The Hype Cycle of Working Memory Training

Thomas S. Redick

Thomas S. Redick, Purdue University, Department of Psychological Sciences, 703 Third Street, West Lafayette, IN 47907, tredick@purdue.edu, (765) 494-5132
Abstract

Seventeen years and hundreds of studies after the first working memory training journal article was published, evidence for the efficacy of working memory training is still wanting. Numerous studies show that individuals who repeatedly practice computerized working memory tasks improve on those tasks and closely related variants. Critically, although individual studies have claimed improvements in untrained abilities and behaviors, systematic reviews of the broader literature show that studies producing large, positive findings are often those with the most methodological shortcomings. The current review discusses the past, present, and future status of working memory training, including consideration of factors that might influence working memory training and transfer efficacy.

Keywords: working memory, training, transfer
The Hype Cycle of Working Memory Training

“Does working memory training work?”. Over the past 16 years, this question has been asked with increasing frequency by researchers, parents, educators, science journalists, and consumers interested in improving their mental functioning. The idea is enticing – a cognitive workout involving the maintenance and manipulation of information, with minimal risk and time commitment, could produce measurable improvements in academic behaviors and daily activities. As with many questions in psychology, the answer is not a simple ‘yes’ or ‘no’, but one that requires a more nuanced consideration of both the similarities and differences across studies, along with the quality of evidence within each study.

Working memory is the system involved in storing information over short periods of time, usually to accomplish a goal and in the face of interference from other external stimuli and internal representations competing for our attention. Over 40 years of research has demonstrated that working memory is strongly and positively related to cognitive abilities in academic settings (e.g., reading comprehension) and in the workplace (e.g., multitasking). In typically-developing individuals, working memory capacity increases until early adulthood, and then subsequently declines throughout middle and late adulthood. In various clinical populations (e.g., schizophrenia, attention-deficit/hyperactivity disorder (ADHD)), working memory impairment is robust and perhaps a core feature of the disorder. Because individuals lower in working memory capacity are more likely to have below-average attention, comprehension, and long-term memory abilities compared to their higher-working memory peers, psychologists have long contemplated methods to improve human working memory. Working memory training research piggy-backed on advances in computer technology in an attempt to provide such an intervention.
Table 1

Common Types of Working Memory Training Tasks

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex span</td>
<td>Also known as processing-and-storage tasks; items to be serially recalled (e.g., letters) are interleaved with secondary decisions (e.g., math operations)</td>
</tr>
<tr>
<td>$n$-back</td>
<td>Items presented sequentially; subject is to make decision about whether current item matches the stimulus $n$ items ago</td>
</tr>
<tr>
<td>Running span</td>
<td>Items presented sequentially, and typically subject does not know in advance how many items will be presented each trial; subject is instructed to recall the final $n$ items, where $n$ is typically less than the number of items presented</td>
</tr>
</tbody>
</table>

The working memory tasks used in training research vary in stimuli (e.g., letters, digits, objects, locations) and structure (e.g., span tasks requiring serial recall of stimuli, $n$-back tasks requiring decisions about whether the current item matches the one presented $n$ items back; see Table 1). While there is substantial variation in the details of any particular working memory training study, the design of most working memory training experiments is simple. Subjects are assigned to training or control groups, and those in the training group perform multiple sessions of computerized working memory training task(s) across different days. Before and after the training intervention is conducted, all subjects complete measure(s) to assess transfer; the critical question to answer is whether or not the pretest-posttest change in the transfer tasks differs as a function of the group assignment. Researchers have often tried to distinguish between near and far transfer (Barnett & Ceci, 2002), although this is not always consistent across studies. Conceptually, near transfer occurs when the training group has improved more than the control
group on tasks in the pretest-posttest sessions that are similar in terms of methodology or content to the training materials. The goal of most working memory studies, however, is to demonstrate far transfer, which occurs when the training group shows larger gains than the control group on outcomes that are dissimilar to the training procedures.

I argue that the working memory training literature has followed what’s been labeled in other contexts, such as new technological innovations, as the ‘hype-cycle’ (‘Gartner Hype Cycle’; see also Baker, Ware, Schweitzer, & Risko, 2017; Linden & Fenn, 2003). A hype cycle occurs when a new technology enters a field (the ‘Innovation Trigger’), and expectations rapidly increase for what the technology may be able to do (the ‘Peak of Inflated Expectations’). Typically, this initial excitement is followed by a rapid decrease in expectations (the ‘Trough of Disillusionment’). Some technologies, through additional investment and research (the ‘Slope of Enlightenment’) eventually are deemed commercially viable and useful innovations (the ‘Plateau of Productivity’), although at levels much more modest than the initial peak. While some technologies make the progression to the ‘Plateau of Productivity’ and are ultimately widely adopted (e.g., cloud computing), others never escape the ‘Trough of Disillusionment’ (e.g., broadband over power lines).

