Case Study 1: An Evidence-Based Practice Review Report

Theme: School Based Interventions for Learning

Does Cogmed working memory training improve working memory for school age children with Special Educational Needs?

Summary

In recent years there has been a great deal of research analysing the effectiveness of Cogmed working memory training (CWMT). CWMT is a computer based Working Memory (WM) intervention that consists of a number of games that are designed to be played five days a week for five weeks. Much of the research analysing the effectiveness of CWMT has focussed on it's ability to help individuals with Attention Deficit Hyperactivity Disorder (ADHD). However, there are many types of Special Educational Needs (SEN) that are associated with WM difficulties. Therefore, this systematic literature review assessed the effectiveness of CWMT in improving WM in children with SEN other than ADHD. Database searches found six studies that were included in the review. These studies were analysed according to their methodological quality and relevance, and relevance to topic. When looking at the effect sizes, it was found that CWMT generally improved working memory scores, although effect sizes differed according to different studies and were generally quite small. Visuo-spatial working and short-term memory seemed to show more improvement post training than verbal working and short-term memory. It was concluded that CWMT was an effective WM intervention for children with SEN other than ADHD. However, many of the studies were underpowered and therefore future research should aim to have bigger sample sizes. Also, more efforts should be made to ensure there are adequate matched control groups.

Introduction

What is CWMT?

CWMT is a computer based program that aims to improve a person's WM. The program was designed by Torkel Klingberg, a professor of cognitive neuroscience based at the Karolinska Institute in Sweden ("Cogmed Working Memory Training | History", 2015). Cogmed, as a company, was founded in 2001 and since 2010 has been distributed by Pearson Education (Pearson, 2010). CWMT is not available to the public and can only be accessed through an organisation.

The training consists of different tasks that take the form of simple computer games. Participants are presented eight tasks each day and the training is supposed to be administered for 25 days. It is available to people of all ages with different versions designed for younger children, older children and adults. It is stipulated that those completing the program have a 'CogMed coach' who has received training and assists the individual through the training offering advice, support and incentives.

Pearson claim that "Cogmed is an evidence-based program for improving working memory" that "aims to improve attention by increasing working memory capacity" ("Cogmed UK Official Minisite". n.d.). It is also stated that "Cogmed acts as a "primer" for improved learning, allowing the student to build the cognitive platform needed to learn successfully" (Pearson Inc., 2014).

Basis in Psychological Theory

WM can be broadly defined as the ability to store and manipulate information within one's short-term memory (STM). The theory was first put forward by Baddeley and Hitch (1974) and suggests that there are a number of components of WM that interact to manipulate and process information. Baddeley and Hitch's (1974) original model suggested that there were three components; The phonological loop, the visuo-spatial sketchpad, the episodic buffer and the central executive. Each component is responsible for a different type of information processing for instance; the phonological loop is responsible for the manipulation of phonological and verbal information and the visuo-spatial sketchpad stores information based on its visual or spatial features. Although the model has changed and evolved over time, many of the key features outlined by Baddeley and Hitch (1974) remain. There has been research linking components of WM to particular neural areas (Eriksson et al., 2015). Neuroimaging studies have found high levels of activation in the posterior parietal cortex and the Broca's Area when completing a verbal WM task (Smith & Jonides, 1998) and lesions to the temporal lobe are associated with deficits in visual WM but not verbal WM (Owen et al., 1996).

It has long been documented that the brain is adaptive and can change according to stimulation received from the environment. Research has found that deaf participants had increased activation in neural areas associated with tracking motion when presented with a task that required them to monitor moving stimuli (Bavelier et al., 2000). This suggests that sensory deprivation had influenced the structure of their brains to better attend the visual stimuli.

Although this theory of neuroplasticity suggests the brain can restructure itself due to sensory deprivation, it also suggests that the brain can respond positively to stimulation. Research has found that musicians respond differently when processing multi-sensory stimuli (Pantev, Paraskevopoulos, Kuchenbuch, Lu, & Herholz, 2015) and training in golf increases neural activity within the motor cortex when presented with a kinetic motor imagery task (Bezzola, Mérillat, & Jäncke, 2012). In recent years, a number of computer programs have been designed to specifically train particular cognitive faculties. Evidence has been found to suggest that brain training games can improve processing speed, attention and WM in people with aphasia (Bruce, Edmundson, Aviet, & Willison, 2010) and can have positive benefits for healthy elderly people (Willis et al., 2006).

Relevance to EP practice

Deficits in WM are associated with reduced performance on reading comprehension measures (Grant, Daneman, & Carpenter, 1980), poor numeracy ability (Kyttala, Aunio, & Hautamaki, 2010) and low academic achievement (Gathercole & Pickering, 2000). Other theorists have suggested that WM has a salient influence on an individual's ability to learn (Cowan, 2014). Therefore, it is important for Educational Psychologists to be aware of how WM can act as a barrier to education and what can be put in place to improve a child's WM.

Research has found that CWMT improved student progress when implemented amongst typically developing children in an English school (Holmes & Gathercole, 2013). A number of reviews have already been produced reviewing the effectiveness of CWMT. The results so far have been mixed, Shipstead, Redick, and Engle (2012) conducted a review and suggested that there was not currently enough substantial evidence to suggest CWMT significantly improves WM. Similar findings were found using a meta-analysis (Hulme & Melby-Lervåg, 2012). Barkley (2013) conducted a review and found that, following WM training, there were immediate positive effects but no evidence to support long-term effects and more general effects such as scholastic achievement. However, Shinaver, Entwistle and Söderqvist (2014) reviewed a great deal of research and found evidence to suggest that WM training can improve WM and attention, and this effect is retained over time. Spencer-Smith and Klingberg (2015) conducted a meta-analysis and found that

CWMT was associated with improved attention in daily life. A UCL student participating on the educational and child psychology doctorate course also conducted a systematic literature review on research regarding WM training with young people with ADHD (Atkins, 2014).

A great deal of these previous reviews have focussed on either general populations or whether CWMT was a successful intervention for people with ADHD. At the time of writing this review, there was no systematic literature review assessing CWMT's effectiveness with improving WM amongst children with Special Educational Needs (SEN) other than ADHD.

Review question

Does CWMT improve WM in school age children with SEN other than ADHD?

Critical review of the evidence base

Literature search

A literature search was conducted in December 2015 using three online electronic databases; Proquest, Pubmed and ERIC. Proquest and Pubmed were chosen as they have a vast selection of publications and have been used in other systematic literature reviews. ERIC was also chosen as it was more specific to education research and has also been used in systematic literature reviews.

In these databases, the search term "Cogmed" was used across all fields. 106 publications were found. After duplicates were removed, 89 studies were screened for eligibility. The inclusion and exclusion criteria detailed in table 1 were used to scrutinise the research.

Table 1

Table showing inclusion/exclusion criteria with rationale

	Inclusion Criteria	Exclusion Criteria	Rationale
1. Publication Type	Publication is published in a peer reviewed journal.	Publication is not published in a peer reviewed journal.	Peer reviewed journals will contain higher quality studies and will contain fewer errors as they have been reviewed.
2. Participants	a) Participants are children of English schooling age (4 to 16 years old).	a) The majority of participants were under the age of 4 or over the age of 17.	a) The review question specifies that the training must be given to children.
	b) Participants have any type of SEN other than ADHD.	b) Over 50% of participants have been diagnosed as having ADHD	b) The review question specifies that the participants have SEN other than ADHD.

	<u>Inclusion Criteria</u>	<u>Exclusion Criteria</u>	<u>Rationale</u>
3. Methodolog y	The study is an experimental study that uses pre and post measures.	The study is not an experimental study.	This review is aiming to establish the evidence base and as a result the research must be conducted in a scientifically robust manner
4. Intervention	CWMT is the only intervention that is used with the participants.	A different WM intervention is used or CWMT is used in conjunction with another form of training.	The review question is specific to CWMT and not WM training programmes in general.
5. Outcome	Studies that assess whether the participants' WM has improved.	Studies that assessed other outcome measures.	The review is aiming to establish whether WM training specifically improves WM.
6. Language	Publication is written in English.	Publication is written in a language other than English.	To ensure there were no errors in translation.

