Target article
Cogmed working memory training: Does the evidence support the claims?
Zach Shipstead, Kenny L. Hicks, Randall W. Engle ∗
School of Psychology, Georgia Institute of Technology, 654 Cherry Street, Atlanta, GA 30332, United States

Abstract
Cogmed working memory training is sold as a tool for improving cognitive abilities, such as attention and reasoning. At present, this program is marketed to schools as a means of improving underperforming students’ scholastic performance, and is also available at clinical practices as a treatment for ADHD. We review research conducted with Cogmed software and highlight several concerns regarding methodology and replicability of findings. We conclude that the claims made by Cogmed are largely unsubstantiated, and recommend that future research place greater emphasis on developing theoretically motivated accounts of working memory training.

© 2012 Society of Applied Research in Memory and Cognition. Published by Elsevier Inc. All rights reserved.
Table 1: Cogmed training studies.

<table>
<thead>
<tr>
<th>Population studied</th>
<th>Authors</th>
<th>Type of control n group</th>
<th>Working memory</th>
<th>Reasoning/IQ</th>
<th>Attn control</th>
<th>Sustained Attn</th>
<th>ADHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children (ADHD)</td>
<td>Beck et al. (2010)</td>
<td>No-contact</td>
<td>51</td>
<td></td>
<td></td>
<td>Parent ratings</td>
<td>Teacher ratings</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Free recall SM</td>
<td>Parent ratings</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Teacher ratings</td>
<td></td>
</tr>
<tr>
<td>Children (ADHD)</td>
<td>Gibson et al. (2010)</td>
<td>Visuo-spatial vs. verbal training</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children (ADHD)</td>
<td>Holmes et al. (2010)</td>
<td>None</td>
<td>25</td>
<td>SS digit</td>
<td>WASI verbal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SS dot matrix</td>
<td>WASI performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SS digit backward</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CS MR X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children (ADHD)</td>
<td>Klingberg et al. (2002)</td>
<td>Non-adaptive</td>
<td>14</td>
<td>SS visual forward</td>
<td>Raven Stroop</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Klingberg et al. (2005)</td>
<td>Non-adaptive</td>
<td>44</td>
<td>SS span board</td>
<td>Raven Stroop</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SS digit forward</td>
<td>Raven Stroop</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SS span board</td>
<td>Raven Stroop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children (ADHD)</td>
<td>Mezzacappa and Buckner (2010)</td>
<td>None</td>
<td>8</td>
<td>SS digit backward</td>
<td>SS spatial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children (cochlear implants)</td>
<td>Kronenberger, Pisoni, Henning, Colson, and Haazard (2011)</td>
<td>None</td>
<td>9</td>
<td>SS digit forward</td>
<td>SS digit forward</td>
<td>SS spatial</td>
<td>Raven Stroop</td>
</tr>
<tr>
<td>Children (low birth weight)</td>
<td>Holmes et al. (2009)</td>
<td>Typically developing</td>
<td>30</td>
<td>SS verbal</td>
<td>SS spatial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children (low WMC)</td>
<td>Holmes et al. (2009)</td>
<td>Non-adaptive</td>
<td>42</td>
<td>SS verbal composite</td>
<td>WASI verbal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SS spatial composite</td>
<td>WASI performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CS counting recall</td>
<td>WORD WOND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children (SEBT)</td>
<td>Roughan and Hadwin (2011)</td>
<td>No-contact</td>
<td>17</td>
<td>SS digit forward</td>
<td>Raven</td>
<td>Go/no go</td>
<td></td>
</tr>
<tr>
<td>Children (special education)</td>
<td>Dahlin (2011)</td>
<td>Control group</td>
<td>57</td>
<td>SS digit backward</td>
<td>SS span board</td>
<td>Raven Stroop</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>from Klingberg et al. (2005)</td>
<td></td>
<td>SS span board</td>
<td>Raven Stroop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children (Typically developing)</td>
<td>Bergman Nutley et al. (2011)</td>
<td>Non-adaptive</td>
<td>101</td>
<td>SS visual forward</td>
<td>SS odd one out</td>
<td>Raven – 3 sets</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SS word span</td>
<td></td>
<td>Block design</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SS digit span</td>
<td></td>
<td>Raven</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SS span board</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SS operation span</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CS word span</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children (typically developing)</td>
<td>Shavelson et al. (2008)</td>
<td>Non-adaptive</td>
<td>37</td>
<td>SS word span</td>
<td>Block design</td>
<td>Stroop-like CPT</td>
<td></td>
</tr>
<tr>
<td>Older adults</td>
<td>Thorrell et al. (2009)</td>
<td>Computer games</td>
<td>62</td>
<td>SS span board</td>
<td>Raven</td>
<td>Go/no go</td>
<td>PASAT</td>
</tr>
<tr>
<td></td>
<td>Brehmer et al. (2012)</td>
<td>Non-adaptive</td>
<td>45</td>
<td>SS digit</td>
<td>Raven Stroop</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SS span board</td>
<td>Raven Stroop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young adults</td>
<td>Brehmer et al. (2012)</td>
<td>Non-adaptive</td>
<td>55</td>
<td>SS digit</td>
<td>Raven Stroop</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SS span board</td>
<td>Raven Stroop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young adults</td>
<td>Klingberg et al. (2002)</td>
<td>Children with/ADHD</td>
<td>4</td>
<td>SS visual forward</td>
<td>Raven Stroop</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>from Exp. 1</td>
<td></td>
<td>SS span board</td>
<td>Raven Stroop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young adults</td>
<td>McNab et al. (2009)</td>
<td>None</td>
<td>13</td>
<td>SS digits backward</td>
<td>SS syllables forward</td>
<td>Ravens – DNR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SS visual forward</td>
<td>SS simple span</td>
<td></td>
<td>PASAT – DNR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SS span board</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young adults</td>
<td>Olesen et al. (2004) Exp 1</td>
<td>No-contact</td>
<td>3c</td>
<td>SS span board</td>
<td>Raven Stroop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroke patients</td>
<td>Westerberg et al. (2007)</td>
<td>No-contact</td>
<td>18</td>
<td>SS digit span</td>
<td>Raven Stroop</td>
<td></td>
<td>PASAST</td>
</tr>
</tbody>
</table>