The hype cycle pattern seems to apply to working memory training progress. Initial research asserted that, after about 15 hours of practicing computerized working memory tasks, working memory training could increase fluid intelligence (Klingberg, Forssberg, & Westerberg, 2002; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008) and reading comprehension (Chein & Morrison, 2010), reduce ADHD symptoms in children (Klingberg et al., 2005) and alcohol consumption in heavy drinkers (Houben et al., 2011), and alter the neurophysiology of frontal lobe connectivity (Dahlin et al., 2008). These highly-cited articles influenced the development
and commercialization of numerous working memory training programs, and spurred further working memory training research in virtually every area of psychology (for review, see Redick, 2015; Redick, Shipstead, Wiemers, Melby-Lervåg, & Hulme, 2015). However, after the initial flurry of positive studies, a second wave of largely negative, challenging research that failed to demonstrate far transfer emerged (Chooi & Thompson, 2012; Harrison et al., 2013; Redick et al., 2013; Sprenger et al., 2013; Thompson et al., 2013), some of which highlighted methodological limitations of the initial studies on the topic (Shipstead, Redick, & Engle, 2012). Currently, working memory training seems to be entering a third wave, in which meta-analyses, which are quantitative reviews that combine the results of multiple studies, have indicated working memory training effects are highly specific to relatively narrow outcomes (Melby-Lervåg, Redick, & Hulme, 2016) and/or to certain training programs (Au et al., 2015). The bulk of the evidence indicates working memory training is not effective in the ways claimed by the initial research.

Unfortunately, the strongest claims that working memory training “works” often come from studies with the most problematic design and data issues (Table 2). These include problems such as small sample sizes, passive control groups, subjective and/or unblinded rating measures, and hypothesis-inconsistent pretest and posttest patterns (Dougherty et al., 2015; Melby-Lervåg et al., 2016; Rapport et al., 2013; Redick, 2015; Redick et al., 2015; Shipstead et al., 2012; Simons et al., 2016). For example, Melby-Lervåg et al. (2016) reviewed published and unpublished working memory training studies, looking at effects on a variety of cognitive and academic outcomes. Overall, similar to other recent meta-analyses (Sala & Gobet, 2017; Soveri, Antfolk, Karlsson, Salo, & Laine, 2017), we found that training produced benefits to working memory tasks similar to the training materials, but not to other outcomes (Figure 1). Specifically,
when looking at whether working memory training improves fluid intelligence, or the ability to
do novel reasoning often with nonverbal materials, Melby-Lervåg et al. (2016) examined 108
results as a function of the study’s sample size (at least 20 subjects per group vs. fewer subjects)
and type of control group (passive vs. active). We found that the studies with small sample sizes
and passive control groups produced an average effect size of $g = .33$ – that is, based on these
methodologically weaker studies, one could conclude that training improved fluid intelligence by
$1/3$ of a standard deviation. However, in the more methodologically sound studies with larger
sample sizes and active control groups, the average effect size was $g = .01$ – no effect
whatsoever.
Table 2

Limitations in Working Memory Training Studies

<table>
<thead>
<tr>
<th>Issue</th>
<th>Consideration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td>If a control group is passive and does not do any study-related activity between pretest and posttest, there is the potential for differential outcomes based on expectancies, demand characteristics, placebo effects, etcetera that may be less likely (but not necessarily eliminated) when using an active control group (Boot et al., 2013)</td>
</tr>
<tr>
<td>Sample size</td>
<td>Small sample sizes per group can inflate effect sizes and produce noisier data patterns (Button et al., 2013)</td>
</tr>
<tr>
<td>Atypical data patterns</td>
<td>Significant ($p &lt; .05$) results can occur but be inconsistent with the predicted results, if for example both the training and control group show a pretest-posttest decrease in IQ scores but the decrease is significantly larger in the control group (Redick, 2015)</td>
</tr>
<tr>
<td>Single outcome measures</td>
<td>Ideally, cognitive abilities as broad as working memory, fluid intelligence, and executive function would be measured by more than one task, because the outcome scores reflect both construct-relevant and irrelevant variance; multiple indicators of the construct help with this issue (Shipstead et al., 2012)</td>
</tr>
<tr>
<td>Subjective outcomes</td>
<td>Subjective ratings, particularly self- or other-reports completed by individuals who know whether the target of the ratings completed</td>
</tr>
</tbody>
</table>
the training program or not, are prone to biases (Rapport et al., 2013).

