Updating and working memory training: immediate improvement, long-term maintenance, and generalizability to non-trained tasks

Ang Su Yin[1], Kerry Lee[1]*, Fiona Cheam[2], Kenneth Poon[1], Juliana Koh[1]

[1]National Institute of Education, Nanyang Technological University, Singapore

[2]Ministry of Education, Singapore

A published version of this manuscript can be downloaded from i:10.1016/j.jarmac.2015.03.001.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

Acknowledgements

This work was supported by grants awarded to the corresponding author from the Office of Educational Research, National Institute of Education (OER17-11KL) and to Fiona Cheam from the Ministry of Education. The funding sources took no part in the design of the study or the interpretation of findings. Views expressed in this article do not necessarily reflect those of the National Institute of Education or the Ministry of Education. We thank the students who participated in the study and the school administrators who provided access and assistance. We also thank Rebecca Bull and Alethea Kerk for their invaluable contributions to this study.

*Queries concerning this article should be addressed to the corresponding author, Kerry Lee, National Institute of Education, 1 Nanyang Walk, Singapore 637616. Email: Kerry.Lee@nie.edu.sg. Telephone: +65 6219 3888, Fax: +65 6896 9845.

Abstract

Despite the popularity of working memory (WM) and updating training, recent reviews have questioned their efficacy. We evaluated a computer-based training programme based on the Running Span and Keep Track paradigms. We assigned 111 7-year-olds with poor WM and mathematical performances to updating training, one of two control groups, or a fourth group, who were administered Cogmed, a commercially available programme. At the immediate posttest, updating training produced only marginal improvements relative to control, but this was sustained and became significant six months post-training. Cogmed training resulted in substantial improvement at immediate posttest, but became marginal at delayed posttest. Neither type of training resulted in better performance in mathematics or generalised to other WM tasks that differed more markedly from those used during training. These findings suggest that relations between WM or updating capacity and mathematics performance may be moderated by factors that do not benefit directly from improved capacity.

Keywords: working memory; updating; cognitive training; math difficulties

Working memory and updating training: immediate improvement, long-term maintenance, and generalizability to non-trained tasks

Working memory (WM) and updating predict children's performances in reading (e.g., Gathercole & Pickering, 2000) and mathematics (e.g., Bull & Scerif, 2001; Lee, Pe, Ang, & Stankov, 2009; St Clair-Thompson & Gathercole, 2006). WM, defined as processes or structures that allow information to be maintained and manipulated simultaneously (Baddeley, 2000), is often deemed a fundamental capacity that affects how well other higher cognitive functions are performed. Updating refers to the ability to monitor and refresh information in WM (Miyake et al., 2000). Although not conceptually synonymous, measures of WM and updating are highly correlated (Schmiedek, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009; St Clair-Thompson, 2011). In addition to correlational findings, experimental studies have found that accuracy of mathematical task performance is dependent on the availability of WM resources (Fürst & Hitch, 2000; Lee & Ng, 2009). A question of continuing interest is whether academic performance can be improved by increasing WM or updating capacity. In this study, we designed and evaluated the efficacy of a computerized updating training programme.

Improving Working Memory or Updating Capacities

In a seminal study, Klingberg et al. (2005) tested a computerized game-based training programme and found that it improved WM performances on both measures that were structurally similar and those that differed from tasks used during training (near and far transfer, respectively). The present version of the programme, Cogmed, consists of 12 visuo-spatial and verbal memory tasks. Because there are already a number of recent reviews on the efficacy of Cogmed and other WM related training (Melby-Lervåg & Hulme, 2013; Shipstead, Redick, & Engle, 2012b; Wass, Scerif, & Johnson, 2012); here, we focused largely on studies conducted with children and which are of direct relevance.

A number of Cogmed studies have found near transfer effects, but far transfer effects have proven elusive. Dunning, Holmes, and Gathercole (2013), working with 8 year olds, found Cogmed improved performance on a range of untrained WM tasks, but not tasks based on classroom activities or other cognitive assessments. They also found improvement in verbal WM to be sustained for 12 months in a subgroup of participants. In contrast, Holmes et al. (2009), working with 10 year olds, found no immediate far transfer effects to reading, mathematical reasoning, or intelligence, but found an improvement in mathematical reasoning scores six months after training. However, because their control group was not retested at delayed posttest, it is unclear whether the improvement can be attributed to training. K. I. E. Dahlin (2013) found that 9- to 12-year-old boys with attention deficits performed better than a control group in mathematics after Cogmed training. Improved posttest performance on visuo-spatial WM was also observed in the treatment group. Again, as the control group was not administered WM tasks at either pretest or posttest, it is not clear whether improvement in mathematics is due to improved WM.

A more recent study by Holmes and Gathercole (2014) found that teacher-administered Cogmed training improved performance on various WM tasks in 8 to 9 year olds. It also improved English and mathematics scores in low achieving 9 to 11 year olds. However, in their first experiment, they used a pretest-posttest design without a control group. This makes it difficult to attribute improvement in WM to training. In their second experiment, a matched group design was used without assessment of performances before and after training. Furthermore, WM was not assessed. Again, these methodological limitations make it difficult to know whether differences in the criterion measures are due to training. Bergman-Nutley and Klingberg (2014) found that Cogmed improved performances in WM, arithmetic, and a following-instructions tasks in 7 to 15 year olds with poorer WM

and attention capacities. Their findings are encouraging, but the use of typically-developing children in the control group added some uncertainty to the interpretation of findings.

In at least one study, updating abilities were found to predict mathematics performance better than did WM (Lee et al., 2012). In updating tasks, participants are typically asked to monitor, remember, and make decisions regarding stimuli that are presented sequentially. In the n-back task, for example, participants are asked to indicate whether the current stimulus is the same as the one that appeared *n* items earlier. Jaeggi, Buschkuehl, Jonides, and Shah (2011) trained 9 year olds using a programme based on a spatial n-back task and found improved performance on the training task. However, compared to an active control group trained on general knowledge and vocabulary questions, there were no post-training differences in fluid intelligence. In contrast, their earlier studies conducted with adults using single and dual n-back tasks found far transfer effects on fluid intelligence (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Jaeggi et al., 2010). The efficacy of n-back training is controversial. A number of researchers (Chooi & Thompson, 2012; Redick et al., 2013; Thompson et al., 2013) have not been able to replicate Jaeggi et al.'s (2008) findings. Furthermore, methodological limitations in Jaeggi et al. (2008) render the findings difficult to interpret (for details, see Moody, 2009; Redick et al., 2013).

