison, 2010; Klingberg et al., 2005). Likewise, within the adaptive *n*-back training literature, subjects who train on single or dual *n*-back typically show transfer to untrained versions of the task (Jaeggi, Studer-Leuthi, et al., 2010; Seidler et al., 2010).

However, as mentioned in the introduction, *n*-back tasks and span measures of WM are weakly related to each other (Jaeggi, Buschkuehl, et al., 2010; Jaeggi, Studer-Leuthi, et al., 2010; Kane, Conway, Miura, & Colflesh, 2007; Oberauer, 2005), despite both types of measures correlating with fluid intelligence tasks such as RAPM. If there are no (or few) overlapping processes between the two types of WM measures, then improvement on one task (*n*-back) would not likely improve performance on the other (span tasks). Table 6 summarizes the relevant studies that have examined *n*-back training and transfer across WM task types. There are many differences among the studies (training procedures, sample sizes, transfer tasks, etc.), but the results show that transfer across types of WM tasks is inconsistent. Clearly, further work is necessary to understand what different WM processes *n*-back and span tasks tap and how these processes overlap with other constructs such as fluid intelligence (Burgess, Gray, Conway, & Braver, 2011; Schmiedek, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009).

Variables That Affect Transfer

If WM training, and more specifically dual or single *n*-back training, can actually cause real improvements in fluid intelligence, then the diversity of transfer results across studies indicates that there are important boundary variables that can mediate or moderate training effectiveness. A recent study with children who trained on an adaptive single *n*-back task identified the amount of *n*-back improvement as a critical variable determining whether or not transfer to intelligence was observed (Jaeggi et al., 2011). When Jaeggi et al. (2011) halved the *n*-back training group based on the amount of improvement observed, the children with the biggest gain showed transfer relative to an active-control group, whereas the children with smaller gains did not. We therefore attempted a similar analysis, by dividing our dual *n*-back subjects into high- and low-improvement groups, using a median split on the difference score of mean dual *n*-back level in Sessions 19–20 versus Sessions 1–2. This post hoc analysis is limited by sample size (only 12 and 12 subjects in the high- and low-improvement groups, respectively), but with that caveat in mind, no significant Group (high dual *n*-back improvement, low dual *n*-back improvement, no-contact control) effects were obtained for the fluid intelligence, multitasking, and WM composite standardized gains (see Table 7). A similar median-split analysis for the visual search group (15 and 14 subjects in the high- and low-improvement groups, respectively) also produced no significant Group effects on the composite standardized gains (see Table 7).

We also correlated the amount of training gain and transfer gain for the same four standardized gain composites (see Table 7). Dual *n*-back improvement was not associated with fluid intelligence gains; it was marginally correlated with WM capacity improvement, but surprisingly, visual search improvement was also correlated with improvement on the verbal fluid intelligence tasks (see supplemental materials, Figure S4). Other WM training studies (Chein & Morrison, 2010; Jaeggi et al., 2011) reporting significant correlations between training change and transfer change suffer from the same limitations as our data for such correlational analyses: small sample sizes and the influence of subjects who performed worse at posttest than pretest on the transfer tasks (i.e., negative value for transfer gain) and performed worse at the end of training than the beginning of training on the training task (i.e., negative value for training gain). Indeed, in our data, the correlation between visual search change and verbal fluid intelligence change was no longer significant, $r(28) = .25$, $p = .20$, after removing the lone subject who had negative values on both variables.

Other studies reporting a relationship between WM training gain and fluid intelligence test improvement have been equivocal. In a training study of young adults using adaptive versions of a neutral and an emotional dual *n*-back task, Schweizer, Hampshire, and Dalgleish (2011) reported a nearly significant correlation between training gain and RSPM gain. However, Loosli, Buschkuehl, Perrig, and Jaeggi (2012) trained children using an adaptive WM span task and tested fluid intelligence using the Test of Nonverbal Intelligence. There was no transfer, and also no relationship between training gain and matrix reasoning test improvement from pre- to posttest in this study. A potential relationship between the amount of training improvement and the amount of transfer is intuitively appealing. In fact, the data from previous dual *n*-back training studies could be reanalyzed to see whether the amount of training improvement affected the amount of intelligence transfer. Of course, if such associations hold up to replication, it would then be important to understand why only some individuals benefit from the training intervention (it would also be important to provide clear evidence that this correlation reflected a causal relation between gain scores).