Note: Bold indicates that the original paper reports significant transfer of this task. When no mention is made of forward/backward this either indicates composite score, or both forward and backward tests were significant; n, number of participants included in posttest; Attn, attention; SS, simple span; CS, complex span; DNR, did not report; SM, secondary memory; CPT, continuous performance task; CRT, choice reaction time.

*a* Relative to baseline scores.

*b* Control group’s performance declined.

*c* Trained group only.

*d* Not reported as a measure of ADHD, but included in this column based on separate research.

As highlighted in the opening quotations, the critical information is not the volume of research, but the findings. Although some studies have produced promising results (in particular, Klingberg et al., 2002, 2005), we contend that the overall picture is bleak. In short, many claims made by Cogmed are based on findings that have not replicated, are not readily attributable to increased WM capacity, or simply have not been thoroughly studied. Moreover, the claim that Cogmed actually increases WM capacity has yet to receive careful examination. However, before these issues can be meaningfully explored, we must first develop...
Fig. 1. Examples of complex and simple span tasks. The (a) reading span task requires test takers to read each sentence, then judge whether or not it makes sense. After several sentences have been read, the test taker is signaled to remember the last word of each sentence, in proper serial position. The (b) word span task, on the other hand, simply presents a short list of words that the test taker must recall in proper serial order.

<table>
<thead>
<tr>
<th>Complex Span</th>
<th>Simple Span</th>
</tr>
</thead>
<tbody>
<tr>
<td>The dog ran over the tractor (y/n)</td>
<td>tractor</td>
</tr>
<tr>
<td>The man in the suit walked to work (y/n)</td>
<td>work</td>
</tr>
<tr>
<td>Monkeys enjoy Christmas vacation (y/n)</td>
<td>vacation</td>
</tr>
</tbody>
</table>

Answer: tractor, work, vacation  

| tractor, work, vacation |

an understanding of what WM is, and how WM capacity is measured.

1. Working memory

There is no consensus definition of WM (Miyake & Shah, 1999), however most researchers would agree that WM is not simple memory over the short term. Rather, WM is a dynamic memory system that is critical in environments in which attention must constantly shift between sources of information. A person's WM capacity therefore reflects several cognitive mechanisms, such as active maintenance and updating of specific goals and information, as well as retrieval of critical information following distraction (e.g., Cowan, 2001; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000; Unsworth & Engle, 2007b).

The first test to reliably measure individual differences in WM capacity was Daneman and Carpenter’s (1980) “reading span” (Fig. 1a). This complex span task requires people to read a sentence and judge whether or not it makes sense. After several sentences have been read, the test taker must recall the last word of each, in serial order.