Other training studies have utilised a running span paradigm. Typically, participants are asked to recall the last items at the end of a sequence of presented stimuli. Because participants are not told the length of the sequence, updating of old with new is needed for successful performance. Working with adults, E. Dahlin, Nyberg, Bäckman, and Neely (2008) found training improved performance on both a running span task (similar to that used in training) and on an untrained n-back task. Zhao, Wang, Liu, and Zhou (2011) used non-adaptive running memory span training and found improvements in 9 to 11 year olds' performances on a fluid intelligence task.

In a recent study, Karbach, Strobach, and Schubert (in press) trained 8-year-olds using two adaptive WM complex span tasks. They found that the training group performed better on an untrained visuo-spatial WM task and a reading ability task compared to a control group trained on a non-adaptive version of the tasks. The effect on the visuo-spatial WM task was sustained for three months. However, there were no training effects on switching, inhibition or mathematics performance.

Improving capacities in young children. Few studies have examined WM or updating training effects in younger children. In recent reviews, both Wass et al. (2012) and Melby-Lervåg and Hulme (2013) found that younger children benefited more from such training. However, both reviews located only a handful of studies that involved younger children. Thorell, Lindqvist, Bergman-Nutley, Bohlin, and Klingberg (2009) administered a Cogmed-based WM training programme to 4-year-olds and found improvement in both trained and untrained tests of WM and attention. Bergman-Nutley et al. (2011) administered the same WM training programme to another sample of 4-year-olds and found that it improved performance on WM tasks, but not on fluid intelligence measures. They did, however, find that training using non-verbal reasoning tasks improved fluid intelligence.

In more recent studies, Chacko et al. (2014) and van Dongen-Boomsma, Vollebregt, Buitelaar, and Slaats-Willemse (2014) evaluated the effects of Cogmed on children with attention deficit and hyperactivity disorder. Although Chacko et al. found improvement in short term memory performance, neither study found improvement in WM. Another recent study on 5 year olds by Foy and Mann (2014) found that Cogmed improved visuo-spatial and verbal WM performance and self-regulation as measured by the Head-Toes-Knees-Shoulders test, but not on measures of emergent reading skills.

Apart from Cogmed training, two recent studies have examined the efficacy of other training modalities. Kroesbergen, van ’t Noordende, and Kolkman (2014) trained 5-year-olds

using WM games that were either domain general or that contained numerical content. They found both types of training improved performances on one of the WM measures and speed of quantity discrimination. Numerical WM training also improved performances on an early numeracy test. Goldin et al. (2014) trained 6 year olds on WM, planning, and inhibition. They found that compared to an active control group, trained children performed better in a cognitive flexibility, but not an inhibitory measure. This finding is unexpected because children were not specifically trained on cognitive flexibility. The trained children also performed better on school based language and mathematics tasks than did children in the control condition. Notably, no tests of WM were administered.

The Current Study

Despite a number of positive findings, there are significant concerns over the replicability of findings and the methodology used in some of the studies (e.g., Gibson, Gondoli, Johnson, Steeger, & Morrissey, 2012; 2013; Shipstead, Hicks, & Engle, 2012; Shipstead, Redick, & Engle, 2012a). Shipstead, Redick, et al. (2012b), for example, argued that a number of studies failed to provide adequate measures of WM by indexing it with only one task , confusing short term with working memory, and using only subjective measures. They also questioned the adequacy of control groups that failed to control for the kind of sustained attention needed for WM training. One of the gaps in the literature is the relatively small number of longer term studies. Because an improvement in WM or updating capacities may only improve children's capacities to learn, longer term evaluation is needed to test whether better learning results in better academic performance. Furthermore, as pointed earlier, there is a dearth of studies on very young children, who are most likely to benefit from WM or updating training.

Given these observations, we designed and tested an updating training programme for children. We created seven games based on two commonly used updating paradigms: running

span (Morris & Jones, 1990), and keep track (Yntema, 1963). In the training component of the games, children were asked to remember the last one to four presented exemplars (running span) or the last exemplar from one to four presented categories of objects (keep track). The number of exemplars to be recalled increased adaptively in each game when children's performances improved.

We examined the efficacy of training by contrasting it against two control conditions. For the active control condition, we designed a suite of games that were similar to those used in the training condition, with the exception that children were not asked to remember the stimuli used during game-play. These games allowed us to examine directly whether any training effects were due merely to playing computer games or to updating training. Because we focused on children identified as having difficulties in mathematics, which at this stage included fluency in counting, we explicitly excluded content that required counting or arithmetic in both the training and active control games. In the passive control condition, we had no contact with the children apart from the pretest and posttest.

The main aim of the study was to investigate whether our training improved updating capacity. We were interested in both near and far transfer effects. WM and updating were indexed by a battery of four tasks. A secondary aim was to compare the efficacy of our updating training with that of a more established protocol. For this purpose, we selected Cogmed. Given the extant findings, both sets of training were expected to produce near transfer effects, but we did not have sufficient information on which to base expectations regarding whether one would produce greater improvement than the other.

Method

Participants and Design

The experiment was based on a 4 (Training condition: Updating, Cogmed, active control, versus passive control) x 3 (Assessment time: pretest, immediate posttest, delayed

posttest) full factorial split-plot design. Using G*Power 3.1.0 (Faul, Erdfelder, Lang, & Buchner, 2007), an a priori power analysis showed that a total sample size of 125 was required (effect size f = .25, 85% power, α = .05). We screened 237 children, who were in their first year of formal schooling and who were identified by their teachers or by a school-administered readiness assessment as having difficulties with mathematics. Children identified by the readiness assessment were enrolled in a learning support programme that provided instructions by specially trained teachers as part of the children's weekly lessons. All children participated with parental consent and received small tokens of appreciation.