Clearly, the amount of *n*-back improvement observed is only one possible variable that might affect the presence or amount of transfer. Others include the pretraining ability level of the sample, the size of the samples, the number and duration of training sessions, the transfer tests used, the administration method of the transfer tests, the spacing of the training sessions, the motivation of the subjects, the subjects' knowledge about the goals of the study, and the experimenters' influences on subjects' behavior. Because training studies are difficult to conduct in terms of time (here 23 sessions per training group subject) and can be financially costly (here $352 per training group subject), it is critical that as many factors as possible are ruled out.

Table 6
Working Memory (n-Back) Training and Transfer Effects Across Task Types

Study	Training	*n*-back transfer	Span transfer
Jaeggi et al. (2008)[a]	Dual	*Dual*	*Reading*
Jaeggi, Studer-Leuthi, et al. (2010)	Single, dual	*Single*	Operation
Seidler et al. (2010)	Dual	*Single, dual*	*Operation*
Li et al. (2008): Young/older adults	Single	*Single/single*	Operation/operation, rotation/rotation
Schmiedek et al. (2010): Young/older adults	Single[b]	*Single*/single	Reading/reading, counting/counting, rotation/*rotation*

Note. Tasks in italics produced significant ($p < .05$) Group (training vs. control) × Session (pretest vs. posttest) interactions.
[a] Significant transfer for 19-session group reported in Jaeggi (2005). [b] Task was part of a battery of training tasks administered to subjects.

Table 7
Standardized Gain Composite Transfer Results Based on Amount of Training Gain

Composite	Pretest to posttest			Correlation	
	F	p	η_p^2	r	p
Dual n-back (n = 24)					
Fluid intelligence (spatial)	0.83	.44	.039	.24	.26
Fluid intelligence (verbal)	0.68	.51	.032	−.19	.37
Multitasking	0.58	.56	.029	.30[a]	.17
Working memory capacity	1.68	.20	.076	.39	.06
Visual search (n = 29)					
Fluid intelligence (spatial)	0.66	.52	.028	−.10	.60
Fluid intelligence (verbal)	0.32	.73	.014	.36	.06
Multitasking	0.84	.44	.037	−.07[b]	.71
Working memory capacity	2.21	.12	.088	−.19	.33

[a] N = 23. [b] N = 28.

Although a detailed discussion of all the aforementioned variables is outside the scope of the current article (for a review, see Shipstead et al., 2012), three variables warrant further comment here. First, perhaps our lack of transfer represented a lack of motivation by our participants. Motivation is not easily measured, and we agree that subjects who are not motivated to perform either the training or transfer sessions would severely impact the ability to detect improvements as a function of training. Indeed, research has shown that motivation can account for nonability variance in performance on intelligence tests (Duckworth, Quinn, Lynam, Loeber, & Stouthamer-Loeber, 2011). Note, however, that in addition to self-reported effort not differing between our two training groups, we observed significant increases in performance on the training tasks for both the dual n-back and visual search groups and significant session effects on six of the transfer measures. Moreover, the amount of dual n-back practice gain in our sample (mean n, Session 1 = 2.3; mean n, Session 20 = 4.1) was slightly greater than the dual n-back group in Anguera et al. (2012), but lower than Jaeggi et al. (2008; 19-session group) and Jaeggi, Studer-Leuthi, et al. (2010). Although it is potentially important to eventually understand why different patterns of improvement are observed across WM training studies, we argue that no matter the cause, there is nothing fundamentally different about the performance observed in our current experiment compared to the existing studies using adaptive dual n-back training.