Accuracy on complex span tasks predicts individual differences in reading comprehension (Daneman & Carpenter, 1980; Turner & Engle, 1989), SAT performance (Turner & Engle, 1989), the ability to solve novel problems (referred to as general fluid intelligence or Gf; Engle, Tuholski, Laughlin, & Conway, 1999; Kane, Hambrick, & Conway, 2005). What is particularly important to the makers of WM training software is that individual differences in WM capacity also predict real-word performance in areas such as multi-tasking (Buhner, König, Prick, & Krumm, 2006; Hambrick, Oswald, Darowski, Rensch, & Brou, 2010), emotion regulation (Kleider, Parrott, & King, 2009; Schmeichel, Volokhov, & Demaree, 2008), and mind-wandering (Kane, Brown, et al., 2007). Children with low WM capacity have relatively poor verbal and mathematical abilities (Cowan et al., 2005; Gathercole & Pickering, 2000; Swanson & Beebe-Frankenberger, 2004) and have difficulty following directions (Engle, Carullo, & Collins, 1991; Gathercole, Durling, Evans, Jeffercock, & Stone, 2008). Moreover, low WM capacity is associated with ADHD (Willcut, Doyle, Nigg, Faranea, & Pennington, 2005), particularly inattentive symptoms (Diamond, 2005). The implication for training is that if performance in all of these domains is somehow limited by WM capacity, then a program that increases WM capacity should result in improvements in all of these areas.

The complex span task contrasts with the more basic simple span task (Fig. 1b), in which to-be-remembered items are presented without interruption. Although complex and simple span tasks require many common cognitive processes (Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008; Unsworth & Engle, 2007a), the simple span is a less reliable measure of complex cognition, and has a tenuous history as a WM task (Daneman & Carpenter, 1980; Engle & Oransky, 1999; Perfetti & Lesgold, 1977; Turner & Engle, 1989). This is likely due to decreased demands on attention control (Barrouillet, Bernardin, & Camos, 2004; Engle, 2002), as well as a decreased need to retrieve information from long term memory (Unsworth & Engle, 2007a). Additionally, because they do not have an interpolated processing task, simple spans afford test takers greater opportunity to strategically chunk to-be-remembered information into memorable units (Cowan et al., 2005; see also, Chase & Ericsson, 1982).

Regardless of the exact cognitive processes that are accentuated by the interpolated task, it is well established that simple span tasks provide relatively poor reflections of intellectual ability (Ackerman, Beier, & Boyle, 2005; Daneman & Merikle, 1996; Engle et al., 1999; Turner & Engle, 1989). As such, their use as WM tasks should generally be avoided (but see Unsworth & Engle, 2007a for exceptions). This is an issue to which we will return.

2. Cogmed training

There are three types of Cogmed training, each of which is performed for roughly 40 min a day, 5 days a week, for 5 weeks. Cogmed JM and RM are respectively designed for preschoolers and older children. These programs include several visuo-spatial and verbal memory tasks that have been embedded in videogames. An example of a visuo-spatial task is “Asteroids”. This task presents a field of several free-floating asteroids, a subset of which lights up, one-at-a-time. The child then reproduces the sequence via mouse-click. An example of a verbal task is “Input Module”. In this task a sequence of auditory digits is played. The child then reproduces the sequence in reverse order, using a number-pad that is displayed on a robot’s arm.

Adults perform Cogmed QM. These tasks are conceptually similar to JM and RM, with the exception that information is presented in relatively simple displays (e.g., rotating grids, rather than asteroids). The ostensible key to all of these programs is that they adapt to user performance. If the trainee is doing well, the to-be-remembered list will increase by one item. If the trainee is struggling, the to-be-remembered list will decrease by one item. Through this method of always having trainees perform at the limit of their ability, it is assumed that WM capacity will grow (Klingberg, 2010).

2.1. Transfer of training

During training, people typically advance through several levels of the task. This phenomenon, in and of itself, does not provide evidence that WM capacity has increased. Improvements on the training task are directly confounded with task-specific practice (Chase & Ericsson, 1982). Thus, researchers must show that training “transfers” to the performance of untrained WM tasks.

These untrained tasks are conceptually included in a pretest–posttest battery, and improvements must be shown relative to a control group. Several types of control groups are present across Cogmed studies (see Table 1). A basic no-contact control group only performs the pre- and posttest. This type of group controls for the effects of repeated testing. Many Cogmed studies employ control groups that train on non-adaptive tasks, which never present lists longer than 2–3 items. Thus the
control group is active within the training environment, but never exposed to the critical manipulation. The intent is to further control placebo-type effects (Klingberg, 2010; Shipstead, Redick, & Engle, 2010; Shipstead, Harrison, & Engle, 2012; Shipstead, Redick, & Engle, 2012).\(^1\)

Demonstration of near transfer of training to untrained WM tasks is of the utmost importance: WM is the presumed mechanism of change through which broad cognitive benefits are realized. In the absence of evidence of increased WM capacity, it is unclear why training should benefit other areas of peoples’ lives.