The rationale for each of the criteria are shown in the table above. No date restriction criterion were used because CWMT was first released in 2001 and therefore it was decided that any research would be relevant. Publications were first screened by title and then by abstract. The final 12 articles were screened for eligibility by analysing the full text. A flow diagram outlining the number of studies that were excluded for each set of criteria is provided in figure 1. A list of the included studies is provided in table 2 and a list of full text excluded studies with the rationale for their exclusion is included in Appendix A.

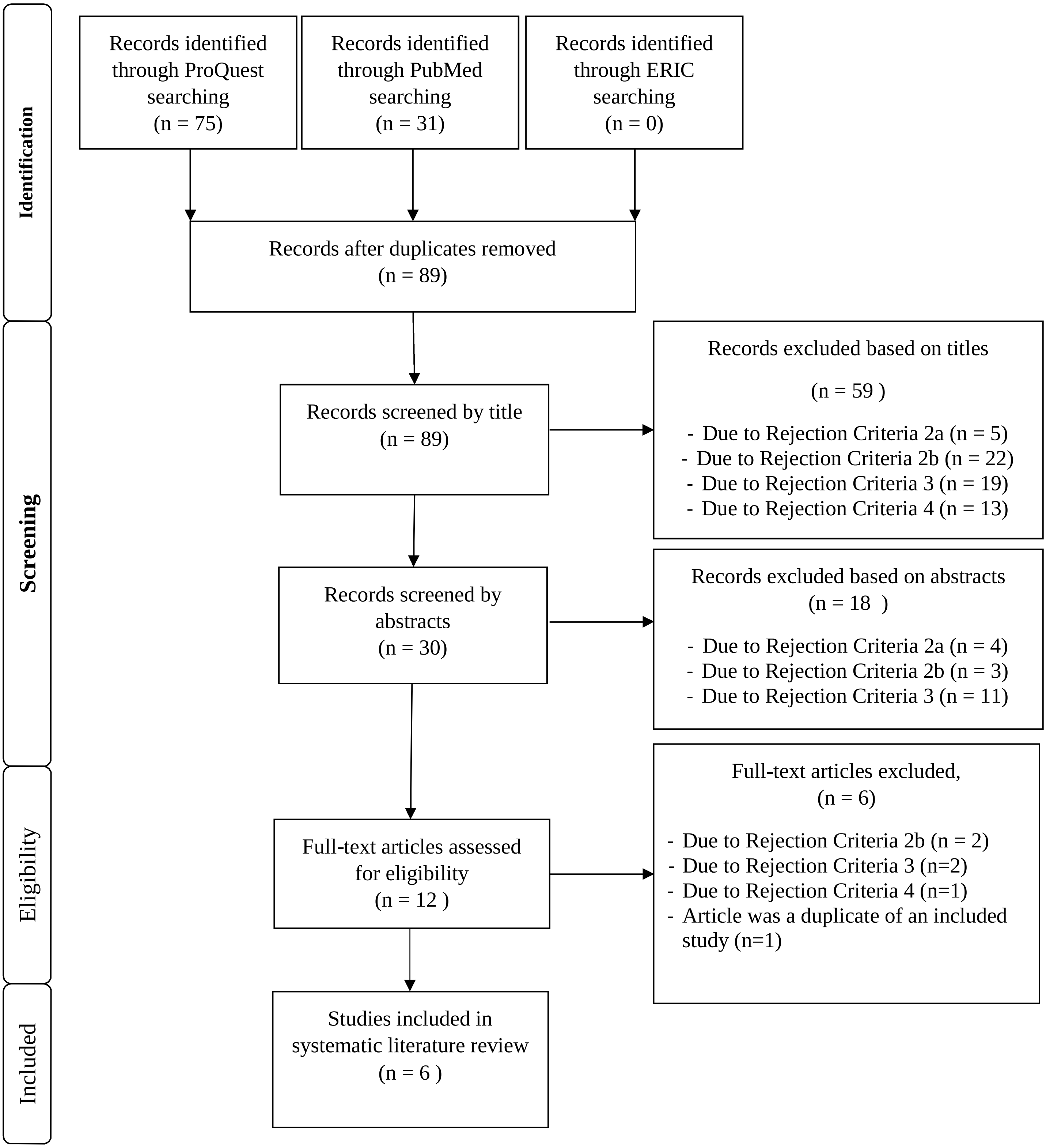

Figure 1. Flow diagram outlining how many studies were rejected during each stage of the screening

Table 2

List of included studies

Bennett, S. J., Holmes, J., & Buckley, S. (2013). Computerized memory training leads to sustained improvement in visuospatial short-term memory skills in children with down syndrome. *American Journal on Intellectual and Developmental Disabilities*, *118*(3), 179–192. http://doi.org/10.1352/1944-7558-118.3.179

Grunewaldt, K. H., Skranes, J., Brubakk, A.-M., & Lähaugen, G. C. C. (2015). Computerized working memory training has positive long-term effect in very low birthweight preschool children. *Developmental Medicine and Child Neurology*, 1–7. http://doi.org/10.1111/dmcn.12841

Holmes, J., Butterfield, S., Cormack, F., van Loenhoud, A., Ruggero, L., Kashikar, L., & Gathercole, S. (2015). Improving working memory in children with low language abilities. *Frontiers in Psychology*, *6*(April), 519. http://doi.org/10.3389/fpsyg.2015.00519

Kerr, E. N., & Blackwell, M. C. (2015). Near-transfer effects following working memory intervention (Cogmed) in children with symptomatic epilepsy: An open randomized clinical trial. *Epilepsia*, 56(11), 1784–1792. http://doi.org/10.1111/epi.13195

Kronenberger, W., & Pisoni, D. (2011). Working memory training for children with cochlear implants: A pilot study. *Journal of Speech, Language, and Hearing Research, 54*(August), 1182–1197. http://doi.org/10.1044/1092-4388(2010/10-0119)1182

Løhaugen, G. C. C., Antonsen, I., Håberg, a, Gramstad, a, Vik, T., Brubakk, a M., & Skranes, J. (2011). Computerized working memory training improves function in adolescents born at extremely low birth weight. *Journal of Pediatrics*, *158*(4), 555–561.e4. http://doi.org/10.1016/j.jpeds.2010.09.060

Quality and relevance appraisal

To establish the quality and relevance of the included research, the weight of evidence protocol outlined by Gough (2007) was used. This protocol stipulates that all reviewed research is weighted according to; methodological quality, methodological relevance and relevance to topic. To establish methodological quality the included studies were coded based on Kratochwill's (2003) Evidence-based intervention coding protocol. This protocol was produced by the APA task force and provided a stringent protocol to assess the quality of each of the included studies. The protocol was modified for use in this review, with some irrelevant items being discarded. Details of the discarded items and the rationale for their exclusion are included in Appendix B. Criteria used to code studies on methodological relevance and relevance to topic were devised by the author, details of these criteria are included in Appendices C and D. WoE D was based on the sum of the scores for WoE A,B and C, the rationale for the descriptors is included in Appendix D. The weighting of each of the included studies is outlined in table 3.

Table 3

Weighting of Evidence by study and criteria

Studies	WoE A - Methodological quality		WoE B - Methodological relevance		WoE C - Relevance to topic		WoE D – Overall score	
(Bennett, Holmes, & Buckley, 2013)	2.50	Medium	3.00	High	3.00	High	8.50	High
(Grunewaldt, Skranes, Brubakk, & Lähaugen, 2015)	2.63	High	2.00	High	1.67	Medium	6.29	High
(Holmes et al., 2015)	2.38	Medium	1.00	Low	2.33	High	5.71	Medium
(Kerr & Blackwell, 2015)	2.75	Medium	3.00	High	2.67	High	8.42	High
(Kronenberger & Pisoni, 2011)	2.00	Medium	1.00	Low	2.33	High	5.33	Medium
(Løhaugen et al., 2011)	2.00	Medium	2.00	Medium	2.67	High	6.67	High

Participants

The included studies were conducted in a number of different countries including; England (Bennett et al., 2013; Holmes et al., 2015), Norway (Grunewaldt et al., 2015; Løhaugen et al., 2011), Canada (Kerr & Blackwell, 2015) and the USA (Kronenberger & Pisoni, 2011). The participants' average ages ranged from 5 (Grunewaldt et al., 2015) to 14 (Løhaugen et al., 2011), although there was one study where the mean age of the participants was not reported (Kerr & Blackwell, 2015).