Children were screened using a WM and an updating task (Block Recall and Animal Updating, see below for details). We selected only those who performed at or below the 20th percentile based on local data (Lee, Bull, & Ho, 2013). Of the 118 children (from 17 primary schools across Singapore) who satisfied the inclusion criteria, one dropped out during training, two were withdrawn by their parents prior to the delayed posttest, and three left school after the immediate posttest and were not contactable. The data from one child, who was unresponsive throughout the study, were excluded from the analyses.

The final data set contained 111 children. Based on parameters used to conduct the initial power analysis, the final sample (M_{age} = 81.73 months; SD = 3.2 months; Range = 76 – 89 months; 69 boys) provided 80% power. However, post-hoc power analyses based on the observed effect sizes of f = .50 and f = .29 (for the omnibus interaction and main training effects, respectively), yielded power of 99% and 94%. Children were quasi-randomly assigned to the four groups: Cogmed (n = 25), updating (n = 32), active control (n = 28), and passive control (n = 26). Where possible, we assigned children from each school to each of the four conditions.

Materials

All children completed a battery of WM, updating, and mathematical proficiency tasks before, immediately after, and six months after training. Because mathematical performance is likely affected by language proficiency and intelligence, these were measured at the start of the study and were used as covariates. The intelligence measure was also administered at posttest.

Working memory and updating tasks. The Animal Updating task was based on the running memory paradigm by Pollack, Johnson and Knaft (1959, as cited in Morris & Jones, 1990). Children were shown a series of animal pictures one at a time on a computer screen and were asked to recall the last two animals (increased to three in a later block) at the end of each trial. Children were not told how many items were to be presented. One point was given for every animal recalled in the correct order (maximum: 60).

In the Letter Rotation task (Ang & Lee, 2008), children were asked to determine whether a capital letter was presented in the normal orientation or as a mirror image. They were also asked to remember the cardinal position aligned with the top of the letter for later recall. One letter at a time appeared in one of seven cardinal positions. The number of positions to be remembered increased progressively. The dependent measure was the total number of positions recalled in the correct order (maximum: 85).

In the Block Recall task (Pickering & Gathercole, 2001), the researcher tapped out sequences of increasing length on nine blocks, positioned randomly on a board. The children were asked to recall and tap the blocks in the same order. Each correctly recalled sequence was given one point (maximum: 54). Although sometimes regarded as a short term memory task, performance on this task has previously been shown to require executive resources, especially in children (e.g., Ang & Lee, 2008). Rudkin, Pearson, and Logie (2007) and Vandierendonck, Kemps, Fastame, and Szmalec (2004) showed that even in adults,

performance on the task was impaired by executive suppression, which again suggests that it draws significantly on executive resources.

The Backward Letter Recall task, was modified from the Backward Digit Recall task (Pickering & Gathercole, 2001) and used consonants in place of digits. Children were asked to recall, in reverse order, sequences of two to seven letters with each correctly recalled sequence scoring one point (maximum: 36). Although deemed a measure of short term memory in adults, St Clair-Thompson (2010) showed that it loads onto WM for children, but short term memory for adults.

Mathematics tasks. The Numerical Operations task from the Wechsler Individual Achievement Test (WIAT-II; Wechsler, 2002) assessed children's abilities to identify and write numbers, count, and solve arithmetic problems. The addition and subtraction fluency tasks were adapted from the WIAT-III (Wechsler, 2009). Children were asked to solve as many mathematical equations out of 48 equations as they could in 1 minute. Consistent with local practice, the equations were presented in a "horizontal" instead of the original "vertical" format. Raw scores were used as dependent measures.

Covariates. The Raven's Coloured Progressive Matrices (Raven, Raven, & Court, 1998) was used as a measure of fluid intelligence. Reading abilities were measured by the Schonell Graded Word Reading Test (Schonell, 1942) in which children were asked to read as many words as possible from a list of 100 words that increased in difficulty. Vocabulary was measured by the Bilingual Language Assessment Battery (BLAB; Rickard Liow & Sze, 2008), which consisted of a set of 100 English words, arranged in increasing difficulty. Children heard each word and had to select a picture, out of a set of four, which best described the word.

Training programmes. Seven computerised games were developed for the updating training. Four games were based on the running span paradigm and three games were based on the keep track paradigm. Figure 1 shows screenshots of one of these games.

<insert Figure 1 here>

All the games started with movement based gameplay that provided the "context" for training. In the Post Bear game, children helped the protagonist deliver mail and were asked to avoid contact with the alien inhabitants by moving the protagonist accordingly. In another game, the protagonist was an ant and the children were asked to help it catch various falling objects. All games were adaptive and all children started at level one, in which they only had to remember the last exemplar in a sequence or the last exemplar from one category. In the Post Bear game, for example, children were asked to remember the last alien encountered by the postman. In the ant game, children were asked to remember the last object in each category of objects that they helped caught. Performing accurately in four of six trials in any given level resulted in progression to the next level. At level two, children were asked to remember the last two exemplars or the last exemplar from each of two categories. There were four levels in each game.

Only one game was available at the commencement of training. A new game was made available when children passed level two. Children who failed to achieve this performance threshold were required to play 36 trials at level two before they could move on to another game. Once the children had unlocked all games, they were allowed to conduct the remaining training using any of the games (with encouragement to alternate between games from each paradigm).

The active control games contained the same themes and game-play, but the children were not asked to remember any stimuli. The children had to play 36 trials of each game

before they could move on to the next game. Again, the children could choose between the various games after all the games were unlocked.

The Cogmed programme for children was administered according to published protocol (see www.cogmed.com). During each session, children were administered eight WM training tasks, with each task consisting of 15 trials. Task difficulty was adapted to each child's current WM span on a trial-by-trial basis. With a pool of 12 tasks, different sets of eight tasks were administered in each session. The programme contained a variety of visuospatial and verbal WM training tasks. In the Input Module task, for example, children were asked to recall a sequence of digits in backward order. The digits were presented on a keypad as part of a robotic arm. In the Rotating Data Link task, children were asked to recall, in sequence, lamps that lit up in a 4 x 4 grid one at a time. The grid rotated 90 degrees in between presentation and recall, so that the lamps were in different locations during recall.