Improved scores on untrained WM tasks are therefore critical to the interpretability of training studies; however, it is the potential for broad mental changes that make these studies meaningful. Thus, transfer batteries typically include several tests of reasoning, attention, or ADHD-related symptoms. These signs of far transfer of training form the basis of claims such as those discussed in the introduction.

**Does Cogmed improve reasoning ability?** Cogmed’s claims regarding improved reasoning are subdued, relative to those of Mindsparke (2011), Lumosity (2011), and Jungle Memory (2011), all of whom claim to increase IQ or school grades dramatically. Regardless, the second set of the opening quotes (i.e., Cogmed, 2011f) directly states that Cogmed will improve reasoning ability, and several relevant studies have been conducted. However, before exploring the results of Cogmed studies, it is important to understand the challenges involved in measuring change to mental abilities.

Oftentimes when researchers want to measure a person’s reasoning ability, they will use Raven’s Progressive Matrices (Raven, 1990). This problem-solving task presents the test-taker with a series of 8 abstract objects arranged within a 3 × 3 matrix. The ninth space is left blank and the test-taker must choose which of several options completes the sequence. Jensen (1998) estimates that 64% of the variance in Ravens performance can be accounted for by general intelligence (e.g., problem solving skills that permeates all aspects of reasoning), placing Ravens among the most valid measures of reasoning ability.

While 64% is impressive, the converse is that 36% of variation in Ravens performance is attributable to factors other than intelligence. For instance, because of their spatial-arrangement, matrix-selling tasks (such as Ravens) are biased in favor of individuals with high visuo-spatial short term memory (Kane et al., 2004). Thus, if a training program simply improves visuo-spatial memory, it may lead to higher scores on matrix-selling tasks (Moody, 2009). However, these performance improvements would not necessarily qualify as an increase in general reasoning ability, since they might be absent if non-spatial test materials were used (e.g., verbal reasoning tests). In short, the source of a training effect cannot be pinpointed using a single task.

It is therefore critical that researchers include a variety of tests within a transfer battery. For instance, if the goal is to increase fluid intelligence, then the battery might include Ravens, along with a test that requires mental manipulation (e.g., PaperFolding; Ekstrom, French, Harman, & Dermen, 1976), as well as a test that involves reasoning with verbal material (Letter Sets; Ekstrom et al., 1976). When trainees show significant post-test improvements across these tasks, it can be argued strongly that a general ability has improved. However, when pretest–posttest differences are found using single indicators, it is not clear whether the improved performance reflects change to an underlying ability, or change to peripheral factors such as the development of modality-specific training (Moody, 2009).

Studies: Klingberg, Forssberg, and Westerberg (2002) provided the initial test of the Cogmed training paradigm. Following 24 sessions of WM training, 7 ADHD-diagnosed children improved their performance on Raven’s Coloured Progressive Matrices (Raven, 1995; a version of Ravens designed for children), relative to a non-adaptive control group. This finding replicated in a double blind study that involved a larger sample (44 children overall).

Although these results were encouraging, a series of failed replications followed. In studies respectively involving children with ADHD and low WM capacity, Holmes et al. (2010) and Holmes, Gathercole, and Dunning (2009) found no evidence of immediate transfer to a battery of novel reasoning and academic achievement tests (i.e., Wechsler Abbreviated Scales of Intelligence; Wechsler, 1999; Wechsler Objective Number Dimensions; Wechsler, 1996; Wechsler Objective Reading Dimensions; Wechsler, 1993). Additionally, Dahlin (2011) found no evidence of improved Ravens performance in a sample of children who were enrolled in special education classes (though she found some evidence of improved reading comprehension). One apparent exception was the study of Roughan and Hadwin (2011), which reported that children with social, emotional, and behavioral difficulties improved their performance on a composite score of Ravens and vocabulary, relative to a no-contact control group. However, while the researchers report a large effect of training (n = .44), the control group showed pretest–posttest declines on these tasks that were of a numerically greater magnitude than improvements shown by the trained group. Thus, the meaningfulness of this finding is questionable.

Turning to typically developing preschoolers, Thorell, Lindqvist, Bergman, Bohlin, and Klingberg (2009) found no evidence of improvement on a block-design task (i.e., recreate a pattern with blocks; WPPSI; Wechsler, 1995), and Bergman Nutley et al. (2011) reported null effects using three versions of Ravens and a block-design task. Similarly, Shavelson, Yuan, and Alonso (2008) found no evidence of transfer to Ravens with a sample of typical middle school children.