Second, given the previous literature on aptitude–treatment interactions (Cronbach & Snow, 1977), and broad cognitive differences between individuals high and low in WM capacity, it would not be unreasonable to think that WM training (treatment) may also vary in its effectiveness depending on a number of factors, including initial WM level (aptitude). Although the current study was not set up to address the presence of aptitude–treatment interactions directly, we examined initial ability level (fluid intelligence and WM capacity) to see whether pretest transfer task scores were related to training gain (a similar analysis to one reported in Jaeggi et al., 2011). We found that there were no significant correlations between the amount of dual n-back gain and pretest scores on the WM capacity factor, $r(24) = .14$, $p = .51$, fluid intelligence–spatial factor, $r(24) = .14$, $p = .51$, or fluid intelligence–verbal factor, $r(24) = .29$, $p = .17$. Note, however, that the sample sizes are rather small for such analyses, and the dual n-back gain variable was a difference score. Other research on intelligence interventions examined the role of preexisting individual differences in specific intellectual abilities (for review, see Carroll, 1993). Studies by Salomon (1974) and Kyllonen, Lohman, and Snow (1984) demonstrated that pretraining individual differences in verbal and/or spatial abilities interacted with the type of training program, indicating that certain training methods were more or less effective for certain individuals. This is an important consideration for future WM training research as well.

Third, given the diverse nature of WM training programs and procedures being used in research and in commercial applications, empirical analyses of the best practices would be helpful to maximize training efficacy. How many sessions of training are necessary? Schmiedek et al. (2010) administered 100 training sessions, whereas other studies have exhibited transfer after only three sessions (Borella et al., 2010). Many WM training programs (Cogmed, n-back) use an adaptive procedure to continually challenge subjects. However, other studies have shown transfer with a static training regimen (Schmiedek et al., 2010). In addition, some WM training procedures involve practicing on many different tasks during and across practice sessions (Klingberg et al., 2005; Schmiedek et al., 2010), with the idea that diverse practice will consequently cause broader and more general cognitive transfer. In contrast, the single and dual n-back training studies use the same task throughout training. Finally, physical exercise training regimens have been related to cognitive improvements, especially in older adults (Colcombe & Kramer, 2003). More research examining the combined efficacy of exercise and WM training (e.g., Fabre, Chamari, Mucci, Masse-Biron, & Prefaut, 2002) might lead to more effective training procedures and provide some information about the underlying physiological mechanisms of WM training.

What Does WM Training Actually Train?

As indicated by the subjective survey responses, subjects believed that certain aspects of their cognition had been affected by the experiment, even though objectively none of the transfer measures reflected differences over and above practice effects. Our questionnaire findings thus appear to indicate so-called illusory placebo effects, whereby trained subjects report subjective improvement in the absence of any objective improvement (see Pratkanis, Eskenazi, & Greenwald, 1994). The potential for such illusions should raise interpretative concerns whenever WM-trained subjects are compared to no-contact controls, as attendant motivational and self-efficacy changes might improve transfer task performance of trained subjects, even in the absence of any underlying "ability" improvements (Shipstead et al., 2010, 2012). This might include persistence in solving the difficult items of intelligence tests; instead of giving up when problems become harder later in the test, the individuals that believe the training has improved their abilities may be more likely to continue to attempt to solve the problem.

Of course, we cannot rule out the possibility that dual *n*-back training actually did change individuals' daily life abilities, because we did not attempt to verify these behaviors in this study. Other cognitive training studies have shown little to no evidence for positive transfer to the performance of everyday functioning (e.g., Willis et al., 2006). However, in a recent Cogmed training study with young and older adult subjects, transfer was observed to simple span tasks similar to those included in the training program, but not to the Stroop task or RSPM (Brehmer et al., 2012). Interestingly, subjects in the adaptive training groups self-reported fewer cognitive problems after completing the study, as evidenced by a significant Group × Session interaction on the Cognitive Failures Questionnaire. For both our study and Brehmer et al. (2012), the self-report results could be interpreted as reflecting subjects' implicit ideas about the processes involved in the training tasks, because practice improvements were observed on the training tasks.