Procedure

All sessions were conducted in supplement to the children's usual school activities. Because game-play was designed to be intensive and continuous in the updating and active control conditions, we restricted each session to 30 minutes. Each Cogmed session took 45 minutes. For all conditions, children were asked to attend 25 sessions, conducted over 8 weeks (three to four session a week) , including a week of school holidays when no training could be conducted. To keep the children motivated during training, they were given stickers after every session. On average, children in the Cogmed, updating, and active control groups attended 24, 23 and 22 sessions respectively.

The immediate posttest was conducted one to two weeks after the last session. The delayed posttest was conducted approximately six months after termination of training. With the exception of the pretest and posttest, the passive control group continued with their usual activities and had no contact with the researchers.

Results

The data were screened for outliers and normality of distribution. Scores that were greater than 3 *SD* from the mean were replaced with scores at 3 *SD* to temper their effects on the mean while preserving their status as outliers (this affected approximately 0.9% of the data points). No multivariate outliers were detected (see Table 1 for descriptive statistics).

<insert Table 1 about here>

To examine whether children's performances on the updating training tasks improved during the course of training, we conducted preliminary analyses on their recall accuracy at the start of training (i.e., data from the two and third sessions) and compared them to the last two training sessions. The dependent variable was the percentage of trials in which the to-be-remembered information was recalled accurately. A paired samples t test showed a significant improvement (M = 34.11%, SD = 15.07% versus M = 43.09%, SD = 18.09%, $t(30) = 2.04$, $p = .05$. Because we did not have access to the raw data generated during Cogmed training, we had to rely on training indices computed automatically by the programme. Similar to Updating training, there was a significant improvement in the training index from the first two days of training (M = 52.30, SD = 7.15) to the last two days (M = 71.34, SD = 10.90); $t(24) = 11.31$, $p < .001$.

The effects of training on WM and updating capacities were examined using a 4 (Training condition) x 3 (Assessment time) multivariate analysis of variance (MANOVA), with the four WM and updating measures as dependent variables. There was a significant interaction effect, $F(24, 290.63) = 2.10$, $p < .01$, $\eta_p^2 = .14$. To identify the locus of the interaction, one-way MANOVAs were conducted, separately for each time point. At pretest, there were no differences amongst the four groups, $F(12, 275.45) = .68$, $p = .77$, $\eta_p^2 = .03$. At immediate posttest, there were significant group differences, $F(12, 275.45) = 4.26$, $p < .01$, $\eta_p^2 = .14$. Univariate tests showed that performance on the Block Recall task was most

affected, $F(3, 107) = 14.14$, $p < .01$, $\eta_p^2 = .28$. Pairwise contrasts showed that the Cogmed group performed significantly better than did the other groups, $p < .01$. Although the univariate ANOVA result was not significant, it is interesting to note that pairwise contrasts showed that the Updating group tended to perform better than did the passive control group on Animal Updating, $p_{\text{uncorrected}} = .04$, $p_{\text{corrected}} = .17$.

We also found significant group differences at the delayed posttest, $F(12, 275.45) = 1.85$, $p = .04$, $\eta_p^2 = .07$. Univariate tests showed that performance on Animal Updating was most affected, $F(3, 107) = 3.03$, $p = .03$, $\eta_p^2 = .08$. The Updating group performed better than did the Cogmed ($p = .02$), active control ($p < .01$), and passive control groups ($p = .05$). Again, though the univariate ANOVA result was not significant, pairwise contrasts showed that the Cogmed group tended to perform better on the Backward Letter Recall task than did the Updating ($p_{\text{uncorrected}} = .03$, $p_{\text{corrected}} = .12$) and passive control groups ($p_{\text{uncorrected}} = .03$, $p_{\text{corrected}} = .14$). Figure 2 shows group differences in performances on the Block Recall and Animal Updating tasks.

Using the same MANOVA model, we examined the effects of training on the mathematics tasks. Children improved across testing sessions, but the magnitude of improvement was not affected by training. Although the majority of our children were enrolled in the LSM programme, 23% were identified by teachers as performing poorly in mathematics, but who were not enrolled in the LSM programme. To take account of possible differences in their performances and responsiveness to training, we re-ran all the analyses using LSM enrolment status as an additional independent variable. The same pattern of findings was obtained.

<insert Figure 2 here>

Performances on all three covariates were significantly correlated with mathematics ($r = .29 – .46$); the BLAB and the CPM were also correlated with Animal Updating ($r = .22 –$

.32) and Letter Rotation (r = .23 – .36); the Schonell was correlated with Backward Letter Recall (r = .25). To investigate whether findings regarding the training were affected by individual differences in the covariates, we conducted separate MANCOVAs for the WM and mathematics measures. The overall pattern of findings did not vary.

Although most of the children attended the scheduled training, there were differences in how well they performed during training. Children in the updating group who remembered more items during training performed better on both the Animal Updating and Letter Rotation tasks at post-test (r = .39 – .52).

Discussion

The Updating group performed better on the updating capacity measure six months after training than did both control groups and the Cogmed group. A similar improvement was found between the Updating and passive control groups at immediate posttest, but it failed to attain significance. In contrast, the Cogmed group significantly out-performed all other groups on the visuo-spatial WM measure at the immediate posttest. However, this advantage was not maintained at delayed posttest. Though the Cogmed group tended to perform marginally better than the Updating and passive control groups on the verbal WM task at delayed posttest, the contrasts were not significant after correction for multiple comparisons. We found no differences between the various groups on the mathematics tasks.

Our findings are generally in line with the literature showing that WM or updating training produces improvement in tasks closely related to those used in training, but does not generalise to other types of executive, reasoning, or academic tasks (Melby-Lervåg & Hulme, 2013). Neither our updating training nor Cogmed improved children’s performance on the mathematics tasks. Inclusion of the language and intelligence measures as covariates did not alter the overall pattern of findings. A supplementary analysis also showed that neither training improved performance in our measure of intelligence. Although some earlier studies

suggested that Cogmed or updating training improved intelligence, this was not replicated in more recent studies (Chooi & Thompson, 2012; Redick et al., 2013; Thompson et al., 2013; Zhao et al., 2011). Cogmed improved visuo-spatial WM performance at the immediate posttest, but this was not sustained after 6 months. Dunning et al. (2013) produced a similar pattern of finding on a verbal WM task; showing improvement immediately after training, but not one year later.