Studies with adults have shown even less promise. Although Klingberg et al. (2002) and Olesen, Westerberg, and Klingberg (2004) report training-related improvements on Ravens performance, the generalizability of these studies is limited by small sample sizes (respectively, 4 and 3 participants were trained) as well as inappropriate control groups (respectively, ADHD-diagnosed children, and no-contact). McNab et al. (2009) tested 13 trained participants on Ravens but chose not to report the data (as reported in the online supplemental material). Finally, a study by Brehmer, Westerberg, and Backman (2012) found no evidence of transfer to Ravens for younger (20–30 years) or older (60–70 years) adults.\(^3\)

**Summary:** Although initial studies indicated that Cogmed training might improve reasoning abilities, subsequent studies have provided a series of failed replications. It is clear that this program does not improve reasoning skills. Thus, while Cogmed’s claims in this area are relatively subdued,\(^3\) any mention of improved reasoning ability should be viewed as false.

**Does Cogmed train attention?** One hypothesis for why WM is critical to many areas of life is that WM capacity is essentially a reflection of a person’s executive attention (Kane, Conway, Hambrick, & Engle, 2007). This is a relationship upon which Cogmed is clearly reliant (Cogmed, 2011d, 2011f). However, WM is not related to all aspects of attention. For instance, individual differences in WM capacity do not predict a person’s ability to perform a

---

\(^1\) These groups, however, may not control for the effect of being adaptively trained. This is an issue that we will raise in the Discussion section.

\(^2\) A subsets of these results is reported in Brehmer et al. (2011).

\(^3\) For contrast, see Mindsparke’s (2011) website which implies that completing their program may result in membership in MENSA, a society for individuals with IQ in the 98th percentile.
visual search task (Kane et al., 2006), nor do they predict the amount of information a person can pre-attentively subitize (Tuholski, Engle, & Baylis, 2001). Rather, individual differences in WM capacity are most apparent when sustained, self-initiated, control (Kane, Blecley, Conway, & Engle, 2001; Kane & Engle, 2003; Unsworth, Schrock, & Engle, 2004) or focus (Conway, Cowan, & Bunting, 2001; Heitz & Engle, 2007; Poole & Kane, 2009; Shipstead, Harrison, et al., 2012) is needed.

A relevant example involves the relationship between individual differences in WM capacity and performance on the Stroop task (1935). The Stroop task (Fig. 2) requires test-takers to rapidly and accurately name the hue in which a word is written. This task is easy when the hue and word are congruent (Fig. 2a), but relatively difficult when the word and hue are incongruent (Fig. 2b). The increase in reaction time and decrease in accuracy that occurs when people encounter incongruent stimuli is known as the “Stroop effect.”

Surprisingly, WM capacity does not predict performance on Stroop tasks that are entirely composed of incongruent items (e.g., Fig. 2b; Hutchison, 2007; Kane & Engle, 2003). It is only when congruent trials are introduced to the task (Fig. 2c) that low WM capacity individuals become relatively slow and inaccurate in their responding to incongruent trials (Hutchison, 2007; Kane & Engle, 2003).

Kane and Engle (2003) interpreted this trend as a sign that when a prepotent response (i.e., word reading) is always unsupportive of a goal (i.e., hue stating; e.g., Fig. 2b), the constant conflict reinforces the appropriate behavior. However, as conflict decreases (e.g., Fig. 2c), low WM capacity individuals have greater difficulty keeping the appropriate goal in mind and begin to use word-information in their responding. Thus, while WM-related attention may reflect some type of executive control mechanism, individual differences in WM capacity will not be apparent in all attention tasks.

Studies: Some studies have found an association between Cogmed training and improvement on Stroop tasks (Klingberg et al., 2005, 2002; Olesen et al., 2004), whereas others have not (Brehmer et al., 2012; Dahlin, 2011; Westerberg et al., 2007). However, in reference to the Stroop tasks that were used in several of these studies (i.e., Klingberg et al., 2005, 2002; Olesen et al., 2004), Klingberg (2010) states that “control congruent trials were not included . . . (p. 319)”. This is problematic: If performance on 100% incongruent Stroop tasks (e.g., Fig. 2b) is unrelated to WM capacity (Hutchison, 2007; Kane & Engle, 2003), then increased WM capacity cannot readily explain transfer. In other words, these two performance-variables are independent: Changing one (WM capacity) should not affect the other (all-incongruent Stroop).