The sample sizes in the included studies differed considerably ranging from 9 (Kronenberger & Pisoni, 2011) to 77 (Kerr & Blackwell, 2015). This affected the statistical power of the studies as many of the studies were underpowered. Whether the study had a sufficient sample size was determined in the WoE A appraisal.

The special educational needs of the children were determined through a variety of methods. All of Bennett et al.,'s (2013) participants were diagnosed with Downs syndrome. Kerr and Blackwell, (2015) included children who had a confirmed diagnosis of epilepsy for at least six months without any other "developmental disability". Kronenberger and Pisoni (2011) included children who had severe or profound bilateral hearing loss and had received cochlear implants before the age of three.

The studies conducted on children with very low birthweight (Grunewaldt et al., 2015; Løhaugen et al., 2011) based their sample on children that had been admitted to a certain hospital or neonatal intensive care unit within a specified time frame. However, Grunewaldt et al., (2015) excluded children with co-morbid conditions such as epilepsy and severe cerebral palsy and Løhaugen et al., (2011) included a child with cerebal paulsy and two childen with reduced hearing. Holmes et al., (2015) determined whether children had low language abilities using cutoffs from a standerdized assessment measuring vocabulary and a subtest from another assessment measuring the child's ability to recall a sentence.

Design

All of the included studies took a quantitative approach assessing differences in score pre and post intervention. All of the studies used a between participants design, other than Kronenberger and Pisoni (2011) who used a within participants design and Grunewaldt et al., (2015) who used a stepped wedge design. The studies that used between participants designs all used no intervention control groups. As the children were aware they were receiving an intervention, it could be argued that this does not control for the Hawthorne effect. If control groups received an alternative intervention that did not aim to improve WM for instance, an emotional literacy intervention, this would have controlled for the Hawthorne and novelty effect.

Bennett et al., (2013) and Kerr and Blackwell, (2015) used random allocation to determine whether the sample received the intevention or not. Grunewaldt et al., (2015) compared long term effects of the training with a control group of children who were born at a very low birthweight a year after the intervention group, this may have influenced the results as the control group were a year younger than the intervention group. Two of the included studies did not use matched control groups. Løhaugen et al., (2011) used a control group that consisted of children born at a healthy birthweight (comparing them to children born at very low birthweight) and Holmes et al., (2015) used a control group of children who had normal language abilities (comparing them with a sample with low language abilities). This non-random, non-matched group allocation could be seen as a confounding variable as the groups were not equivalent and therefore any comparisons would be speculative.

A power analysis was conducted to establish whether the included studies had sufficient sample sizes. No previous meta-analyses were found that analysed the effectiveness of CWMT with samples of children with SEN other than ADHD. The power analysis figure was drawn from two meta analyses. The first analysed the effectiveness of CWMT in reducing inattention in daily life among people with ADHD and found an effect size of d=-0.47 (Spencer-Smith & Klingberg, 2015). It is important to note here that, in this study a reduction of inattention in daily life was seen as a positive outcome and therefore the achieved effect size was negative. This was interpreted as a medium effect of reducing inattention. The other meta analysis provided an effect size of d=0.50 for improvements in non-verbal reasoning among participants with WM problems (Hulme & Melby-Lervåg, 2012). It was established that CWMT would roughly achieve a medium effect size, as both the previous meta-analyses attained a medium effect size, with Spencer-Smith & Klingberg (2015) showing a reduction of inattention (d=-0.47) and Hulme & Melby-Lervåg (2012) showing improvements in non verbal reasoning (d=0.50). It is imortant to note that the effect sizes were not combined as one of the meta analyses was based on how CWMT can reduce symptoms of inattention and therefore the effect size was negative. As both the meta analyses achieved a medium effect size and the majority of the measures included in this review searched for an improvement in pretest-posttest WM scores, the power analysis was based on the effect size d=0.5. Hulme and Melby-Lervåg's (2012) overall meta analysis effect size was d=0.15, but this was not taken into account in this power analysis, as a number of their included studies were conducted on typically developing adult populations and were therefore not relevant to this systematic literature review. The effect size that was included in this review was calculated from the studies that specifically used participants with WM difficulties.

Intervention

The implementation of the intervention also differed across studies. CWMT is designed to be delivered five days a week for five weeks totalling 25 sessions over 35 days. Some of the studies ensured and reported that all students completed the full 25 sessions (Bennett et al., 2013; Grunewaldt et al., 2015; Kerr & Blackwell, 2015; Løhaugen et al., 2011). Holmes et al. (2015) only specified that the participants completed 20 sessions. Kronenberger & Pisoni, (2011) only had six participants complete the full 25 sessions but included the data from any child that completed 20 sessions.

In the majority of studies, the participants completed the training at home. Two studies (Bennett et al., 2013; Holmes et al., 2015) had participants complete the training in school and one study did not specify the setting of the training (Grunewaldt et al., 2015). As the theme for this review was ‘school based interventions for learning’, the studies that were conducted in a school received higher WoE B scores.

In the majority of the studies, an adult who had received some sort of training or support supervised the children. Holmes et al., (2015) did not specify whether the adults supervising the children had received any training and Grunewaldt et al., (2015) did not specify any detail regarding those supervising the children. The implementation of the intervention was scored according to whether the intervention was delivered in full, where the training was conducted and how the participants were supervised. Details of the criteria and scoring system can be found in appendix C.

Outcome measures

To establish WM scores a number of different assessments were used. The majority of the studies used more than one measure of WM, only two used one measure (Holmes et al., 2015; Løhaugen et al., 2011). This is an advantage, as using more than one measure allows for a number of constructs to be assessed in different manners. Kerr and Blackwell, (2015) used four measures; The Working Memory Test Battery for Children, Wechsler Intelligence Scale for Children Fourth Edition (WISC-IV), Test of everyday attention for children and an attentional capacity test. Two studies used the Automated Working Memory Assessment (AWMA) (Bennett et al., 2013; Holmes et al., 2015). This assessment is a computer based WM assessment that consists of eight components that provide scores for verbal STM, verbal WM, visuo-spatial STM, and visuo-spatial WM. Two studies used the Wechsler Memory Scale (Grunewaldt et al., 2015; Løhaugen et al., 2011), this measure provides total scores for a participant's spatial span and digit span that give an indication of a participants' visual WM and verbal WM. Three studies used components of the Wechsler Intelligence Scale for Children (WISC-IV) (Grunewaldt et al., 2015; Kerr & Blackwell, 2015; Kronenberger & Pisoni, 2011). The spatial span backwards subcomponent was frequently used as a measure of a participant's visuo-spatial WM. The Behavior Rating Inventory of Executive Function (BRIEF) is a measure of executive functioning and has a sub component that gives a score for WM. This measure was used in two studies (Bennett et al., 2013; Kronenberger & Pisoni, 2011) where it was given to parents to complete. Although this has the advantage of assessing the child's WM through different methods and different environments, Bennett et al., (2013) acknowledge that it is a subjective measure as it is completed by the child's parents.

Findings

Table 4 shows all the effect sizes from the data provided from all the included studies. Bennett et al., (2013), Holmes et al., (2015), Kerr and Blackwell (2015) and Løhaugen et al., (2011) provided mean scores, pre and post intervention for their treatment and control group, their effect sizes were calculated using Morris' (2008) formula suggestion. Kronenberger and Pisoni (2011) did not include a control group and therefore Becker's (1988) formula for within person change was used. Grunewaldt et al., (2015) used a stepped wedge randomised control trial design therefore Becker's (1988) formula was used on the overall pre and post means and standard deviations. The descriptors were drawn from Cohen's (1992) suggestions for what would constitute small, medium and large effect sizes.

Bennett et al., (2013) found large effects for participants' visuo-spatial STM. Medium effect sizes were found in scores of verbal WM and visuo-spatial WM. A small effect was found for WM scores as rated by the BRIEF and no effect was found in verbal STM scores. This study had the highest WoE D score of all the studies included in this review and therefore the findings from this study should be considered as accurate and a fair representation of how CWMT could improve WM outcomes amongst children with Down's syndrome.

Grunewaldt et al., (2015) found a large effect for participants' spatial span backwards scores and a medium effect for the total spatial span scores. Small effect sizes were found in the spatial span forwards subcomponent and the digit span backwards component and no effect was found in digit span backwards scores. This study was rated with a high WoE D score and therefore can be generally regarded as methodologically sound. However, the study received one of the lowest WoE C ratings; this was predominantly due to the lack of detail provided regarding how the intervention was implemented. Important details were not specified, such as whether the children completed the training at home or in a school setting and whether the children were monitored or supervised. Therefore, the results of the study cannot be generalised to a particular context.