It is unclear why the effects of both types of training were restricted to tasks fairly similar to those used in training. One possibility is that relations between WM or updating capacity and mathematics performance may be moderated by factors that do not benefit directly from improved capacity. Dunning et al. (2013) suggested that additional guidance, instructions or reinforcement are required to help participants apply their newly developed skills to novel tasks or situations. This could be especially true for children. Children have reported that they closed their eyes to concentrate, pay special attention to presented information, or rehearsed information to improve their performance on WM training tasks (Holmes et al., 2009). Indeed, St Clair-Thompson, Stevens, Hunt, and Bolder (2010) found that training children on a memory strategy game improved both WM and mental arithmetic. Such training is not uncontroversial and it has been argued that the use of explicit strategies training may result in improving WM performance without improving core WM capacity. Indeed, this could potentially explain our findings. Children in both the Updating and Cogmed conditions may simply have worked out strategies to perform the trained tasks more efficiently, with the underlying capacities remaining unchanged. From a practical perspective, this may matter little if strategies used in these training tasks can be generalised and applied to more distant academic tasks. In contrast to St Clair-Thompson et al. (2010), our findings provide little evidence for such transfer. However, it should be noted that even in

St Clair-Thompson et al. (2010) study, the effect of training was found on one mental arithmetic task, but not on another standardised mental arithmetic task.

Another explanation for the lack of generalisation rests in differences in timescales of WM versus mathematics tasks. The typical WM task requires keeping information in mind for seconds. In contrast, it is often the case that each mathematics question requires a longer period of sustained attention. It is possible that even with improved WM capacities, children cannot sustain or extend that capability over a longer period of time. Though speculative, another possibility is that bringing online the improved capacity is itself effortful; it is only through sustained practice that it can be successfully applied.

One concern raised in recent reviews (Melby-Lervåg & Hulme, 2013; Shipstead, Redick, et al., 2012b) is the lack of appropriate control groups in previous studies. In this study, we used both an active and a passive control groups. The active condition controlled for the possible effects of sustained engagement with a computer-based task on attention. The passive condition controlled for test-retest and other developmental changes associated with the passage of time. Another concern with prior studies is that of statistical power. Specifically, that some studies may not have sufficient power to detect training effects. Given the cost of training and the existence of pre-existing intervention for our target population, we erred on the side of caution and adopted conventional parameters ($\alpha = .05$, power = .80, Cohen, 1988) in estimating our required sample size. Our findings show that with regards to the major effect of interest (i.e., the interaction between training condition and assessment time), the observed effect size is large. A post hoc analysis shows that the present sample size provides 99% power.

In our training protocol, we specifically avoided the use of numerical and mathematical content because our participants had known difficulties in mathematics. We were concerned that using such content may have decreased the accessibility and increased

difficulty of the training tasks. One potential drawback of our approach is that the use of non-mathematically related stimuli limited transfer to the performance of mathematical problems. This concern is allayed by recent findings from Kroesbergen et al. (2014), who found that there were no differences in the effects of domain-specific versus domain-general WM training in improving pre-schoolers' early numerical skills. Nonetheless, it may be useful to investigate a staged approach in which initial training is conducted with domain-general stimuli, followed by similar training conducted using domain-specific stimuli, and further reinforced by explicit instructions on methods to reduce load on WM or updating resources.

One caveat to our Cogmed findings is that though overall exposure was maintained, training was not administered as intensively as recommended. Due to school time-tabling issues, instead of five times a week, it was administered three to four times a week for most participants. Recent findings suggest that dosage affects the magnitude of the training effect (Alloway, Bibile, & Lau, 2013), however, it is not clear whether the same dosage, administered under different schedule (e.g., over 5 versus 7 weeks) also affects training outcome. Because such logistical issues will likely be encountered when programmes are implemented in wider school settings, if such variation in intensity does markedly reduce efficacy, it has serious implications for parameters under which training can be considered efficacious. In comparison to our Updating training, the amount of time spent on Cogmed was longer (30 versus 45 minutes per session, respectively). Thus, it is possible that the superior effect of Cogmed at immediate posttest is partially attributable to the duration of training.

Practical Application

The findings showed that our updating training resulted in improved updating capacity six months after termination of training. However, the improvement is limited to a task similar to that used in training and did not transfer to better mathematics performance. The

effect of Cogmed was similarly restricted to another task similar to those used in training. If the aim of training is better mathematics outcome, given these limitations, it seems premature to recommend either type of training. In this study, we focused on methods to remediate problems associated with poor WM or updating capacity. In the absence of clear evidence of transfer, another approach is to accommodate to the problem (Gathercole & Alloway, 2007). This may take the form of structuring mathematics problems to reduce the number of steps that needs to be performed in sequence, the use of intermediate points at which partial solutions are to be written down to reduce the amount of information that needs to be kept in mind, and the use of standardised heuristics to reduce the amount of effortful processing needed for solving routine problems.

Conclusion

More research is needed to determine the additional guidance or training processes that are needed for successful transfer to mathematics performance. One potential direction is to develop a better understanding of the underlying processes involved in producing the training effects and to compare that with processes associated with better mathematics performance. Using neuroimaging methods, a comparison of changes in neural activation when children undergo training in mathematics versus in updating training may provide insight on whether critical processes needed for mathematics improvement is targeted by updating based training.

References

Alloway, T. P., Bibile, V., & Lau, G. (2013). Computerized working memory training: Can it lead to gains in cognitive skills in students? *Computers in Human Behavior, 29*(3), 632-638. doi: 10.1016/j.chb.2012.10.023

Ang, S. Y., & Lee, K. (2008). Central executive involvement in children's spatial memory. *Memory, 16*(8). doi: 10.1080/09658210802365347

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Bergman-Nutley, S., & Klingberg, T. (2014). Effect of working memory training on working memory, arithmetic and following instructions. *Psychological Research, 78*(6), 869-877. doi: 10.1007/s00426-014-0614-0

Bergman-Nutley, S., Soderqvist, 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*(3), 591-601. doi: 10.1111/j.1467-7687.2010.01022.x

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19*(3), 273-293.