Alternatively, there is evidence that Cogmed training may improve attentional stamina or vigilance. Westerberg et al. (2007) and Brehmer et al. (2012) both report training-related improvements on the Paced Auditory Serial Attention Task (Gronwall, 1977), which requires test-takers to attend to a long series of digits and continually sum the most recently presented with the digit that was presented one-back. Likewise, Thorell et al. (2009) found training-related improvements on continuous performance and go/no-go tasks (attend to a series of items and make a specific response when a specific item is shown). In both cases, trained children omitted fewer responses, relative to baseline, and relative to control participants.

Summary: There is evidence that Cogmed training will improve “attentional stamina” (as claimed in the opening quotes). Whether this is related to increased WM capacity, or is a product of completing a month of training on an attention-demanding task (i.e., any training task will do) is unclear. On the other hand, the evidence that Cogmed training may improve attentional control is not only equivocal in terms of replication, but is based on tasks that do not even relate to WM. Thus, whereas this training program may increase the time that a person can apply attention to a specific task, there is no reasonable evidence to suggest it will improve attention as it relates to selecting appropriate information or controlling impulses.

Does Cogmed alleviate ADHD-related symptoms? Although the Cogmed website repeatedly references the relationship between ADHD and low WMC (Cogmed, 2011a, 2011f), specific claims regarding Cogmed’s efficacy are avoided. Instead, user testimonials imply that training will alleviate myriad issues that ostensibly stem from ADHD, such as hostility (Cogmed, 2011i) and poor performance in school (Cogmed, 2011b). However, such claims are either not supported, or not addressed, by the available research.

Objectives of change: The adaptive technique used by Cogmed showed early promise when Klingberg et al. (2002) found training-related reductions in the number of hyperactive head-movements made by a small sample (5 trained) of ADHD-diagnosed children (single-blind). This effect, however, did not replicate in a larger (18 trained) double-blind study (Klingberg et al., 2005).

In addition to recording head-movements, Klingberg et al. (2002) also included a choice reaction time task (i.e., press one of two buttons when a specific stimulus is presented). This task is meaningful since the response times of children with ADHD tend to be more variable on a trial-by-trial basis than those of typically developing children (Westerberg, Hirvikoski, Forssberg, & Klingberg, 2004). Klingberg and colleagues (Westerberg et al., 2004) interpret this variability as an indication that children with ADHD cannot allocate attention to a task consistently. It is thus significant to note that Klingberg et al. (2002) did not find evidence of training-related improvement to choice reaction time performance.

Gibson, Gondoli, Johnson, Steeger, Dobrzenski, et al. (2011) took a different approach to measuring the effect of Cogmed training on ADHD-related symptoms. Relative to typically developing children, children with ADHD are impaired in their ability to retrieve information from secondary, or long term, memory (Gibson, Gondoli, Flies, Dobrzenski, & Unsworth, 2010). The amount of information a child can maintain in primary memory (immediate or short term memory), on the other hand, is not symptomatic of ADHD. Using this criterion, Gibson, Gondoli, Johnson, Steeger, Dobrzenski, et al. (2011) measured the effect of Cogmed training on retrieval from both primary and secondary memory in a free recall task (see Tulving & Colotla, 1970). Although training did increase the amount of information ADHD-diagnosed children could maintain in primary memory, retrieval from secondary memory was unaffected. This indicates that the aspects of WM that are trained by Cogmed are not related to ADHD.
Subjective change: Several Cogmed studies have asked parents or teachers to rate changes to ADHD-related symptoms. Although this is intended to extend findings beyond the laboratory setting (Klingberg, 2010), the downside is that subjective reports can be wholly driven either by a participant’s expectations of what should happen or by what the participant believes the experimenter expects to happen (Orne, 1962, 1972). For instance, if people are told they are participating in a sensory deprivation study, the simple act of sitting in a normal office may lead them to report feeling restless and disoriented (Orne & Scheibe, 1964). More relevant to the present discussion, people are also apt to perceive intellectual change in themselves (Conway & Ross, 1984) or others (DeLoache et al., 2010) following training interventions, even when no objectively measurable differences are present.

It is therefore critical that trainees or raters remain blind to condition assignment. This will allow for equivalent expectations of outcome to be formed across conditions. However, in Cogmed studies, raters are not always blind to whether the children are in the training or control group, and a predictable pattern is present. When parents or teachers are aware that they are rating the symptoms of children who have received training, they tend to report improved symptoms (Beck, Hanson, Puffenberger, Benniger, & Benniger, 2010; Gibson, Gondoli, Johnson, Steeger, Dobrzenski, et al., 2011; Mezzacappa & Buckner, 2010). On the other hand, when raters are blind they tend not to perceive change (Beck et al., 2010; Klingberg et al., 2005). The exception to this trend was parents in Klingberg et al. (2005), who, despite the study’s double-blind design, reported training-specific improvements in ADHD symptoms. Given the singularity of this finding, it should be interpreted cautiously (particularly since teachers did not report change). For instance, Klingberg et al. (2005) do not specifically report the degree to which parents were privy to the children’s advancement through the training program. Thus, unlike the non-adaptive control condition, parents with children in the adaptive group may have interpreted improved performance on the training task as evidence of generalized improvement.

