Case Study 1: An Evidence-Based Practice Review Report

Theme: School (Setting) based interventions for children with Special Educational Needs (SEN)

How Effective is Computer-Based Working Memory Training (CWMT) at Reducing Behavioural Symptoms of Attention Deficit Hyperactivity Disorder (ADHD) in Children?

1.0. Summary

Computer-based working memory training (CWMT) has evolved since Klingberg et al.'s (2002) landmark feasibility study into its use. CWMT aims to reduce the symptoms of Attention Deficit Hyperactivity Disorder (ADHD) by increasing children's working memory (WM) capacity.

Children with ADHD have been associated with phonological and visuospatial WM deficits located in specific areas of the pre-frontal cortex and parietal lobe (Silk et al., 2008; Olesen et al., 2004). WM capacity was previously thought to be a fixed entity (Lieberman, 2012), but advances in technology, particularly functional magnetic resonance imaging (fMRI), have demonstrated plasticity in WM functioning rendering it potentially susceptible to training (Olesen et al., 2004; Klingberg, 2010).

Theoretical models of ADHD indicate that CWMT should reduce child symptoms of ADHD by increasing WM capacity, resulting in increased information processing power and executive function (EF) (Klingberg, 2010; Sonuga-Barke et al., 2014). These are skills which enhance success at school in terms of children being able to focus during lessons, sit relatively still without fidgeting, achieve academically and conform to expected behavioural norms (Gwernan-Jones et al., 2016). However, these theoretical models have been very difficult to evidence empirically.

Despite this lack of empirical evidence, the market for CWMT has grown significantly over a relatively short period of time. Several international CWMT companies have emerged targeting children with ADHD, with Cogmed (https://www.cogmed.com/) developed by Torkel Klingberg, "the most widely used software training program" (Simons et al., 2016, p114). This has led to researchers striving to obtain conclusive empirical evidence to support claims that CWMT increases WM capacity for children with ADHD. A systematic literature review (SLR) was inconclusive and raised questions regarding the impact of inflated ratings due to ineffective blinding of participants (Sonuga-Barke et al., 2014). A subsequent meta-analysis addressed this concern by analysing the scores of 'typically unblinded' and 'probably blinded' raters separately, and concluded that "cognitive training had limited effects on ADHD symptoms according to assessments based on blinded trials" (Cortese et al., 2015, p164). However, neither of these reviews were exclusively focused on behavioural symptoms of ADHD, which is unique to this review.

The aim of this review was to conclusively ascertain the impact of CWMT on the behavioural symptoms of children with ADHD, as it is these symptoms that are most disruptive to children's school experience. A focused approach was adopted to source a homogenous group of studies. Consequently, this review only included six studies, all of which were randomised controlled trials (RCTs) and used the Cogmed CWMT program.

The findings of this review indicated that CWMT improved elements of real-world EF but had little or no effect on the wider problematic behavioural symptoms of children with ADHD, probably due to EF being one of many endophenotypes of ADHD. The main limitations of this review were the heterogenous nature of participant's ADHD status, the prevalence of under-powered studies, and variable primary outcome measures. This led to the recommendation that further research either focuses on improving the quality of RCTs in this area, or employs qualitative research and mixed methodologies to further explore this phenomenon, as these may be more effective and discerning than RCTs.

2.0. Introduction

2.1. ADHD

2.1.1. What is ADHD?

ADHD usually manifests in children between the ages of two and seven years old, and is characterised by inattentive behaviour, impulsivity, restlessness, poor social relationships, school under-achievement and a propensity towards immediate gratification (Moore et al., 2017; Taylor et al., 2004). It affects approximately three to five percent of children and two

percent of adults in the United Kingdom (UK), and is four times more prevalent in boys than girls (CAMHS Advisory Group, 2018).

Children with acute symptoms of ADHD are often prescribed mild psychoactive drugs which are effective for most in the short term (Faraone & Buitelaar, 2010). However, their long-term effects are unknown, there is an increased likelihood that these children will go on to abuse drugs and alcohol (Groenman et. al., 2013), and knowing the right time to stop medicating is problematic (van de Loo-Neus et al., 2011). Furthermore, the life-time prognosis for children with ADHD is concerning, as they exhibit higher incidences of "psychiatric outcomes including markedly elevated rates of antisocial, addictive, mood and anxiety disorders" (Biederman et al., 2006, p167). Consequently, there have been attempts to find non-pharmaceutical alternatives to treat children with ADHD, such as eliminating certain foods from their diet, behavioural and therapeutic interventions and CWMT (Sonuga-Barke et al., 2013), as well as multimodal treatments which combine pharmaceutical and non-pharmaceutical treatments (Murray et al., 2008).