Holmes et al., (2015) found small negative effects for scores of verbal WM, visual-spatial STM and WM, suggesting that the individuals in the intervention group performed worse at that post-test stage. A small positive effect was found for verbal STM scores. Due to the low number of participants, this study was underpowered. As a result, there may have been effects that were not found and the extent of the effects they found may have been understated. This study also received a low WoE B score, suggesting that it lacked methodological relevance. This was primarily because the study did not include a control group and therefore the findings should be interpreted cautiously.

Kerr and Blackwell (2015) had a very large sample size and were also rated with one of the highest WoE D scores. Therefore, the results of this study should be considered in high regard as they were obtained in a methodologically rigorous manner. They found large effects for the counting recall subcomponent of the WM test battery and the visual attention component of the WISC-IV. They found medium effects for visual WM and small effects for all three measures of visual verbal WM.

Kronenberger and Pisoni (2011) found mostly medium effect sizes for measures of immediate verbal phonological memory, visual-spatial sequential memory capacity and visual-spatial WM. They found a large effect for the WM component of the BRIEF and visual-spatial WM component of the WISC-IV. A small effect was found for immediate verbal WM capacity. When considering these findings, one must take into account that this study received the lowest WoE D score suggesting that overall the study was lacking, particularly in methodological quality and relevance. This was primarily due to the small sample size and the lack of a control group.

Løhaugen et al., (2011) did not find any effects for any of the measures they used. This was an interesting finding as the study received a high WoE D rating, suggesting that the methodology employed in the study met a high standard and was relevant to the topic being studied.

Table 4;

Table showing effect sizes for each included study

Study	Assessment used	Components	Effect size	Effect size descriptors	Overall weight of Evidence
(Bennett et al., 2013)	AWMA	Verbal STM	0.16	None	High
		Verbal WM	0.57	Medium	
		Visuo-spatial STM	0.89	Large	
		Visuo-spatial WM	0.70	Medium	
	BRIEF	WM	-0.36	Small	
(Grunewaldt et al., 2015)	Wechsler memory scale	Visual WM - Spatial Span forwards	0.31*	Small	High
		Visual WM - Spatial Span Backwards	0.81*	Large	
		Visual WM - Spatial Span total	0.63*	Medium	
	WISC	Verbal WM - Digit span forwards	0.00*	None	
		Verbal WM - Digit span backwards	0.29*	Small	
(Holmes et al., 2015)	AWMA	Verbal STM	0.44	Small	Medium
		Visuo-Spatial STM	-0.07	None	
		Verbal WM	-0.44	Small	
		Visuo-Spatial WM	-0.27	None	
(Kerr & Blackwell, 2015)	WM Test battery for children	Basic auditory WM	0.08	None	High
		Auditory WM	0.86	Large	
		Visual-Verbal WM	0.48	Small	
		Counting recall	1.07	Large	

Study	Assessment used	Components	Effect size	Effect size descriptors	Overall weight of Evidence
	WISC IV	Visual Attention	0.98	Large	
		Visual WM	0.75	Medium	
	Test of everyday attention for children	Visual-Verbal WM	0.30	Small	
	Attentional capacity test	Visual-Verbal WM	0.27	Small	
(Kronenberger & Pisoni, 2011)	WISC IV	Immediate verbal phonological memory	0.60*	Medium	Medium
		Immediate verbal WM capacity	0.32*	Small	
		Visual-spatial sequential memory capacity	0.55*	Medium	
		Visual-spatial WM	0.84*	Large	
	BRIEF	WM	-0.48*	Large	
(Løhaugen et al., 2011)	Wechsler memory scale	Verbal WM -Digit Span	-0.22	None	High
		Verbal WM -Letter number sequencing	-0.10	None	
		Visuo-Spatial memory - Spatial Span	0.16	None	

* = Effect size calculated using Becker's (1988) within person change formula

All other effect sizes calculated with Morris' (2008) formula

Conclusions and recommendations

Findings

The aim of this systematic literature review was to establish whether CWMT improved WM scores for children who have SEN other than ADHD. Most of the studies found positive effect sizes showing the participants generally improved on measures of WM after the intervention. Although, the effect sizes were mixed and some studies found small negative effects. All of the studies were underpowered and therefore may have failed to detect, or underrepresented, an effect that was present. When looking at the effect size table, a number of the studies found stronger effects on measures of visual STM and WM. The average visual STM and WM effect size was 0.52, whereas the average verbal STM and WM was 0.28. The small and negative effect sizes found in Løhaugen et al., (2011) and Holmes et al., (2015) could suggest that children with low language abilities may be less responsive to CWMT as both the participant groups had low scores on language assessments.

Strengths of studies

The included studies all achieved high or medium methodological quality and overall weight of evidence scores. This shows that the majority of the included studies were methodologically sound and provided strong evidence to support their claims. Many of the studies used matched control groups, which adequately controlled for practice effects and extraneous variables.

A large proportion of the studies used valid reliable assessments that are commonly used to measure WM in research. The studies also frequently used more than one measure of WM, with two of the studies including measures that were completed by parents.

Weaknesses

One of the studies did not have a control group and used a within participants design which does not control for the practice effect and subsequently may have influenced the results. Another study did not use a matched control group and compared the performance of children with low language abilities to children with typical language skills. As a result, there was no matched control group for the children with low language abilities and the findings may have been influenced by extraneous variables.

None of the studies included an active control group and therefore some of the results may have been influenced by the Hawthorne effect. Because the participants were aware that they were receiving an intervention, this may have influenced their performance in the post-test assessment.

All of the included studies were underpowered for the effect size that is typically associated with Cogmed training. Many of the studies used very small sample sizes to rate the effectiveness of the intervention. The majority of the studies had between 9 and 20 participants in the intervention condition. The only study that had a large sample size was Kerr and Blackwell (2015) and even then, the sample was not big enough to meet statistical power for the analyses they used. This may have been partially due to adjusted alpha that was used due to multiple comparisons. Although small sample sizes and underpowered studies can be seen as a criticism for the included research, it is important to take into account that children with SEN are a vulnerable group and therefore obtaining large sample sizes will always be hard to achieve.

Practical Recommendations

CWMT is an effective intervention for improving WM scores for children who have SEN other than ADHD. Particularly children who have difficulties with their visuo-spatial memory, for instance children who struggle to remember how far things are away from them, get confused by maps or have difficulty learning how to use a keyboard.

Research recommendations

Further research could be used to determine whether there is a differential effect if CWMT is delivered in school or at home. Future studies should aim to include an active control group, for example if the control group were to receive a computer based literacy intervention and studies with larger sample sizes.

References

Atkins, T. (2014). *Are Working Memory Training programmes effective interventions for young people with ADHD?* (Doctoral Case study). Retrieved from http://www.ucl.ac.uk/educational-psychology/resources/CS1Atkins.pdf.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Barkley, R. a. (2013). Can Working Memory Training Ameliorate ADHD and Other Learning Disorders? A Systematic Meta-Analytic Review, 1–6.