More generally, we think that the self-report strategy results highlight the lack of knowledge about what is being trained in dual *n*-back and other WM training programs. As outlined elsewhere (Shipstead et al., 2010), understanding the mechanisms responsible for transfer in WM training studies is an important goal. Such understanding may require further task-analytic studies of the training procedures in order to isolate the cognitive processes that are being trained or improved. As indicated in our survey data, self-reported strategy use differed not only between subjects but also within subjects; more generally, it seems reasonable to ask whether a subject who has attained an *n* of 10 via dual *n*-back training is engaging many (any?) of the same mental processes and strategies in the task as is a subject who attains an *n* of only 4. A fundamental understanding of the processes involved in the performance of the dual *n*-back and other WM training programs is thus important to understanding why transfer occurs (if and when it does). Research on the physiological (Jaeggi et al., 2007) and psychometric (Jaeggi, Buschkuehl et al., 2010; Jaeggi, Studer-Leuthi, et al., 2010, Study 1; Redick, Shipstead, et al., 2012) properties of the dual *n*-back are informative, but these studies did not use the adaptive versions of the dual *n*-back tasks, nor did they assess performance across multiple sessions.

Finally, we strongly advocate the full report of all transfer tasks and experimental conditions assessed in WM training studies so that any significant transfer results can be interpreted within the context of measures that did not show significant transfer. The transfer results for any one particular measure must be statistically analyzed and interpreted within the full pattern of results, avoiding selective reporting and uncorrected comparisons. Having full knowledge of previous WM training procedures and results (whether statistically significant or not) will help future researchers identify the mechanisms responsible for transfer and understand the boundary conditions of WM training. On a broader level, in order to minimize the file-drawer problem, publication of studies that do not find transfer adds to the overall context of interpreting the efficacy of WM training.

Conclusion

A critical reexamination of dual *n*-back training studies indicated the need for replication and extension of previous research, in line with the recommendations of Buschkuehl and Jaeggi (2010), Shipstead et al. (2010), and Sternberg (2008). Subjects in an adaptive dual *n*-back training group were compared to both an adaptive visual search training (placebo control) group and a no-contact control group. Despite significant improvements on the training tasks, subjects showed no positive transfer to fluid intelligence, multitasking, WM capacity, crystallized intelligence, or perceptual speed tasks. The current study presents a pessimistic view of the effects of dual *n*-back practice, but future research might identify variables that maximize the potential intellectual benefits of WM training.

References

Anguera, J. A., Bernard, J. A., Jaeggi, S. M., Buschkuehl, M., Benson, B. L., Jennett, S., . . . Seidler, R. D. (2012). The effects of working memory resource depletion and training on sensorimotor adaptation. *Behavioural Brain Research, 228,* 107–115. doi:10.1016/j.bbr.2011.11.040

Arthur, W., Jr., & Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices test. *Educational and Psychological Measurement, 54,* 394–403. doi:10.1177/0013164494054002013

Basak, C., Boot, W. R., Voss, M. W., & Kramer, A. F. (2008). Can training in a real-time strategy video game attenuate cognitive decline in older adults? *Psychology and Aging, 23,* 765–777. doi:10.1037/a0013494

Berger, F. R., Gupta, W. B., Berger, R. M., & Skinner, J. (1990). *Air Force Officer Qualifying Test (AFOQT) Form P: Test manual* (AFHRL Tech. Rep. 89-56). Brooks Air Force Base, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Bergman Nutley, S., Söderqvist, S., Bryde, S., Thorell, L. B., Humphreys, K., & Klingberg, T. (2011). Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: A controlled, randomized study. *Developmental Science, 14,* 591–601. doi:10.1111/j.1467-7687.2010.01022.x

Borella, E., Carretti, B., Riboldi, F., & De Beni, R. (2010). Working memory training in older adults: Evidence of transfer and maintenance effects. *Psychology and Aging, 25,* 767–778. doi:10.1037/a0020683