2.1.2. Psychological Theory of ADHD

There is much that is unknown about ADHD, which is in part due to its heterogenous nature (Wahlstedt et al., 2009). Moreover, ADHD is often co-morbid with conditions such as Oppositional Defiant Disorder (ODD), emotional disorders or specific learning difficulties (Taylor et al., 2014), which further complicate diagnosis and assigning appropriate interventions.

Technological advances in fMRI mean that ADHD can also be defined in terms of neuropsychological impairments. Cognitive networks affected

include inhibitory EF, timing and working memory networks (Sonuga-Barke et al., 2014), although it should be noted that EF deficits are not consistently represented across ADHD populations (Barkley & Murphy, 2010). Motivational networks affected include intrinsic motivation and reward-punishment networks. Energetic processes, such as effort and arousal have also been implicated (Sonuga-Barke et al., 2014).

2.2. Working Memory

2.2.1. Psychological Theory of Working Memory

According to the Baddeley and Hitch model (1974; as cited in Lieberman, 2012), WM has three components; the phonological loop (e.g. for processing audio for about two seconds of spoken material), the visuo-spatial sketchpad (e.g. for processing images) and the central executive, which controls what information is processed, and combined, with information retrieved from long-term memory (LTM). The central executive coordinates information flow from the two lower components (Baddeley et al., 1991).

Thus, WM can be defined as "a store where we can combine information from the phonological loop, the visuo-spatial sketchpad and the long-term memory" (Lieberman, 2012, p332), under the control of the central executive. Put more simply, WM is the ability to hold and process information in the short term before integrating it with and committing it to LTM.

2.2.2. Working Memory and ADHD

The information stored in the WM is transient, and so it is sometimes referred to as short-term memory (STM). Vocabulary development (Gathercole et al., 1999) and early stage word learning (Jackson et al., 2016) were shown to be

strongly associated with phonological STM capacity. This has implications for children with ADHD and may explain why they often experience reading and academic difficulties.

Our visual-spatial memory of pictures is much more developed and effective than the phonological loop responsible for the memory of words. This is known as the 'picture superiority effect' (Defeyter et al., 2009) and may explain why children with ADHD have difficulties with mathematics, planning, deferred gratification and other abstract activities which involve visualisation.

Baddeley et al. (1991) provided evidence for the central executive controlling the phonological loop and the visuo-spatial sketchpad, with the central executive being required to coordinate activity between the two components of WM. Consistent with a central executive or EF deficit is that children with ADHD are often impulsive, inattentive and restless. Thus, deficits in all three components of WM link individually and collectively to symptoms of ADHD.

2.3. Computer-Based Working Memory Training (CWMT) Intervention

2.3.1. Psychological Theory of the Intervention

CWMT aims to take advantage of brain plasticity and increase WM capacity in children with ADHD using adaptive computer programs to either increase the efficiency of pre-cortical and parietal domains, or increase their neural volume. The theory is akin to working out in a gym to increase muscle strength: neural networks are more likely to develop using adaptive programs which respond to user input and maintain neural network operation at full capacity, rendering the need for them to either become more efficient or increase their volume to meet the demand.

According to the Baddeley and Hitch model (1974; as cited in Lieberman, 2012), with increased WM capacity, children with ADHD should have increased EF, phonological and visuospatial capabilities. This should result in better social and emotion regulation and thus relationships with others, increased focus in the classroom and overall improvements in behaviour.

2.3.2. The Basis for Change

Cogmed is a CWMT program consisting of brain-training computer games designed to increase children's WM capacity. The embedded activities include inhibition, visuospatial and forward and backward recall tasks (Klingberg et al., 2005). It is a five-week program involving five days training per week for approximately 40 minutes per day where the trainee completes 90 trials which focus on various components of working memory (Roche & Johnson, 2014). Children are supervised during training and so an adult, usually a parent, is also trained in software use. Internet access is required as training scores are uploaded to a server where the trainee's progress is monitored. Ideally, monitoring should be conducted by a Cogmed certified coach who not only trains the adult and the child, but who communicates with the adult at least weekly to ensure compliance and engagement with the programme, and addresses any technical issues or queries that may arise (Roche & Johnson, 2014).