![Figure 1. Effect sizes from Melby-Lervåg et al. (2016) meta-analysis comparing various forms of working memory training versus active control groups. Outcomes shown in blue reflect pretest/posttest measures that are similar to that training materials in methodology and/or stimulus type (near transfer), while outcomes shown in red reflect measures that are different from training content (far transfer). Error bars indicate 95% confidence intervals. Numeric values under each outcome represent the number of different comparisons included in the calculation of that outcome’s effect size.](image-url)
Individual differences

Although the conclusion about the current state of working memory training from Melby-Lervåg et al. (2016) is that it is not an effective intervention, a contrarian view might argue that lumping many different studies together is not an ideal method to isolate the specific combination of factors that make working memory training work. That is, perhaps there are moderators based on individual differences in cognitive abilities and personality traits that would maximize the efficacy of working memory training (e.g., Jaeggi, Buschkuehl, Jonides, & Shah, 2011; Jaeggi, Buschkuehl, Shah, & Jonides, 2014; Katz, Jones, Shah, Buschkuehl, & Jaeggi, 2016). Multiple studies (e.g., Guye, De Simoni, & von Bastian, 2017; Jaeggi et al., 2014; Sprenger et al., 2013; Thompson et al., 2013) have examined the potential role of variables such as grit, implicit theory of intelligence (i.e., fixed vs. growth mindset) and need for cognition, but those studies have found no evidence that such variables influence transfer after working memory training (although most studies had rather small samples to detect such moderated effects). Examining cognitive ability, recent research with my colleagues has shown that individuals that are higher in working memory at pretest show the largest improvements on the training tasks (Foster et al., 2017; Wiemers, Redick, & Morrison, 2018). For example, Foster et al. (2017) pre-screened a large number of subjects for their working memory functioning, and then created separate training and control groups of high- and low-working memory subjects. Foster et al. found that high-working memory subjects, compared to low-working memory subjects, showed higher performance on day one of training and larger improvements on the training tasks from the first to the last day. Critically, only limited near transfer was observed, with no evidence of far transfer to fluid intelligence for either high- or low-working memory subjects.
Some researchers have proposed a dose-dependent relationship between training and transfer (Jaeggi et al., 2008; 2011), such that those who receive more training and/or improve more on the training tasks should also be the subjects that show the largest transfer gains. Thus, our findings (Foster et al., 2017; Wiemers et al., 2018) that high-working memory subjects gained more on the training tasks than the low-working memory subjects might be troubling if the goal of training is to help those with working memory impairments. However, the evidence linking training gain and transfer task improvement is weak. Some studies have reported a positive relationship between training gain and transfer gain (e.g., Chein & Morrison, 2010; Jaeggi et al., 2011). However, Tidwell et al. (2014) demonstrated that correlating training gain scores and transfer gain scores (or similar responder analyses) are uninformative and should be ignored if reported. In addition, many of the individual training studies examining the relationship between training and transfer using any statistical approach suffer from exceedingly small sample sizes for such correlational analyses. At the meta-analytic level, Au et al. (2015) found no relationship between the rate of training improvement and the amount of change in intelligence scores from pretest to posttest.

An intriguing finding worthy of further study is that the spacing of the training sessions might be relevant for the amount of transfer produced (von Bastian & Oberauer, 2014). For example, Wang, Zhou, and Shah (2014) tested four separate working memory training groups, each of which completed 20 training sessions, versus a control group. The critical manipulation was the spacing of the training sessions – the four different groups completed between 1-10 sessions per day. Although the amount of training task improvement did not vary, Wang et al. argued that transfer to fluid intelligence was directly related to the training schedule – subjects with spaced sessions showed more pretest-posttest improvement than subjects with massed
sessions. Replicating this finding with larger samples and an active-control group comparison would be an important step in addressing the potential role of spacing and scheduling of training sessions.

**Expectations**

As mentioned above, one reason for using active-control groups and using objective performance outcomes and/or blinded subjective reports is the difficult role played by subject expectations in intervention studies, including working memory training research. A survey asking naïve subjects their beliefs about computerized cognitive training found that the majority of subjects thought they would improve their cognitive functioning (Rabipour & Davidson, 2015). In addition, subjects often are recruited from informational sessions or fliers that specify the purpose of the study is to improve cognition. Foroughi, Monfort, Paczynski, McKnight, and Greenwood (2016) tested subjects using two different types of recruitment materials that either emphasized that brain training would enhance cognition or included more neutral language that did not mention brain training at all. Foroughi et al. (2016) found that transfer to fluid intelligence was larger for the subjects recruited using the brain training advertisements compared to the neutral fliers. However, the results of Foroughi et al. (2016) are weakened by their use of a small sample size and administration of only one session of n-back training.