Summary: Any insinuation that Cogmed training alleviates the symptoms of ADHD is disconcerting. In terms of objective measures of change, the only relevant effect thus far reported was a reduction in hyperactive behaviors (Klingberg et al., 2002), but this did not replicate in a later study (Klingberg et al., 2005). Although many studies have reported improved symptoms using subjective measures of change, rarely have raters been blind to condition assignment. Thus, expectation-of-outcome provides a parsimonious account of most findings (cf. Conway & Ross, 1984; DeLoache et al., 2010; Greenwald, Spangenberg, Pratkanis, & Eskenazi, 1991; Orne & Scheibe, 1964).

Does Cogmed increase WM capacity? Finally, the most basic question is whether or not Cogmed training actually increases WM capacity. Although the literature is focused on “far” transfer effects (e.g., improved reasoning or attention), “near” transfer of training to WM capacity is the mechanism of change through which training effects are proposed to occur. In the absence of near transfer there is no clear reason why abilities should change.

Thus, a coherent explanation of training effects cannot be formed unless experimenters rigorously demonstrate that an intervention actually increases WM capacity. However, the majority of Cogmed studies measure changes to WM capacity using a limited variety of simple span tasks (e.g., Fig. 1b).4 Beyond concerns regarding the reliability of these tasks as measures of WM capacity (cf. Daneman & Carpenter, 1980; Turner & Engle, 1989), Cogmed training is based on simple span tasks (forward and backward recall). Thus, it is not overly surprising that trainees who have spent a month learning to perform several variations of simple span tasks almost always improve their performance on other versions.

A handful of studies have attempted to demonstrate transfer to WM capacity using complex span tasks (e.g., Fig. 1a). Holmes et al. (2009) reported transfer relative to a non-adaptive control group. Holmes et al. (2010) also report improved complex span performance. The latter study, however, did not include a control group and thus its results are confounded by repeated testing. In contrast to these studies, Bergman Nutley et al. (2011) found only marginal improvement, and Shavelson et al. (2008) found no evidence of near-transfer (both relative to non-adaptive control groups).

These results are less than encouraging. Moreover, while complex span tasks are better measures of WM capacity than simple span tasks, they nonetheless share a fair amount of task-specific overlap with the method of training. That is, in either task, participants see a short list and are required to recall all of the items in their proper serial position. Thus, a better test of improved WM capacity might be transfer to different categories of WM task. That is, if Cogmed training does, in fact, increase WM capacity, then transfer should be apparent in several types of WM tasks. Examples of other valid tasks include visual arrays (cf. Cowan et al., 2005), running memory span (cf. Broadway & Engle, 2010), keeping track (cf. Engle et al., 1999; Miyake et al., 2000) and free recall (cf. Unsworth & Engle, 2007b). To date, few studies have made such attempts (but see Gibson, Gondoli, Johnson, Steeger, Dobrzenski, et al., 2011; Westerberg et al., 2007).

The point is, whether “simple” or “complex”, span tasks are only one method of making a person’s WM capacity apparent, and performance on them should not be confused with WM capacity itself. WM capacity is an ability that unites performance on many types of memory tasks. Thus, before a company can claim that their product increases WM capacity, it should demonstrate near transfer in many types of task that are known to reflect WM capacity.

3. Discussion

Does Cogmed training enhance mental abilities? The only unequivocal statement that can be made is that Cogmed will improve performance on tasks that resemble Cogmed training. However, for people seeking increased intelligence, improved focus and attentional control, or relief from ADHD, current research suggests that this training program does not provide the desired result.

3.1. Will the free market decide?

The efficacy of Cogmed is not simply a scientific curiosity. Cogmed is a product that is actively marketed to school systems and to the parents of children with developmental disabilities. Due to the cost to tax payers and consumers,5 it is our opinion that demonstrating the validity of this program is of the utmost importance. Cogmed (or any other commercial training program) must be able to support its claims.

It has been repeatedly suggested to us that such concerns are mitigated by the free market. Products that work will remain in the market place, while those that do not will disappear over time. Setting aside concerns about the initial wave of consumers who have

---

4 Backward simple span, which is often used in Cogmed studies, does index WM capacity in children (St. Clair-Thompson, 2010). However, many Cogmed studies report averages of forward and backward recall. Moreover, Cogmed training provides extensive practice on backward span tasks.