The Cogmed programme is adaptive, and so the level of challenge adjusts in response to trainee input so that a significant level of challenge is maintained. This is to sustain motivation and optimise WM development. There are also reward games built into the programme that can be earned to

motivate trainees to complete sessions. Cogmed is considered suitable for children with ADHD because the adaptive nature, game element and embedded rewards are designed to engage children and maintain their focus, which is usually a challenge for children with ADHD, whilst increasing their WM capacity.

A 'non-adaptive' version of the Cogmed software also exists which can be used in research as an intervention element placebo. The non-adaptive version contains similar games, but the challenge is maintained at a low level, not increasing in response to the user's improved competence.

By changing brain physiology, the adaptive version of Cogmed aims to develop positive near- and far-transfer effects. The former describes transfer of trained WM capabilities to similar untrained tasks, such as visuospatial tasks, and the latter describes transfer of trained WM capabilities to dissimilar, untrained tasks, such as the behavioural symptoms of ADHD. These would manifest as increased focus in lessons, for example, being able to sit still for longer periods of time, reduced restlessness and impulsivity, and improved emotion regulation resulting in more positive relationships with others. It is these behavioural far-transfer effects that are the outcomes of focus for this review.

2.4. Rationale for Using Cogmed in Educational Psychology

Educational Psychologists (EPs) often work with children with ADHD. Cogmed is an intervention which parents could implement at home, which does not require medication nor intensive professional input. Furthermore, after the initial outlay for Cogmed software, it could prove to be cost effective.

Therefore, if Cogmed could effectively increase the WM capacity of children with ADHD and the far transfer effects improved their behavioural symptoms, and subsequently their academic performance, it could transform the lives of children with ADHD.

However, EPs should exercise caution and ensure that the evidence base for interventions is secure (Hague et al., 2020), and consistent with their professional guidelines (BPS, 2018; HCPC, 2016), before recommending them. With limited proven non-pharmaceutical options available for ADHD treatment, EPs should resist succumbing to pressure from parents and professionals, who may be seduced by the potential benefits of using Cogmed and similar CWMT programs, to endorse popular interventions at the expense of their suitability for particular children and how their ADHD manifests (Hague et al., 2020).

2.5. Review Question

Considering the above discourse, the review question is:

How effective is computer-based working memory training (CWMT) at reducing behavioural symptoms of children with Attention Deficit Hyperactivity Disorder (ADHD)?

3.0. Critical Review of the Evidence

3.1. Literature Search

On 12th February 2022, a Boolean literature search was conducted to maximise the number of peer reviewed articles accessed (Table 1) using the following databases: Education Resource Information Centre (ERIC

EBSCO), ProQuest, PubMed, PsychInfo and Web of Science Core Collection (WoS).

Six articles were retained for inclusion in the review from 227 results. The research journey was documented (Figure 1), and the inclusion and exclusion criteria summarised (Table 2). Proquest proved the least informative database, as none of the 20 articles extracted were included after screening by title. WoS proved the most lucrative database, as it yielded 38 articles for inclusion, after screening by title, out of 137 articles extracted.

Table 1

To show the databases, key words and phrases used in literature searches

Boolean Operator	Key Word or Phrase	Reason for Including Key Word or Phrase
	"working memory" AND computer* AND Training OR Program*	Intervention
AND	Child* OR adolescen* OR young OR teen* OR student OR pupil	Population
AND	ADHD OR "attention defici* hyperactiv* disorder" OR hyperkinetic	Outcome

If during screening there was any ambiguity regarding the exclusion of an article, the article was included and scrutinised more closely in the next iteration. This explains why some articles that were excluded at full text screening could arguably have been excluded earlier. When articles were screened at the abstract level, it became apparent that there were sufficient articles to only consider RCTs, so a new exclusion criterium was added such that only RCTs were reviewed. Ten RCTs remained after screening by abstract which were reduced to six after full-text screening (Table 3). Details of the four excluded articles were summarised (Appendix A).