Chacko, A., Bedard, A. C., Marks, D. J., Feirsen, N., Uderman, J. Z., Chimiklis, A., et al. (2014). A randomized clinical trial of Cogmed Working Memory Training in school-age children with ADHD: a replication in a diverse sample using a control condition. *Journal of child psychology and psychiatry, and allied disciplines, 55*(3), 247-255. doi: 10.1111/jcpp.12146

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd. ed.). Hillsdale, NJ: Lawrence Erlbaum.

Dahlin, E., Nyberg, L., Bäckman, L., & Neely, A. S. (2008). Plasticity of executive functioning in young and older adults: immediate training gains, transfer, and long-term maintenance. *Psychology and Aging, 23*(4), 720-730. doi: 10.1037/a0014296

Dahlin, K. I. E. (2013). Working Memory Training and the Effect on Mathematical Achievement in Children with Attention Deficits and Special Needs. *Journal of Education and Learning, 2*(1). doi: 10.5539/jel.v2n1p118

Dunning, D. L., Holmes, J., & Gathercole, S. E. (2013). Does working memory training lead to generalized improvements in children with low working memory? A randomized controlled trial. *Developmental Science, 16*(6), 915-925. doi: 10.1111/desc.12068

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*(2), 175-191.

Foy, J. G., & Mann, V. A. (2014). Adaptive cognitive training enhances executive control and visuospatial and verbal working memory in beginning readers. *International Education Research, 2*(2), 19-43. doi: http://dx.doi.org/10.12735/ier.v2i2p19

Fürst, A. J., & Hitch, G. J. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Memory and Cognition, 28*(5), 774-782.

Gathercole , S. E., & Alloway, T. P. (2007). *Understanding working memory: a classroom guide*. London: Harcourt Assessment.

Gathercole, S. E., & Pickering, S. J. (2000). Assessment of working memory in six- and seven-year-old children. *Journal of Educational Psychology, 92*(2), 377-390.

Gibson, B. S., Gondoli, D. M., Johnson, A. C., Steeger, C., & Morrissey, R. A. (2012). The future promise of Cogmed working memory training. *Journal of Applied Research in Memory and Cognition, 1*(3), 214-216. doi: 10.1016/j.jarmac.2012.06.003

Goldin, A. P., Hermida, M. J., Shalom, D. E., Costa, M. E., Lopez-Rosenfeld, M., Segretin, M. S., et al. (2014). Far transfer to language and math of a short software-based gaming intervention. *Proceedings of the National Academy of Sciences, 111*(17), 6443-6448. doi: 10.1073/pnas.1320217111

Holmes, J., & Gathercole, S. E. (2014). Taking working memory training from the laboratory into schools. *Educational Psychology, 34*(4), 440-450. doi: 10.1080/01443410.2013.797338

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

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*, 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, 108*(25), 10081-10086. doi: 10.1073/pnas.1103228108

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

Karbach, J., Strobach, T., & Schubert, T. (in press). Adaptive working-memory training benefits reading, but not mathematics in middle childhood. *Child Neuropsychology*. doi: 10.1080/09297049.2014.899336

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlstrom, K., et al. (2005). Computerized training of working memory in children with ADHD-A randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry, 44*(2), 177-186.

Kroesbergen, E. H., van ’t Noordende, J. E., & Kolkman, M. E. (2014). Training WM in kindergarten children: Effects on working memory and early numeracy. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence, 20*(1), 23-37. doi: 10.1080/09297049.2012.736483

Lee, K., Bull, R., & Ho, R. M. H. (2013). Developmental Changes in Executive Functioning. *Child Development, 84*(6), 1933-1953. doi: 10.1111/cdev.12096

Lee, K., & Ng, S. F. (2009). Solving algebraic word problems: The roles of working memory and the model method. In P. Y. Lee (Ed.), *Mathematics education: The Singapore journey* (pp. 204-226). Singapore: World Scientific.

Lee, K., Ng, S. F., Pe, M. L., Ang, S. Y., Hasshim, M. N. A. M., & Bull, R. (2012). The cognitive underpinnings of emerging mathematical skills: Executive functioning, patterns, numeracy, and arithmetic. *British Journal of Educational Psychology, 82*(1), 82-99. doi: 10.1111/j.2044-8279.2010.02016.x

Lee, K., Pe, M. L., Ang, S. Y., & Stankov, L. (2009). Do measures of working memory predict academic proficiency better than measures of intelligence? *Psychological Science Quarterly, 51*(4), 403-419.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49*(2), 270-291. doi: 10.1037/a0028228

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49-100.

Moody, D. E. (2009). Can intelligence be increased by training on a task of working memory? *Intelligence, 37*(4), 327-328. doi: 10.1016/j.intell.2009.04.005

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology, 81*(2), 111-121.

Pickering, S., & Gathercole, S. E. (2001). *Working memory test battery for children (WMTB-C)*: Psychological Corporation.

Raven, J., Raven, J., C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales: Section 2 Coloured Progressive Matrices*. San Antonio, TX: Harcourt Assessment, Inc.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., et al. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General, 142*(2), 359-379. doi: 10.1037/a0029082

Rickard Liow, S. J., & Sze, W. P. (2008). *Bilingual Language Assessment Battery (Version 1).* Unpublished manuscript. National University of Singapore. Singapore.

Rudkin, S. J., Pearson, D. G., & Logie, R. H. (2007). Executive processes in visual and spatial working memory tasks. *The Quarterly Journal of Experimental Psychology, 60*(1), 79-100. doi: 10.1080/17470210600587976

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*(4).

Schonell, F. (1942). *Backwardness in the basic subjects*. London, UK: Oliver & Boyd.