Brehmer, Y., Westerberg, H., & Bäckman, L. (2012). Working-memory training in younger and older adults: Training gains, transfer, and maintenance. *Frontiers in Human Neuroscience, 6,* 63. doi:10.3389/fnhum.2012.00063

Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods, 42,* 563–570. doi:10.3758/BRM.42.2.563

Bühner, M., König, C., Pick, M., & Krumm, S. (2006). Working memory dimensions as differential predictors of the speed and error aspect of multitasking performance. *Human Performance, 19,* 253–275. doi:10.1207/s15327043hup1903_4

Burgess, G. C., Gray, J. R., Conway, A. A., & Braver, T. S. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology: General, 140,* 674–692. doi:10.1037/a0024695

Buschkuehl, M., & Jaeggi, S. M. (2010). Improving intelligence: A literature review. *Swiss Medical Weekly, 140,* 266–272.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Chicago, IL: Rand-McNally.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York, NY: Cambridge University Press. doi:10.1017/CBO9780511571312

Cattell, R. B. (1973). *Measuring intelligence with the Culture Fair tests.* Champaign, IL: Institute for Personality and Ability Testing.

Chein, J. M., & Morrison, A. B. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. *Psychonomic Bulletin & Review, 17,* 193–199. doi:10.3758/PBR.17.2.193

Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science, 313,* 1307–1310. doi:10.1126/science.1128317

Colcombe, S., & Kramer, A. F. (2003). Fitness effects on the cognitive function of older adults: A meta-analytic study. *Psychological Science, 14,* 125–130. doi:10.1111/1467-9280.t01-1-01430

Colom, R., ´Angeles Quiroga, M., Shih, P.-C., Martínez, K., Burgaleta, M., Martínez-Molina, A., . . . Ramírez, I. (2010). Improvement in working memory is not related to increased intelligence scores. *Intelligence, 38,* 497–505. doi:10.1016/j.intell.2010.06.008

Colom, R., Martínez-Molina, A., Shih, P.-C., & Santacreu, J. (2010). Intelligence, working memory, and multitasking performance. *Intelligence, 38,* 543–551. doi:10.1016/j.intell.2010.08.002

Conway, A. R. A., & Getz, S. J. (2010). Cognitive ability: Does working memory training enhance intelligence? *Current Biology, 20,* R362–R364. doi:10.1016/j.cub.2010.03.001

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions.* Oxford, England: Irvington.

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,* 720–730. doi:10.1037/a0014296

Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 7716–7720. doi:10.1073/pnas.1018601108

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Elsmore, T. F. (1994). SYNWORK: A PC-based tool for assessment of performance in a simulated work environment. *Behavior Research Methods, Instrumentation, and Computers, 26,* 421–426. doi:10.3758/BF03204659

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,* 309–331. doi:10.1037/0096-3445.128.3.309

Fabre, C., Chamari, K., Mucci, P., Masse-Biron, J., & Prefaut, C. (2002). Improvement of cognitive function by mental and/or individualized aerobic training in healthy elderly subjects. *International Journal of Sports Medicine, 23,* 415–421. doi:10.1055/s-2002-33735

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191. doi:10.3758/BF03193146

Fothergill, S., Loft, S. D., & Neal, A. (2009). ATC-lab^{Advanced}: An air traffic control simulator with realism and control. *Behavior Research Methods, 41,* 118–127. doi:10.3758/BRM.41.1.118

French, J. R. P. (1953). Experiments in field settings. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences* (pp. 98–135). New York, NY: Holt, Rinehart, and Wilson.

Hambrick, D. Z., Oswald, F. L., Darowski, E. S., Rench, T. A., & Brou, R. (2010). Predictors of multitasking performance in a synthetic work paradigm. *Applied Cognitive Psychology, 24,* 1149–1167. doi:10.1002/acp.1624

Highfield, R. (2008, April 28). "Brain training" games do work, study finds. *The Telegraph.* Retrieved from http://www.telegraph.co.uk/science/science-news/3340967/Brain-training-games-do-work-study-finds.html

Jaeggi, S. M. (2005). *Capacity limitations in human cognition: Behavioural and biological considerations* (Unpublished doctoral dissertation). University of Bern, Bern, Switzerland.