Figure 1

PRISMA flowchart to summarise the screening process for articles reviewed

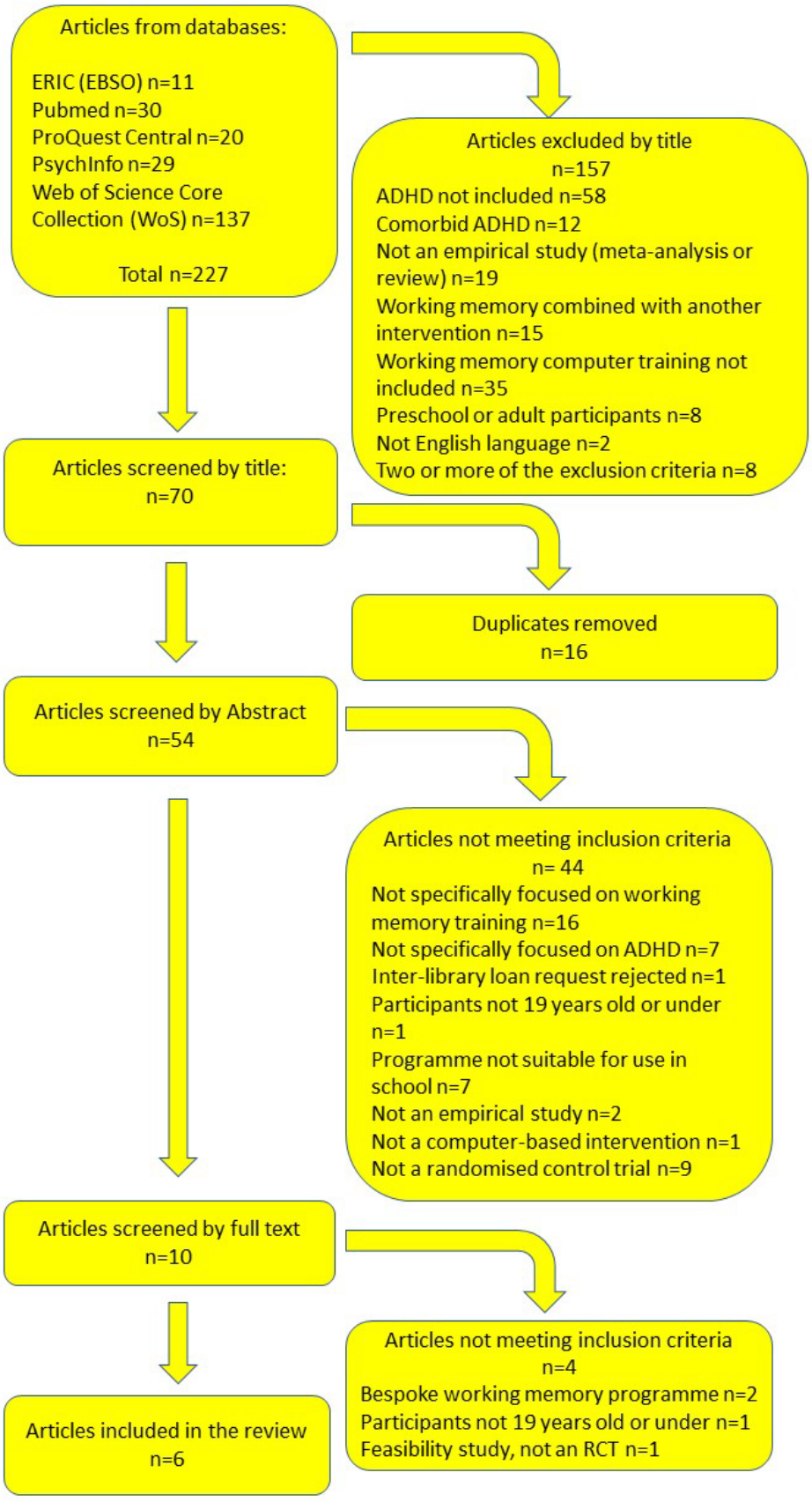

Table 2

To show Inclusion and Exclusion criteria for review articles

		Inclusion	Exclusion	Rationale
1	**Intervention**			
	Working memory (WM)	Specific focus on the effect of WM training only on ADHD.	Combining the WM intervention with other interventions.	To avoid confounding variables resulting from combining CWMT with other interventions such as exercise or maths training.
	Computer-based	Using a commercial software program.	No therapeutic or other WM interventions.	To reduce heterogeneity. Commercial software programmes can be reviewed and the study replicated to establish their effect on children with ADHD.
	Practical application	Can be used in the school, home or community.	Not practical for school, home or community use.	Computer-based interventions such as neuro feedforward and backward require specialist knowledge and equipment rendering them more difficult to implement (Dobrakowski & Lebecka, 2020; Steiner et al., 2014).
2	**Participants**			
	Children with ADHD	Only children with ADHD.	Not comorbid with other conditions such as ODD.	To minimise outcomes due to comorbid conditions and difficulties identifying which outcomes were due to ADHD.
	Age	Up to and including 19 years old.	No participants 20 years old or over.	EPs mainly work with children and young people (CYP) in schools and colleges. Most ADHD diagnoses are for

	Inclusion	Exclusion	Rationale
			children between 3-7 years old (NHS, 2021)
3	**Design and methodology**		
	Empirical studies	Empirical studies only	Not meta-analyses, reviews, reports or dissertations.	Primary, empirical sources used so the data can be interpreted from first principles.
	Quantitative	Pre- and post-intervention data to judge effectiveness of the intervention.	Qualitative studies or research without pre- and post-intervention data.	So that data can be re-analysed, standardised and effect sizes calculated.
	Randomised Controlled Trial (RCT)	Must be an RCT.	Not a RCT.	RCTs rank higher than other empirical studies, such as cohort, case-control and observational studies (Guyatt et al., 1995; 2000) and are appropriate for finding out if an intervention works (Petticrew & Roberts, 2003)
4	**Publication**		
	Peer Reviewed	Peer reviewed published articles only.	Work that is not peer reviewed nor published.	To mitigate against low quality studies and promote the inclusion of high-quality studies.