Shipstead, Z., Hicks, K. L., & Engle, R. W. (2012). WM training remains a work in progress. *Journal of Applied Research in Memory and Cognition, 1*(3), 217-219. doi: 10.1016/j.jarmac.2012.07.009

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

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

St Clair-Thompson, H. L. (2010). Backwards digit recall: A measure of short-term memory or working memory? *European Journal of Cognitive Psychology, 22*(2), 286-296. doi: 10.1080/09541440902771299

St Clair-Thompson, H. L. (2011). Executive functions and working memory behaviours in children with a poor working memory. *Learning and Individual Differences, 21*(4), 409-414. doi: 10.1016/j.lindif.2011.02.008

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology, 59*(4), 745-759.

St Clair-Thompson, H. L., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology, 30*(2), 203-219. doi: 10.1080/01443410903509259

Thompson, T. W., Waskom, M. L., Garel, K.-L. A., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., et al. (2013). Failure of Working Memory Training to Enhance Cognition or Intelligence. *PLoS ONE, 8*(5), e63614. doi: 10.1371/journal.pone.0063614

Thorell, L. B., Lindqvist, S., Bergman-Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science, 12*(1), 106-113. doi: 10.1111/j.1467-7687.2008.00745.x

van Dongen-Boomsma, M., Vollebregt, M. A., Buitelaar, J. K., & Slaats-Willemse, D. (2014). Working memory training in young children with ADHD: a randomized placebo-controlled trial. *Journal of child psychology and psychiatry, and allied disciplines, 55*(8), 886-896. doi: 10.1111/jcpp.12218

Vandierendonck, A., Kemps, E., Fastame, M. C., & Szmalec, A. (2004). Working memory components of the Corsi blocks task. *British Journal of Psychology, 95*(1), 57-79. doi: 10.1348/000712604322779460

Wass, S. V., Scerif, G., & Johnson, M. H. (2012). Training attentional control and working memory – Is younger, better? *Developmental Review, 32*(4), 360-387. doi: 10.1016/j.dr.2012.07.001

Wechsler, D. (2002). *Wechsler Individual Achievement Test* (2nd ed.). Psychological Corp: San Antonio, TX.

Wechsler, D. (2009). *Wechsler Individual Achievement Test* (3rd ed.). San Antonio, TX: Psychological Corporation.

Yntema, D. B. (1963). Keeping track of several things at once. *Human Factors, 5*(1), 7-17. doi: 10.1177/001872086300500102

Zhao, X., Wang, Y., Liu, D., & Zhou, R. (2011). Effect of updating training on fluid intelligence in children. *Chinese Science Bulletin, 56*(21), 2202-2205. doi: 10.1007/s11434-011-4553-5

Table 1

Means and Standard Deviations (in Parentheses) of the Outcome Measures

	Updating			Cogmed			Active Control			Passive Control		
Task	Pre	Post 1	Post 2	Pre	Post 1	Post 2	Pre	Post 1	Post 2	Pre-	Post 1	Post 2
AU	25.00	39.22	41.25	20.28	34.40	35.64	23.54	35.57	34.79	19.35	33.04	36.52
	(11.69)	(10.13)	(7.44)	(10.88)	(10.90)	(9.97)	(11.24)	(11.25)	(9.11)	(10.94)	(13.00)	(10.24)
LR	18.41	21.69	22.09	21.75	20.20	19.52	20.56	19.96	21.29	19.53	18.50	21.31
	(5.93)	(7.21)	(8.95)	(9.20)	(6.96)	(8.23)	(10.70)	(6.64)	(8.64)	(9.73)	(6.22)	(7.46)
BR	19.31	20.81	21.56	19.76	27.20	23.84	19.64	21.50	21.82	19.52	21.58	22.89
	(4.22)	(4.19)	(4.41)	(3.26)	(4.55)	(3.77)	(4.48)	(3.49)	(4.29)	(5.43)	(3.89)	(4.83)
BLR	6.28	7.66	7.56	6.16	8.28	8.88	6.71	7.21	7.71	6.54	7.08	7.55
	(1.42)	(2.48)	(2.12)	(2.24)	(2.25)	(2.56)	(2.54)	(2.46)	(2.12)	(2.25)	(1.62)	(1.94)
NO	9.47	12.38	15.16	9.88	12.00	14.88	9.14	11.93	13.89	9.92	12.50	15.77
	(2.08)	(2.64)	(3.45)	(2.32)	(2.08)	(3.24)	(2.26)	(2.55)	(3.45)	(2.06)	(2.61)	(3.87)
AFlu	6.97	11.59	16.75	8.48	11.84	15.08	6.82	10.18	13.79	6.88	10.43	14.31
	(4.79)	(6.33)	(6.65)	(4.60)	(5.77)	(6.45)	(5.16)	(6.24)	(7.43)	(3.89)	(6.73)	(6.72)

SFlu	3.45	7.96	11.78	4.12	6.32	11.04	4.18	6.89	10.75	3.38	7.62	10.19
	(3.40)	(5.07)	(6.92)	(2.85)	(3.65)	(4.13)	(3.85)	(4.98)	(5.71)	(3.65)	(4.60)	(5.23)

Note. Pre = Pretest; Post 1 = Immediate posttest; Post 2 = Delayed posttest; AU = Animal Updating Task; BR = Block Recall Task; BLR = Backward Letter Recall Task; LR = Letter Rotation Task; AFlu = Additional Fluency Task; SFlu = Subtraction Fluency Task; NO = Numerical Operations Task. Standard Deviations are in Parentheses.