More recently, Tsai et al. (2018) used a more typical multi-session design, comparing n-back training versus a trivia active-control condition. Critically, they manipulated subjects’ expectations for transfer to only a visual n-back task identical to the training task, or to both visual and auditory tasks. Tsai et al. (2018) found that for transfer from n-back training to n-back tasks, the expectation manipulation did not have an effect. That is, regardless of the expectation instructions that subjects received, the n-back training group showed transfer but the control
The results of Tsai et al. (2018) are a good first step in experimentally addressing expectations in working memory training research, but it does not address far transfer (e.g., fluid intelligence, reading comprehension, ADHD symptoms). Indeed, in the meta-analysis by Melby-Lervåg et al. (2016), the nonverbal ability far transfer effect size was significantly larger for passive-control groups versus active-control groups, but not for working memory near transfer. As suggested by Foroughi et al. (2016), one conclusion is that covert recruitment and double-blind research designs should become the norm in future research, instead of the outlier, if the field is truly interested in determining the possible efficacy of working memory training in the absence of placebo/expectancy effects (Green et al., in press; Simons et al., 2016).

**Current and Future Directions**

The review thus far portrays a rather pessimistic view of the working memory training literature. Seventeen years after publication of the first computerized working memory training study, there is substantial evidence that practicing working memory tasks over multiple days improves performance on those and similar tasks, but not to other untrained tasks and materials (Melby-Lervåg et al., 2016). Some researchers have now attempted to combine working memory training with other cognitive enhancement techniques, such as transcranial direct current stimulation (tDCS), in a kitchen-sink approach to maximize the possibility of increasing cognition. However, the early returns on this approach are not promising, as a meta-analysis of the small number of published tDCS+training studies to-date yielded no benefits of tDCS and only near transfer (Nilsson, Lebedev, Rydström, & Lövdén, 2017).

An alternative view is that it is still too soon to judge the fate of working memory training. Returning to the hype cycle metaphor that I mentioned previously, perhaps the early studies reflect the ‘Peak of Inflated Expectations’, and the current status of the field reflects the
section of the cycle that is categorized as the ‘Trough of Disillusionment’. That is, despite the numerous studies conducted and commercialized products already available, perhaps it is still possible to optimize working memory training, by either tailoring it to the individual user, and/or to focus on more realistic near transfer goals instead of increasing intelligence or treating ADHD. This is the view held by someone who thinks working memory training can proceed to the ‘Plateau of Productivity’. While possible, previous research indicates that it is unlikely that working memory training programs currently in use will stand the scrutiny of larger sample sizes, procedures that control for expectations, and transfer to daily activities instead of the controlled nature of in-lab experiments to produce the large far-transfer effects seen in the initial studies in the field.

Ultimately, researchers that continue to conduct working memory training research and develop commercial products based on the research still need to gain a deeper understanding of what actually changes with training. That is, theory-based approaches explaining the mechanism that underlies working memory and far transfer are still lacking. Instead, many studies use a metaphor that appeals to the (incorrect) idea that the brain is a muscle, or that transfer between working memory and fluid intelligence occurs because there is functional overlap in some of the brain areas involved in performance of both types of tasks. The logic of many studies seems to be that working memory training would work by increasing the number of discrete representations one can hold at a time within working memory. But, is it really accurate to state that someone who performs a $n$-back task at level 2 on day one of training is simply engaging more of the identical cognitive processes when they perform $n$-back at level 8 on day eight of training? My colleagues and I tried to determine this using self-report surveys administered at posttest, which suggested subjects engage in multiple strategies across different working memory
training sessions (Redick et al., 2013). Albeit limited evidence, the use of different strategies depending on the current task demands is entirely consistent with the idea that the working memory system permits one to allocate attention as needed (Logie, 2012), and seemingly inconsistent with a dose-dependent relationship between training and transfer. Thus, if working memory training research is to move beyond the hype cycle, theory-driven research grounded in solid methodology is critical for providing stronger evidence for its efficacy.
Acknowledgment

While working on this manuscript, TSR was supported by the National Institutes of Health (Award # 2R01AA013650-11A1) and National Science Foundation (Award # 1632403 and # 1839971).
References


**Recommended Reading**


Chapter that discusses designing working memory training programs and characteristics of the trainer such as personality and motivation.

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). (See References). A meta-analysis that both summarizes the working memory training literature quantitatively across a wide number of outcomes, types of training programs, and groups of people while also discussing factors responsible for contrasting results in the literature.