Bavelier, D., Tomann, a, Hutton, C., Mitchell, T., Corina, D., Liu, G., & Neville, H. (2000). Visual attention to the periphery is enhanced in congenitally deaf individuals. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *20*, RC93.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, *41*(2), 257–278. http://doi.org/10.1111/j.2044-8317.1988.tb00901.x

Bennett, S. J., Holmes, J., & Buckley, S. (2013). Computerized memory training leads to sustained improvement in visuospatial short-term memory skills in children with down syndrome. *American Journal on Intellectual and Developmental Disabilities*, *118*(3), 179–192. http://doi.org/10.1352/1944-7558-118.3.179

Bezzola, L., Mérillat, S., & Jäncke, L. (2012). The effect of leisure activity golf practice on motor imagery: an fMRI study in middle adulthood. *Frontiers in Human Neuroscience*, *6*(March), 1–9. http://doi.org/10.3389/fnhum.2012.00067

Bruce, C., Edmundson, A., Aviet, C., & Willison, L. (2010). Exploring the effects of brain-training exercises on the cognitive and linguistic skills of adults with aphasia. *Procedia - Social and Behavioral Sciences*, *6*, 244–245. http://doi.org/10.1016/j.sbspro.2010.08.122

Cogmed UK Official Minisite. (n.d.). Retrieved December 29, 2015, from http://www.pearsonclinical.co.uk/Cogmed/Cogmed-Working-Memory-Training.aspx

"Cogmed Working Memory Training | History". (2015). Retrieved December 29, 2015, from http://www.cogmed.com/history

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. http://doi.org/10.1037/0033-2909.112.1.155

Cowan, N. (2014). Working Memory Underpins Cognitive Development, Learning, and Education. *Educational Psychology Review*, *26*(2), 197–223. http://doi.org/10.1007/s10648-013-9246-y

Eriksson, J., Vogel, E. K., Lansner, A., Bergström, F., & Nyberg, L. (2015). Neurocognitive Architecture of Working Memory. *Neuron*, *88*(1), 33–46. http://doi.org/10.1016/j.neuron.2015.09.020

Gathercole, S. E., & Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology*, *70*(2), 177–194. http://doi.org/10.1348/000709900158047

Gough, D. (2007). Weight of Evidence : a Framework for the Appraisal of the Quality and Relevance of Evidence, 213–228.

Grant, N. I. E., Daneman, M., & Carpenter, P. a. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. http://doi.org/10.1016/S0022-5371(80)90312-6

Grunewaldt, K. H., Skranes, J., Brubakk, A.-M., & Lähaugen, G. C. C. (2015). Computerized working memory training has positive long-term effect in very low birthweight preschool children. *Developmental Medicine and Child Neurology*, 1–7. http://doi.org/10.1111/dmcn.12841

Holmes, J., Butterfield, S., Cormack, F., van Loenhoud, A., Ruggero, L., Kashikar, L., & Gathercole, S. (2015). Improving working memory in children with low language abilities. *Frontiers in Psychology*, *6*(April), 519. http://doi.org/10.3389/fpsyg.2015.00519

Holmes, J., & Gathercole, S. E. (2013). Taking working memory training from the laboratory into schools. *Educational Psychology*, *3410*(March 2014), 1–11. http://doi.org/10.1080/01443410.2013.797338

Hulme, C., & Melby-Lervåg, M. (2012). Current evidence does not support the claims made for CogMed working memory training. *Journal of Applied Research in Memory and Cognition*, *1*(3), 197–200. http://doi.org/10.1016/j.jarmac.2012.06.006

Kerr, E. N., & Blackwell, M. C. (2015). Near-transfer effects following working memory intervention (Cogmed) in children with symptomatic epilepsy: An open randomized clinical trial. *Epilepsia*, *56*(11), 1784–1792. http://doi.org/10.1111/epi.13195

Kronenberger, W., & Pisoni, D. (2011). Working memory training for children with cochlear implants: A pilot study. *Journal of Speech, , Language, and Hearing Research*, *54*(August), 1182–1197. http://doi.org/10.1044/1092-4388(2010/10-0119)1182

Kyttala, M., Aunio, P., & Hautamaki, J. (2010). Working memory resources in young children with mathematical difficulties. *Scandinavian Journal of Psychology*, *51*(1), 1–15. http://doi.org/10.1111/j.1467-9450.2009.00736.x

Løhaugen, G. C. C., Antonsen, I., Håberg, a, Gramstad, a, Vik, T., Brubakk, a M., & Skranes, J. (2011). Computerized working memory training improves function in adolescents born at extremely low birth weight. *Journal of Pediatrics*, *158*(4), 555–561.e4. http://doi.org/10.1016/j.jpeds.2010.09.060

Morris, S. B. (2008). Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. *Organizational Research Methods*, *11*(2), 364–386. http://doi.org/10.1177/1094428106291059

Owen, a M., Morris, R. G., Sahakian, J. L., Polkey, C. E., Robbins, T. W., & Sahakian, B. J. (1996). Double dissociations of memory and executive\nfunctions in working memory task following frontal lobe\nexcision, temporal lobe excisions or amygdala-hippocampectomy in man. *Brain*, *119*, 1597–1615. http://doi.org/10.1093/brain/119.5.1597

Pantev, C., Paraskevopoulos, E., Kuchenbuch, A., Lu, Y., & Herholz, S. C. (2015). Musical expertise is related to neuroplastic changes of multisensory nature within the auditory cortex. *European Journal of Neuroscience*, *41*(5), 709–717. http://doi.org/10.1111/ejn.12788

Pearson. (2010). Pearson announces Cogmed now part of Pearson Clinical Assessment – Linking Intervention to Assessment [Press release]. Retrieved December 29, 2015, from http://www.cogmed.com/pearson-announces-cogmed-part-pearson-clinical-assessment-linking-intervention-assessment

Pearson Inc. (2014). Cogmed Factsheet [Brochure]. Author. Retrieved December 29, 2015, from http://www.pearsonclinical.co.uk/Education/BestsellingInterventions/Cogmed/Resources/Cogmed-Factsheet.pdf .

Shinaver, C. S., Entwistle, P. C., & Söderqvist, S. (2014). Cogmed WM training: reviewing the reviews. *Applied Neuropsychology. Child*, *3*(February 2015), 163–72. http://doi.org/10.1080/21622965.2013.875314

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*(4), 628–654. http://doi.org/10.1037/a0027473\r10.1037/0033-2909.131.1.30; Aiken, L.S., West, S.G., Invalidity of true experiments: Self-report pretest biases (1990) Evaluation Review, 14, pp. 374-390. , doi:10.1177/ 0193841X9001400403; Alloway, T.P., Can interactive working memory training improving learning? Journal of Interactive Learning Research, , (in press); Alloway, T.P., Alloway, R.G., Jungle memory training program (2008), Edinburgh, United Kingdom: MemosyneAlloway, T.P., Gathercole, S.E., Kirkwood, H., Elliot, J.

Smith, E. E., & Jonides, J. (1998). Neuroimaging analyses of human working memory. *Proc. Natl. Acad. Sci. USA*, *95*(20), 12061–12068. http://doi.org/VL - 95

Spencer-Smith, M., & Klingberg, T. (2015). Benefits of a Working Memory Training Program for Inattention in Daily Life: A Systematic Review and Meta-Analysis. *Plos One*, *10*(3), e0119522. http://doi.org/10.1371/journal.pone.0119522

Willis, S. L., Tennstedt, S. L., Marsiske, M., Ball, K., Elias, J., Koepke, K. M., … Wright, E. (2006). Long-term effects of cognitive training on everyday functional outcomes in older adults. *JAMA: The Journal of the American Medical Association*, *296*(23), 2805–2814. http://doi.org/10.1001/jama.296.23.2805

Appendices

- Appendix A; Excluded studies
- Appendix B; Discarded items from Kratochwill coding protocol
- Appendix C; Woe B criteria and scoring
- Appendix D; Woe C criteria and scoring
- Appendix E; Bennett et al., (2013) Kratochwill protocol
- Appendix F; Kratochwill protocol scoring table

Appendix A; Excluded studies

Study	Rationale for exclusion
Bergman-Nutley, S., & Klingberg, T. (2014). Effect of working memory training on working memory, arithmetic and following instructions. *Psychological Research*, *78*(6), 869–877. http://doi.org/10.1007/s00426-014-0614-0	Due to Rejection Criteria 2b - Study states that ‘A majority of subjects were diagnosed with ADHD’
Dahlin, K. I. E. (2011). Effects of working memory training on reading in children with special needs. *Reading and Writing*, *24*(4), 479–491. http://doi.org/10.1007/s11145-010-9238-y	Due to Rejection Criteria 2b - Study states that more than 60% of children had ‘attention problems’ and the analysis was based on a control group that consisted of 25 children diagnosed with ADHD
Dias, N. M., & Seabra, A. G. (2015). Is it possible to promote executive functions in preschoolers? A case study in Brazil. *International Journal of Child Care and Education Policy*, *9*(6). http://doi.org/10.1186/s40723-015-0010-2	Due to Rejection Criteria 4
Grunewaldt, K. H., Lohaugen, G. C. C., Austeng, D., Brubakk, a M., & Skranes, J. (2013). Working memory training improves cognitive function in VLBW preschoolers. *Pediatrics*, *131*(3), e747–e754. http://doi.org/10.1542/peds.2012-1965	Study was a duplicate of an included study
Pascoe, L., Roberts, G., Doyle, L. W., Lee, K. J., Thompson, D. K., Seal, M. L., … Anderson, P. J. (2013). Preventing academic difficulties in preterm children: a randomised controlled trial of an adaptive working memory training intervention - IMPRINT study. *BMC Pediatrics*, *13*(1), 144. http://doi.org/10.1186/1471-2431-13-144	Was not a completed research article. The publication was providing details of ongoing research, but no results had been published at the time of writing the review.
Titz, C., & Karbach, J. (2014). Working memory and executive functions: effects of training on academic achievement. *Psychological Research*, *78*(6), 852–868. http://doi.org/10.1007/s00426-013-0537-1	Due to Rejection Criteria 3