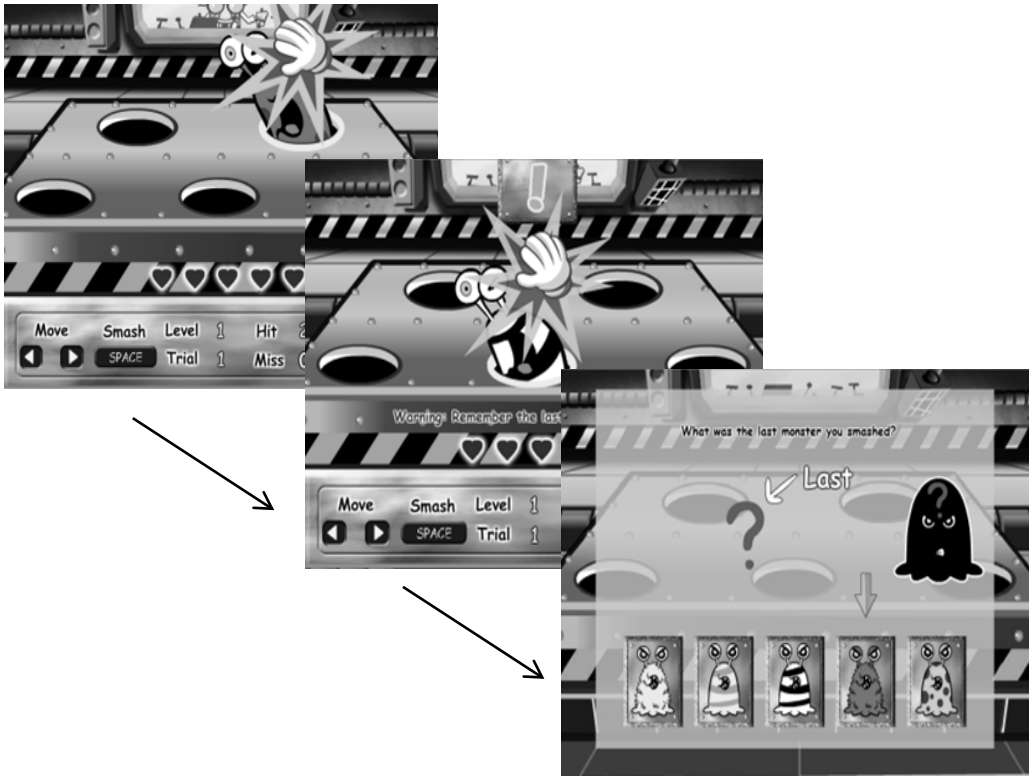

Figure 1. Sequential screenshots of a game based on the running span paradigm. The first screenshot shows game play during which children are asked to smack the monsters as soon as they appear. The second screen shows the warning sign at the top of the screen, which alerts the children to start memorizing the identity of the monsters. Depending on the block, children were informed as to whether they should remember the last, the last two, the last three, or last four monsters. The third screenshot shows the recall screen. The children were asked to recall the last monster and had five response alternatives from which to choose. Original screenshots were in full colour.

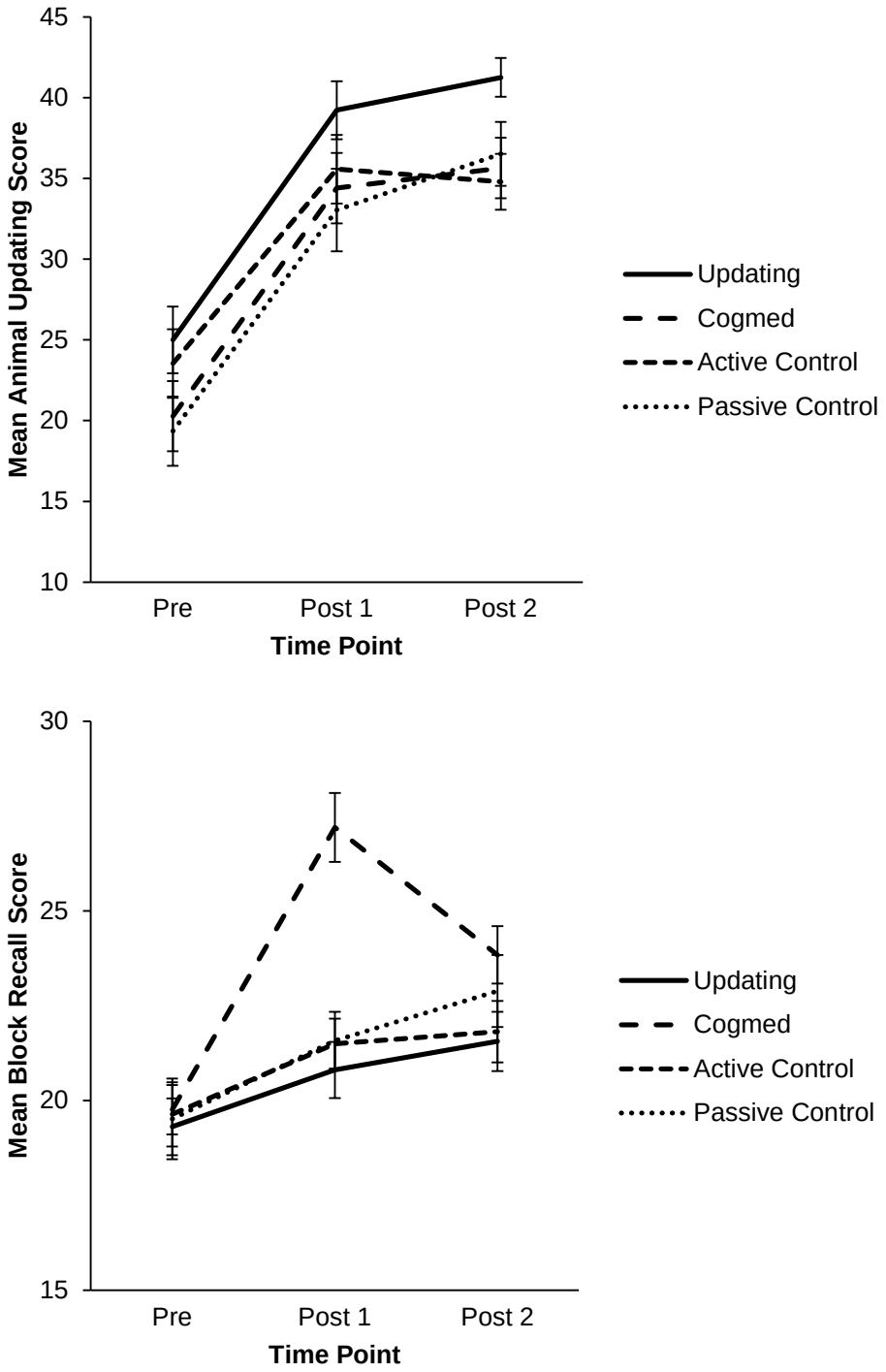

Figure 2. Mean Animal Updating and Block Recall scores by condition and time of test. The error bars depict standard errors.