Jaeggi, S. M., Buschkuehl, M., Etienne, A., Ozdoba, C., Perrig, W. J., & Nirkko, A. C. (2007). On how high performers keep cool brains in situations of cognitive control. *Cognitive, Affective, & Behavioral Neuroscience, 7,* 75–89. doi:10.3758/CABN.7.2.75

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 of the United States of America, 105,* 6829–6833. doi:10.1073/pnas.0801268105

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 of the United States of America, 108,* 10081–10086. doi:10.1073/pnas.1103228108

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the *N*-back task as a working memory measure. *Memory, 18,* 394–412. doi:10.1080/09658211003702171

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. doi:10.1016/j.intell.2010.09.001

Jensen, A. R. (1998). *The* g *factor: The science of mental ability.* Westport, CT: Praeger.

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 21–48). New York, NY: Oxford University Press.

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the *n*-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 615–622. doi:10.1037/0278-7393.33.3.615

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131,* 66–71. doi:10.1037/0033-2909.131.1.66

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,* 749–777. doi:10.1037/0278-7393.32.4.749

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

Klingberg, T., Fernell, E., Olesen, P., Johnson, M., Gustafsson, P., Dahlström, K., . . . Westerberg, H. (2005). Computerized training of working memory in children with ADHD—A randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry, 44,* 177–186. doi:10.1097/00004583-200502000-00010

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology, 24,* 781–791. doi:10.1076/jcen.24.6.781.8395

Kyllonen, P. C., Lohman, D. F., & Snow, R. E. (1984). Effects of aptitudes, strategy training, and task facets on spatial task performance. *Journal of Educational Psychology, 76,* 130–145. doi:10.1037/0022-0663.76.1.130

Li, S.-C., Schmiedek, F., Huxhold, O., Röcke, C., Smith, J., & Lindenberger, U. (2008). Working memory plasticity in old age: Practice gain, transfer, and maintenance. *Psychology and Aging, 23,* 731–742. doi:10.1037/a0014343

Loosli, S. V., Buschkuehl, M., Perrig, W. J., & Jaeggi, S. M. (2012). Working memory training improves reading processes in typically developing children. *Child Neuropsychology, 18,* 62–78. doi:10.1080/09297049.2011.575772

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7,* 107–127. doi:10.1016/0160-2896(83)90023-5

Medalia, A., & Freilich, B. (2008). The Neuropsychological Educational Approach to Cognitive Remediation (NEAR) model: Practice principles and outcome studies. *American Journal of Psychiatric Rehabilitation, 11,* 123–143. doi:10.1080/15487760801963660

Melby-Lervåg, M., & Hulme, C. (2012). Is working memory training effective? A meta-analytic review. *Developmental Psychology.* Advance online publication. doi:10.1037/a0028228

Morrison, A. M., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review, 18,* 46–60. doi:10.3758/s13423-010-0034-0

Oberauer, K. (2005). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General, 134,* 368–387. doi:10.1037/0096-3445.134.3.368

Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H.-M. (2005). Working memory and intelligence—Their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131,* 61–65. doi:10.1037/0033-2909.131.1.61

Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience, 7,* 75–79. doi:10.1038/nn1165

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., . . . Ballard, C. G. (2010). Putting brain training to the test. *Nature, 465,* 775–778. doi:10.1038/nature09042

Perrig, W. J., Hollenstein, M., & Oelhafen, S. (2009). Can we improve fluid intelligence with training on working memory in persons with intellectual disabilities? *Journal of Cognitive Education and Psychology, 8,* 148–164. doi:10.1891/1945-8959.8.2.148

Pratkanis, A. R., Eskenazi, J., & Greenwald, A. G. (1994). What you expect is what you believe (but not necessarily what you get): A test of the effectiveness of subliminal self-help audiotapes. *Basic and Applied Social Psychology, 15,* 251–276. doi:10.1207/s15324834basp1503_3

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* New York, NY: Psychological Corporation.