	Language	English only.	Not English.	No access to translators and translation software may be inaccurate.

As a result of the screening process, six articles were included in the review (Table 3).

Table 3

The Six Articles Included in the Review After Full-Text Screening

Study Code	Reference
1	Bigorra, A., Garolera, M., Guijarro, S., & Hervas, A. (2016). Long-term far-transfer effects of working memory training in children with ADHD: A randomized controlled trial. *European Child and Adolescent Psychiatry, 25,* 853-867. https://doi.org/10.1007/s00787-015-0804-3
2	Chacko, A., Bedard, A. C., Marks, D. J., Feirsen, N., Uderman, J. Z., Chimiklis, A., Rajwan, E., Cornwell, M., Anderson, L., Zwilling, A., & Ramon, M. (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, 55*(3), 247-255. https://doi.org/10.1111/jcpp.12146
3	Green, C. T., Long, D. L., Green, D., Losif, A-M, Dixon, J. F., Miller, M. R., Fassbender, C. & Schweitzer, J. B. (2012). Will working memory training generalize to improve off-task behaviour in children with Attention-Deficit/Hyperactivity Disorder? *Neurotherapeutics, 9,* 639-648. https://doi.org/10.1007/s13311-012-0124-y
4	Hovic, K. T., Saunes, B-K., Aarlien, A. K. & Egeland, J. (2013). RCT of working memory training in ADHD: Long-term near-transfer effects. *PLOS One, 8*(12), 1-9. https://doi.org/10.1007/s13311-012-0124-y
5	Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M, Gustafsson, P., Dahlstrom, K., Gillberg, C. G., Forssberg, H. & Westerberg, H. (2005). Computerized training of working memory in children with ADHD – a randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry, 44*(2). https://doi.org/10.1097/00004583-200502000-00010
6	Van Dongen-Boosma, 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, 55*(8), 886-896. https://doi.org/10.1111/jcpp.12218

3.2. Mapping the Field

3.2.1. Country of Origin

Research represented in this review was conducted in European countries and the United States of America (USA). Incidentally, during the screening process a significant number of articles were encountered from Nordic countries. This may be because there is a reluctance to medicate children with ADHD in these countries, particularly in Sweden (Klingberg et al., 2005) and so there may be more of an impetus to find alternatives (Table 4).