5 The exact cost of Cogmed is difficult to obtain, since Pearson does not officially advertise their prices. Our online search of practitioners indicates that an individual consumer in the United States can expect to pay up to $1500 to complete the program (e.g., In Focus Health, 2011; Quesenbery, 2011; Stepping Stones, 2011) and this was consistent with a quote that was obtained from a Cogmed representative.
been enticed to purchase an expensive and ineffective product, we counter that one of the basic principles of human cognition is a tendency to seek evidence that conforms to our expectations, and ignore anything that is disconfirmatory (confirmation bias). This tendency should only increase following an investment of time and resources (cognitive dissonance). Indeed, the continuing presence of magnet therapy, homeopathy, and perpetual motion machines (cf. Park, 2000) is a testament to the inability of the free market to self-correct when products do not live up to their initial hype.

It is thus important that consumers have access to appropriate information. The availability of detailed contrary opinions allows for informed purchasing decisions, rather than decisions which are based on heuristics such as appeal to authority (e.g., “hundreds of experts . . . are embracing working memory training”; Cogmed, 2011e) and the number of studies conducted (most of which cannot be accessed by non-academics). How can the buyer beware if the buyer does not have access to the appropriate information ahead of time?

3.2. Are non-adaptive control groups appropriate?

As previously mentioned, no-contact control groups do not eliminate the possibility that the trained and control groups are differentially motivated to perform at posttest. In order to deal with this confound, many Cogmed studies use non-adaptive training tasks (see Table 1) that are capped at lists of 2–3 items. On its surface, this type of control group appears reasonable. The tasks ensure that control participants are engaged within the training environment, but are never exposed to the critical manipulation. Therefore, at least two concerns associated with this type of control group. First, the validity of non-adaptive control groups rests upon the premise that repetitive performance of 2–3 item lists is sufficient to convince control participants that they are engaged in cognitive training. Second, the adaptive nature of the training tasks means that only the training group is given explicit feedback via increasing levels. That is, adaptive and non-adaptive groups differ in terms of (a) the subjective challenge posed by the training task and (b) being given tangible evidence that their cognitive abilities are improving. Similar to no-contact control groups, this creates concern that the training and control group do not approach the posttest with the same level of motivation to perform.

We argue that, because the groups differ in terms of explicit feedback regarding task-improvement, non-adaptive control groups cannot be expected to control placebo effects. In order to truly control training-related expectations, both groups should perform tasks in which difficulty explicitly changes in response to performance (both tasks level-up and level-down). Several examples can be found in the literature. For instance, Redick et al. (2012) developed an adaptive visual search task in which participants search for a specific character within a briefly presented display. The size of this display will increase or decrease, depending upon a trainee’s accuracy, Jaeggi, Buschkuehl, Jonides, and Shah, (2011), on the other hand, developed a knowledge-based trainer in which participants answer questions that increase or decrease in difficulty, in response to recent performance. Through these methods, the control group receives feedback regarding performance improvement and thus has equal opportunity to develop a belief that their abilities are improving.

3.3. Future directions

At present we are not convinced that Cogmed provides effective training of WM or associated abilities. Although we and others have expressed doubt about the general state of WM training (Melby-Lervåg & Hulme, 2012; Morrison & Chein, 2011; Redick et al., 2012; Shipstead, Redick, et al., 2012), we do not rule out the possibility that WM training could be made effective. The largest issue seems to be that, while there is logic to WM training (increase WM and improve related abilities), this literature is still struggling to find a theory. Specifically, it is important that research move beyond the desire to show that broad change can be realized through a month of training on a limited set of tasks.

Instead, we endorse a movement toward programs that are based on specific aspects of WM and transfer effects that should logically follow from such training. In particular we highlight the work of Gibson, Gondoli, Johnson, Steeger, and Morrissey (2011). As noted, these researchers have found evidence that Cogmed training specifically improves primary memory (Gibson et al., 2010), but not retrieval from secondary memory. On the basis of the empirical work of Unsworth and Engle (2007a, 2007b) which emphasizes the importance of retrieval from secondary memory, Gibson, Gondoli, Johnson, Steeger, and Morrissey (2011) have begun modifying Cogmed in an attempt to produce specific desired training effects (by creating greater need for retrieval). Under mounting evidence that Cogmed and other training programs have not lived up to the promise of early studies (Melby-Lervåg & Hulme, 2012; Redick et al., 2012) these endeavors represent a sensible course of action.

References