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment.* Advance online publication. doi:10.1027/1015-5759/a000123

Redick, T. S., Shipstead, Z., Meier, M. E., Evans, J., Hicks, K. L., Unsworth, N., . . . Engle, R. W. (2012). *Cognitive predictors of multitasking performance.* Manuscript in preparation.

Richmond, L. L., Morrison, A. B., Chein, J. M., & Olson, I. R. (2011). Working memory training and transfer in older adults. *Psychology and Aging, 26,* 813–822. doi:10.1037/a0023631

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86,* 638–641. doi:10.1037/0033-2909.86.3.638

Salomon, G. (1974). Internalization of filmic schematic operations in interaction with learners' aptitudes. *Journal of Educational Psychology, 66,* 499–511. doi:10.1037/h0036753

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27,* 763–776. doi:10.1037/0012-1649.27.5.763

Salthouse, T. A., Siedlecki, K. L., & Krueger, L. E. (2006). An individual differences analysis of memory control. *Journal of Memory and Language, 55,* 102–125. doi:10.1016/j.jml.2006.03.006

Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1089–1096. doi:10.1037/a0015730

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad abilities in adulthood: Findings from the COGITO study. *Frontiers in Aging Neuroscience, 2,* 27. doi:10.3389/fnagi.2010.00027

Schweizer, S., Hampshire, A., & Dalgleish, T. (2011). Extending brain-training to the affective domain: Increasing cognitive and affective executive control through emotional working memory training. *PLoSONE, 6,* e24372. doi:10.1371/journal.pone.0024372

Seidler, R. D., Bernard, J. A., Buschkuehl, M., Jaeggi, S., Jonides, J., & Humfleet, J. (2010). *Cognitive training as an intervention to improve driving ability in the older adult* (Tech. Rep. No. M-CASTL 2010-01). Ann Arbor: University of Michigan.

Shellenbarger, S. (2011, November 29). Ways to inflate your IQ. *New York Times.* Retrieved from http://online.wsj.com/article/SB10001424052970203935604577066293669642830.html

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

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin.* Advance online publication. doi:10.1037/a0027473

Sternberg, R. J. (2008). Increasing fluid intelligence is possible after all. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 6791–6792. doi:10.1073/pnas.0803396105

Thorell, L. B., Lindqvist, S., Bergman, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science, 12,* 106–113. doi:10.1111/j.1467-7687.2008.00745.x

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. (I.) *Psychological Review, 8,* 247–261. doi:10.1037/h0074898

Thurstone, L. L. (1938). *Primary mental abilities.* Chicago, IL: University of Chicago Press.

Wang, S., & Aamodt, S. (2009, March 10). Guest column: Can we increase our intelligence? *New York Times.* Retrieved from http://judson.blogs.nytimes.com/2009/03/10/guest-column-can-we-increase-our-intelligence/

Westerberg, H., & Klingberg, T. (2007). Changes in cortical activity after training of working memory—A single-subject analysis. *Physiology & Behavior, 92,* 186–192. doi:10.1016/j.physbeh.2007.05.041

Willis, S. L., Tennstedt, S. L., Marsiske, M., Ball, K., Elias, J., Mann Koepke, K., . . . Wright, E. (2006). Long-term effects of cognitive training on everyday functional outcomes in older adults. *Journal of the American Medical Association, 296,* 2805–2814. doi:10.1001/jama.296.23.2805

Zachary, R. A. (1986). *Shipley Institute of Living Scale: Revised manual.* Los Angeles, CA: Western Psychological Services.

Zinke, K., Zeintl, M., Eschen, A., Herzog, C., & Kliegel, M. (2012). Potentials and limits of plasticity induced by working memory training in old-old age. *Gerontology, 58,* 79–87. doi:10.1159/000324240

Received July 3, 2011
Revision received May 8, 2012
Accepted May 17, 2012 ■