Appendix B; Discarded items from Kratochwill coding protocol

Discarded item	Rationale
A1. Theoretical basis**.** Intervention design is based upon a conceptual model that is grounded in theory and applied to the empirical study of the target phenomenon and group.	In this review the intervention is always the same and therefore the design does not differ per publication
A4. Research question guides selection of qualitative, quantitative or mixed methodology of study.	All included studies were quantitative in design
A5.1. Participatory (Research Questions): Involvement of participants in research design. Were relevant target group members consulted in the formulation of the research questions?	As the target group were school age children with SEN, these items were not seen as applicable.
A5.2. Participatory (Methodology**):** Involvement of participants in research methods: Were relevant target group members involved in the formulation of the research methods?	As the target group were school age children with SEN, these items were not seen as applicable
A5.3. Participatory (Implementation): Involvement of participants in implementation of intervention: Were relevant target group members involved in the development and delivery of intervention?	Not applicable as intervention was pre-defined. Those involved in delivering the intervention will be discussed within the review.
A5.4. Participatory (Negotiated Interpretation): Involvement of participants in interpretation of findings: Were relevant target group members involved in the interpretation of the research findings?	As the target group were school age children with SEN, these items were not seen as applicable.
Research Methodology	
C7. Coding	No qualitative data analysis were used in any of the included studies
C8. Interactive process followed	No qualitative data analysis were used in any of the included studies
A1. Characteristics of the data collector	Individuals who delivered the intervention will be discussed at a

Discarded item	Rationale
	different point in the review
A.2. Characteristics of Participants	Cultural differences were not discussed in any of the included research and the characteristics of the participants will be discussed in a different section of the review
B4. Extent of Engagement--The researchers conduct data collection in a manner that guarantees sufficient scope and depth through prolonged engagement (data collection over a sufficient time period to ensure accuracy of representation) and persistent observation (progressively focused to ensure thorough understanding of consistency and variation), respectively.	As the intervention has a defined start and end point, this is not relevant to any of the included research. The extent to which the participants completed the training will be discussed at another point in the review.
B6. Cultural Appropriateness of the Measures. In rating this item, consider the following dimensions: meaning, language, dialect, and response format.	None of the included studies adapted or developed measures to assess outcomes.
D3. Evidence of appropriate statistical analysis for secondary outcomes	This review aimed to establish whether the intervention improved one outcome (WM) and therefore secondary outcomes did not need to be analysed or included in any part of the review.
D6. Cultural Moderator Variables. Investigated effects of cultural moderating variables on the outcome.	None of the included studies investigated cultural moderating variables.
D Table	All of the information required in this table will be discussed in more detail within the review
F Table rows	
F1. Categorical Diagnosis Data	The intervention was not intended to reduce the categorical diagnosis of the participants.
F3. Subjective Evaluation	None of the studies used subjective

Discarded item	Rationale
	evaluation to establish differences in WM scores. Participants' behaviour was not the focus of the intervention The characteristics of the participants in the included studies will be discussed in detail in a different section of this review
F4. Social Comparison:	The intervention was not intended to reduce the categorical diagnosis of the participants.
G2. Participant Characteristics Specified for Treatment and Control Group	The characteristics of the participants in the included studies will be discussed in detail in a different section of this review
G5. Transferability of the intervention.	This will be discussed in detail in a different section of this review
G6. Participant perceptions of benefits of intervention (treatment group)	WM improvements were assessed using standerdized measures and None of the included studies sought the participants' perceptions of the benefits of the intervention
H1. Follow-up assessment	The timings of the follow up assessments will be discussed in detail in a different section of the review
H2. Durability/Generalization over time	This will be discussed in detail in a different section of this review
H2.1.1 Evidence is provided regarding the sustainability of outcomes after intervention is terminated	This is not the focus of the review question

Discarded item	Rationale
H3. Durability/Generalization across settings	This will be discussed in detail in a different section of this review
H3.1 Evidence is provided regarding the extent to which outcomes are manifested in contexts that are different	This is not the focus of the review question
H4. Generalization across persons	This will be discussed in detail in a different section of this review
I. Identifiable Intervention Components (answer I1 through I7)	As intervention was computer based, it could nto be adapted to any significant degree. The implementation of the intervention will be discussed during a different section of the review.
J Implementation fidelity	This entire section was not included as the implementation of the intervention will be discussed in more detail in a different section of the review.
K. Replication (answer K1, K2, K3, and K4)	None of the studies were replications
L. Site of Implementation	This entire section was not included as the implementation of the intervention will be discussed in more detail in a different section of the review.

Appendix C; Woe B criteria and scoring

The included studies were assigned a score of one to three rated according to the following criteria.

Study	**WoE B score**
(Bennett et al., 2013)	3

High (3)

- Participants must be randomly allocated to intervention or control
- Control group must be an active control group that;
  - Receive some type of alternative intervention
  - Are matched according to SEN
- Pre and post measures provided for intervention and control group
- WM was assessed using a standardised psychometric assessment

Medium (2)

- Must have a control group
- Pre and post measures provided for intervention and control group

Low (1)

- Pre and post measures provided

(Grunewaldt et al., 2015)	2
(Holmes et al., 2015)	1
(Kerr & Blackwell, 2015)	3
(Kronenberger & Pisoni, 2011)	1
(Løhaugen et al., 2011)	2

Appendix D; WoE C criteria and scoring, and WoE D rationale

The score for WoE C was calculated from an average of three scores based on different criteria. The scoring criteria are listed below and the individual and average scores are included in the table below.

Score	Intervention delivered	Setting	Supervised
3	Full training delivered	In school	Supervised by trained adult
2	20+ training delivered	At home or other premises	Supervised by an adult (training non specified)
1	No information provided	No information provided	No information provided

Study	WoE C score	Intervention delivered	Setting	Supervised
(Bennett et al., 2013)	3.00	3	3	3
(Grunewaldt et al., 2015)	1.67	3	1	1
(Holmes et al., 2015)	2.33	2	3	2
(Kerr & Blackwell, 2015)	2.67	3	2	3
(Kronenberger & Pisoni, 2011)	2.33	2	2	3
(Løhaugen et al., 2011)	2.67	3	2	3

WoE D rationale

The Kratochwill protocol could have given studies a minimum score of zero and a maximum of three. WoE B and C could have given a minimum score of one and a maximum score of three. As a result the maximum a study could have scored for WoE D would be nine (three for each criteria) and the minimum would be two (zero for WoE A and one for WoE B and C). This meant that the studies scores had a range of seven. The range of seven was divided in three to establish studies that would be deemed Low, Medium and High. The cut-offs for each descriptor are outlined in the table below.

Descriptor	WoE D score	
	Minimum	Maximum
High	6.67	9
Medium	4.34	6.66
Low	2	4.33

Appendix E; Bennett et al., (2013) Kratochwill protocol

[adapted from Task Force on Evidence-Based Interventions in School Psychology, American Psychology Association, Kratochwill, T.R. (2003)]

Coding Protocol

Name of Coder: XXXX Date:25/01/2016

Full Study Reference in proper format:

Bennett, S. J., Holmes, J., & Buckley, S. (2013). Computerized memory training leads to sustained improvement in visuospatial short-term memory skills in children with down syndrome. *American Journal on Intellectual and Developmental Disabilities*, *118*(3), 179–192. http://doi.org/10.1352/1944-7558-118.3.179

Intervention Name (description of study):Cogmed

Study ID Number:01

☐Type of Publication:
☐Book/Monograph
☒Journal Article
☐Book Chapter
☐Other (specify):

1.General Characteristics

A. General Study Characteristics

A2. Ecological validity of the empirical-theoretical (conceptual) basis for the target population has been established.

☐ Ecological validity of conceptual base established through formative research with target population OR conceptual base is culturally derived
☐ Existing theory adapted based on research with target population
☒ Existing theory adapted based on empirical evidence with similar or related populations
☐ Existing theory used without attention to ecological validity

A3. Researcher perspective of the empirical-theoretical basis for the target population

☐ Researcher articulated personal beliefs and relation to the conceptualization of the study
☐ Research articulated personal beliefs with vague link to conceptualization
☐ Research articulated personal beliefs with no link
☒ No discussion of researcher's beliefs related to study's conceptualization

B. General Design Characteristics

B1. Random assignment designs (if random assignment design, select one of the following)
☒ Completely randomized design
☐ Randomized block design (between participants, e.g., matched classrooms)
☐ Randomized block design (within participants)
☐ Randomized hierarchical design (nested treatments

B2. Nonrandomized designs (if non-random assignment design, select one of the following)
☐ Nonrandomized design
☐ Nonrandomized block design (between participants)
☐ Nonrandomized block design (within participants)
☐ Nonrandomized hierarchical design
☐ Optional coding for Quasi-experimental designs

B3. Overall confidence of judgment on how participants were assigned (select on of the following)
☐ Very low (little basis)
☐ Low (guess)
☐ Moderate (weak inference)
☐ High (strong inference)
☒ Very high (explicitly stated)
☐ N/A
☐ Unknown/unable to code

C Participants

C1. Appropriate unit of analysis ☒Yes ☐ No
C2. Familywise error rate controlled ☐Yes ☒ No
C3. Sufficiently Large N ☐Yes ☒ No

Statistical Test: Within group t-test
α level: .05
ES: 0.5
N required: 54

Statistical Test: 2x2 ANOVA
α level: .05
ES: 0.25

N required: 54

C4. Total size of sample (start of study): 25
C5. Intervention group sample size: 10
C6. Control group sample size: 12

C9. Rival Interpretations. Evaluates rival cultural, methodological, statistical, or theoretical explanations and hypotheses.

☐ Investigated multiple rival hypotheses (e.g., cultural, methodological, statistical, and/or theoretical explanations) and provided empirical evidence
☒ Investigated one rival hypothesis (e.g., cultural, methodological, statistical, and/or theoretical explanations) and provided empirical evidence
☐ Considered rival explanations without empirical evidence
☐ No evidence that rival hypotheses were investigated or considered (examined data only from one perspective)

D. Type of Program
☐ Universal prevention program
☐ Selective prevention program
☐ Targeted prevention program
☒ Intervention/Treatment
☐ Unknown

E. Stage of Program
☐ Model/demonstration programs
☐ Early stage programs
☒ Established/institutionalized programs
☐ Unknown

F. Concurrent or Historical Intervention Exposure
☐ Current exposure
☐ Prior exposure
☒ Unknown

II. Key Features for Coding Studies and Rating Level of Evidence/ Support

: ☐ 3 =Strong Evidence ☒ 2=Promising Evidence ☐ 1 =Weak Evidence ☐ 0=No Evidence

Methodology (answer A1 through A5)

A3. Sample appropriate to research methods. Research methods guide sampling procedures.

☒ Clear links established between research methods and sampling, and sampling is *appropriate* to the research methods
☐ Vague or no links established between research methods and sampling, but sampling is *appropriate* to the research methods
☐ Links established between research method and sampling, but sampling is inappropriate to the research

A4. Operationalization. Specifying the link between key abstract constructs (variables) and data collection methods (operations).

☒ Clear links established between constructs and methods, **and** all key constructs are clearly operationalized.
☐ Some, but not all, key constructs are clearly operationalize
☐ Vague reference to link between constructs and methods.
☐ No evidence that key constructs are operationalized.

A5. Integration of data from multiple sources, methods, and investigators

☐ Used multiple sources, methods, and investigators.
☐ Used two of the following: multiple sources, multiple methods, multiple investigators
☒ Used one of the following: multiple sources, multiple methods, multiple investigators
☐ No evidence of multiple sources, methods, or investigators

A. OVERALL Rating for Research Methodology ☒ 3 ☐ 2 ☐ 1 ☐ 0

B. Measurement (answer B1 through B6)

B1 The use of the outcome measures produce reliable scores for the majority of the primary outcomes (see following table for a detailed breakdown on the outcomes)
☐Yes
☐No
☒Unknown/unable to code

B2 Multi-method (at least two assessment methods used)
☒Yes
☐No
☐N/A
☐Unknown/unable to code

B3 Multi-source (at least two sources used self-reports, teachers etc.)
☒Yes
☐No
☐N/A
☐Unknown/unable to code

B5 Validity of measures reported (well-known or standardized or norm-referenced are considered good, consider any cultural considerations)
☐Yes validated with specific target group
☒In part, validated for general population only
☐No
☐Unknown/unable to code

B7 Measures of key outcomes are linked to the conceptual model.
☐ Clear links established between the conceptual model and key outcome indicators
☒ Some, but not all, key outcomes are clearly linked to conceptual model.
☐ Vague reference to links between key outcomes and conceptual model
☐ No evidence that key outcomes are linked to conceptual model.

Overall Rating of Measurement: ☐ 3 ☒ 2 ☐ 1 ☐ 0

Comparison Group

C Comparison Group

C1 Type of Comparison group
- ☐Typical intervention
- ☐Attention placebo
- ☐Intervention element placebo
- ☐Alternative intervention
- ☐Pharmacotherapy
- ☐No intervention
- ☒Wait list/delayed intervention
- ☐Minimal contact
- ☐Unable to identify type of comparison

C2 Overall rating of judgment of type of comparison group
- ☐Very low
- ☐Low
- ☐Moderate
- ☐High
- ☒Very high
- ☐Unable to identify comparison group

C3 Counterbalancing of change agent (participants who receive intervention from a single therapist/teacher etc were counter-balanced across intervention)
- ☐ By change agent
- ☐ Statistical (analyse includes a test for intervention)
- ☐ Other
- ☒ Not reported/None

C4 Group equivalence established

- ☐ Random assignment
- ☐ Posthoc matched set
- ☒ Statistical matching
- ☐ Post hoc test for group equivalence

C5 Equivalent mortality

- ☒ Low attrition (less than 20 % for post)
- ☐ Low attrition (less than 30% for follow-up)
- ☐ Intent to intervene analysis carried out?

Findings. 4 children were not included in the final analysis; one couldn't access the intervention, one did not complete the training and the other 2 were not included because their school had technical problems.

25/4 =16%

Overall Rating of Comparison Group: ☒ 3 ☐ 2 ☐ 1 ☐ 0

D. Primary Outcomes (WM) Are Statistically Significant

D1. Analysis 1 – t tests

- ☒ Appropriate unit of analysis
- ☐ Familywise/experimenter wise error rate controlled when applicable
- ☐ Sufficiently large N **No – 54 participants needed for a within group t-test with an effect size of 0.5**

D1. Analysis 2 – ANOVA

- ☒ Appropriate unit of analysis
- ☐ Familywise/experimenter wise error rate controlled when applicable
- ☐ Sufficiently large N **No – 54 participants needed for a 2x2 ANOVA with a medium effect size of 0.25**

D2 Percentage of **outcomes** that are significant (select one of the following)

- ☐ Significant primary outcomes for at least 75% of the total primary outcome measures for each key construct Significant primary outcomes for at least 75% of the total primary outcome measures for each key construct
- ☒ Significant primary outcomes for between 50% and 74% of the total primary outcome measures for each key construct
- ☐ Significant primary outcomes for between 25% and 49% of the total primary outcome measures for any key construct
- ☐ Significant primary outcomes for between 25% and 49% of the total primary outcome measures for any key construct

D. Secondary Outcomes (executive functioning) Are Statistically Significant

D2. Analysis 1 – t tests

- ☒ Appropriate unit of analysis
- ☐ Familywise/experimenter wise error rate controlled when applicable

☐ Sufficiently large N **No – 54 participants needed for a within group t-test with an effect size of 0.5**

D2. Analysis 2 – ANOVA

☒ Appropriate unit of analysis
☐ Familywise/experimenter wise error rate controlled when applicable
☐ Sufficiently large N **No – 54 participants needed for a 2x2 ANOVA with a medium effect size of 0.25**

D2 Percentage of **outcomes** that are significant (select one of the following)

☐ Significant primary outcomes for at least 75% of the total primary outcome measures for each key construct Significant primary outcomes for at least 75% of the total primary outcome measures for each key construct
☐ Significant primary outcomes for between 50% and 74% of the total primary outcome measures for each key construct
☒ Significant primary outcomes for between 25% and 49% of the total primary outcome measures for any key construct
☐ Significant primary outcomes for between 25% and 49% of the total primary outcome measures for any key construct

Overall Rating of Analysis: ☐ 3 ☐ 2 ☒ 1 ☐ 0

D.5 Overall Summary of Questions Investigated

Main effect analyses conducted Y ☒ N ☐
Specify results –
Significant improvement in pre-post intervention scores for Visuo-Spatial WM and Visuo Spatial STM subtests of the AWMA.

Significant decrease in scores of WM and Shift componenets of the BRIEF.

Moderator effect analyses conducted Y ☐ N ☒
Specify results -
Mediator effect analyses conducted Y ☐ N ☒
Specify results -
Unintended outcomes assessed Y ☒ N ☐
Specify results –
Gains in AWMA scores were not related to the number of tasks completed.
Improvement in AWMA scores were not related to number of tasks completed (although it did approach significance)
Improvement in Cogmed tasks was significantly related to baseline verbal IQ

D. OVERALL Rating for Primary/Secondary Outcomes (select 0, 1, 2, or 3): ☒ 3 ☐ 2 ☐ 1 ☐ 0

Outcome Variables:	Pretest	Posttest			Follow Up		
		Clinically significant improvement from pre-test to post-test:			Clinically significant improvement from post-test to follow up:		
		Y	N	Unknown	Y	N	Unknown
AWMA - Verbal STM			X			X	
AWMA - Verbal WM			X			X	
AWMA – Visuospatial STM		X				X	
AWMA – Visuospatial WM		X				X	
BRIEF - Inhibition			X			X	
BRIEF - Shift		X				X	
BRIEF – Emotional control			X			X	
BRIEF - WM		X				X	
BRIEF – planning organise			X			x	

G. External Validity Indicators

G1. Sampling Procedures (Answer G1.1 through G1.4)

G1.1 Sampling procedures described in detail
☒ Yes
☐ No (incomplete or no evidence)

G1.2. Rationale for sample selection specified [integrate current A1.1-A1.3, & A.3]
☒ Yes
Specify -
☐ No (incomplete or no evidence)

G1.3. Rationale for sample size specified
☐ Yes
Specify -
☒ No (incomplete or no evidence)

G1.4. Evidence provided that sample represents target population
☐ Yes
Specify -
☒ No (incomplete or no evidence)

G1.5 Recruitment procedures congruent with target cultural group. Researcher used culturally appropriate ways/methods to contact, recruit, inform, and maintain participation.

☒ Yes
Specify -
☐ No (incomplete or no evidence)

G1.6. Inclusion/exclusion criteria specified
☒ Yes
☐ no

G1.7. Inclusion/exclusion criteria similar to school practice
☐ Yes
☐ No

G1.8. Specified criteria related to concern
☒ Yes
☐ no

G1. Rating for Sampling (select 0, 1, 2, or 3): ☐ 3 ☒ 2☐ 1 ☐ 0

G3. Adequately reported characteristics of participants/sample. Adequate level of detail in <u>description</u> of participants.
☒ Yes
Specify -
☐ No (incomplete or no evidence)

G4. Details are provided regarding variables that:
G4.1 Have differential relevance for intended outcomes
☒ Yes
Specify -
☐ No (incomplete or no evidence)

G4.2 Have relevance to inclusion criteria
☒ Yes
Specify -
☐ No (incomplete or no evidence)

G. OVERALL Rating for External Validity (select 0, 1, 2, or 3): ☒3 ☐2 ☐1 ☐ 0

Appendix F; Kratochwill protocol scoring table

To produce a score for WoE A, the average score from each of the components of the Kratochwill coding protocol was used. The scores for each section are presented below with the average.

	WoE A	Key features	Methodol ogy	Measure ment	Comparis on group	Analyses	Outcome s	Sampling	External validity
(Bennett et al., 2013)	2.50	2	3	2	3	2	3	2	3
(Grunewaldt et al., 2015)	2.63	2	3	3	3	2	3	2	3
(Holmes et al., 2015)	2.38	2	2	2	2	2	3	3	3
(Kerr & Blackwell, 2015)	2.75	2	3	3	3	2	3	3	3
(Kronenberger & Pisoni, 2011)	2.00	1	3	3	0	2	2	2	3
(Løhaugen et al., 2011)	2.00	2	1	2	1	2	3	2	3

Appendix G; Summary of studies table

Study	Implementation	Design	Participants: Total	Participants: Intervention	Participants: Control	Measures	Results
(Bennett et al., 2013)	Every participant completed 25 days of training	Between participants random allocation	25 children with Down Syndrome.	N=10 Age M= 9.46 SD = 18.09 mths	N=11 Age M= 9.47 SD = 26.75 mths Waiting list	AWMA – Verbal STM – Verbal WM – Visuo-spatial STM – Visuo-spatial WM BRIEF – WM	Improvements in visuo-spatial working and short term memory assessments
(Grunewaldt et al., 2015)	Every participant completed 25 days of training	Stepped wedge	Children that were born at a very low birthweight and were admitted to a particular neonatal intensive care unit between the years of 2005 to 2006.	N =20 Age M= 5.8 SD = 0.49 mths	N =17 Age M= 5.4 SD = 0.29 mths Long term effects determined using participants admitted to the same neonatal intensive care unit the year after the intervention group	Wechsler memory scale – Visual WM – Spatial Span forwards – Spatial Span Backwards – Spatial Span total WISC – Verbal WM – Digit span forwards – Verbal WM – Digit span backwards	Improvements in WM and visual and verbal learning

Study	Implementation	Design	Participants: Total	Participants: Intervention	Participants: Control	Measures	Results
(Holmes et al., 2015)	Every participant completed 20 sessions of training.	Between participants	12 children with standard scores below 86 on a vocabulary test and scaled scores of below 7 on a recalling sentences subtest.	N=12 Age M=9:9 SD=8.4 mths.	N = 15 Age M=9:9, SD=9.5 mths. Matched by: – age – gender – non verbal reasoning scores No intervention.	AWMA – Verbal STM – Visuo-Spatial STM – Verbal WM – Visuo-Spatial WM	Improvements in visuo-spatial STM scores Children with low language scores, showed more improvements in verbal STM. Children with high verbal IQs showed more improvements on visuo-spatial measures.
(Kerr & Blackwell, 2015)	Every participant completed 25 sessions of training.	Between participants, random allocation	Convenience sample of children admitted to epilepsy clinics or through referral from an epilepsy association.	N= 35 Age M=10.6 SD= 2.7mths	N= 35 Age M=11.1 SD= 3.0mths No intervention	WM Test battery for children – Basic auditory WM – Auditory WM – Visual-Verbal WM – Counting recall WISC IV – Visual Attention – Visual WM Test of everyday attention for children Visual	Intervention group showed improvement on scores of visual attention span, auditory WM and visuo-verbal WM.

Study	Implementation	Design	Participants: Total	Participants: Intervention	Participants: Control	Measures	Results
						– Verbal WM Attentional capacity test – Visual-Verbal WM	
(Kronenberger & Pisoni, 2011)	6 participants completed 25 days of training. Other participants completed between 20 and 23 days of training.	Within participants	Participants had severe or bilateral hearing loss from birth and to have received a cochlear implant before the age of 3.	N= 9 Age M=10.2 SD= 2.2 mths	N/A	WISC IV – Immediate verbal phonological memory – Immediate verbal WM capacity – Visual-spatial sequential memory capacity – Visual-spatial WM BRIEF – WM	Participants showed improvements on measures of verbal WM, non-verbal WM, parent reported WM and sentence repetition skills.
(Løhaugen et al., 2011)	Every participant completed 25 sessions of training.	Between participants	Children admitted to a particular neonatal intensive care unit between 1992 and 1993.	N= 16 Age M=14.1 SD= 0.6 mths	N= 19 Age M=14.3 SD= 0.7 mths Children born at healthy birthweight No intervention	Wechsler memory scale – Digit Span – Letter number sequencing – Spatial Span	Intervention group improved on WM tasks. Children with low IQs showed more improvements.