3.2.2. Participants

A total of 360 participants were included in this review, between five and fourteen years old, of which 68% were boys and 32% were girls. The participants were diagnosed with ADHD using a range of criteria depending on which country they were in and the sampling strategy used. Most children were diagnosed using DCM-IV-TR (American Psychiatric Association; APA, 2000), which was the established version at the time, and has now been superseded by DCM-5 (APA, 2013). However, there was variation regarding inclusion of the ADHD sub-types. For example, Klingberg et al. (2005) and Green et al. (2012) included children with the combined or inattentive sub-types, Bigorra et al. (2016) included children with the combined sub-type, and other studies did not state the ADHD sub-type included. Hovic et al. (2013) used F90.0 Hyperkinetic Disorder (ICD-10) (WHO, 1992) to diagnose ADHD, which was equivalent to DCM-IV. However, Chacko et al. 2014 used a combination of parent and teacher rating scales to diagnose ADHD (see Table 4), and did not employ a clinical diagnosis.

The medication status of participants was also heterogenous with three studies excluding medicated children (Bigorra et al., 2016; Klingberg et al., 2005; van Dongen-Boomsma et al., 2014), two studies including medicated children (Green et al., 2012; Hovik et al., 2013), and one study not stating the medication status of the children (Chako et al., 2014). Klingberg et al. (2005) purposely did not recruit medicated children because they predicted that the children's ADHD symptoms would be more acute than those of unmedicated children, and their progress less evident as a result of intervention.

3.2.3. The Cogmed Intervention

All studies used the Cogmed CWMT program, with one study using the junior version (van Dongen-Boomsma et al., 2014) and one employing Intent to Treat (ITT) instead of non-adaptive Cogmed (Hovik et al., 2013).

Table 4

Mapping the field to present an overview of the studies included in the review

Study Number/ Reference	Country of Origin	Age Group (Years)	Number of Participants	Gender %		ADHD Diagnosis	Medication Accepted?	Computer Programme Used	Control Group Activity
				Boys	Girls				
1. Bigorra et al. (2016)	Spain	7-12	66	45	55	*DCM-IV combined Type ADHD	No medication	RoboMemo Cogmed	Non-adaptive RoboMemo
2. Chacko et al. (2014)	U.S.A.	7-11	85	78	22	**Parent and Teacher Ratings	Not stated	Cogmed	Non-adaptive Cogmed
3. Green et al. (2012)	U.S.A.	7-14	26	63	37	*DCM-IV combined and inattentive subtype	20 medicated	RoboMemo Cogmed	Non-adaptive RoboMemo
4. Hovik et al. (2013)	Norway	10-12	75	73	27	***F90.0 Hyperkinetic Disorder (ICD-10) Combined Type	46 medicated	Cogmed	No intervention – intention to treat (ITT)
5. Klingberg et al. (2005)	Sweden	7-12	53	83	17	*DCM-IV, ADHD combined or predominantly inattentive subtype	No medication	RoboMemo Cogmed	Non-adaptive RoboMemo
6. Van Dongen-Boomsma et al. (2014)	Netherlands	5-7	51	67	33	*DCM-IV-TR	No medication	Cogmed JM	Non-adaptive Cogmed JM

*Participants were diagnosed using DCM-IV (APA, 2000) which was later superseded by DCM-5 (APA, 2013).

**Parent and teacher ratings on the Disruptive Behaviour Disorder Rating Scales (DBD; Pelham et al., 1992) and the Impairment Rating Scale (Fabiano et al., 2006) were used to diagnose children's ADHD, as well as parent interviews using the Kiddie-SADS (Kaufman et al., 1996).

*** F90.0 Hyperkinetic Disorder (ICD-10) (WHO, 1992) is equivalent to DCM-IV (APA, 2000).

3.3. Weight of Evidence (WoE)

Harden and Gough's (2012) Weight of Evidence (WoE) framework was used to assess the strengths and limitations of the studies included in this review. It consisted of four weights. WoE A, and the scoring criteria (Appendix D), assessed the methodological quality of the studies and how they compared to studies with similar methodologies. WoE A was evaluated using the Kratochwill (2003) protocol which is particularly suited to RCTs (Appendix H). Adaptations to the protocol to complement the studies reviewed were summarised in Appendix B. WoE B, and the scoring criteria (Appendix E, Table F), assessed how relevant the methodology was to answering the review question. WoE C, and the scoring criteria (Appendix F, Table H) assessed how well focused the study was on answering the research question. WoE D comprised an overall WoE score based on the average scores of WoE A, B and C. Tables summarising the WoE A to D scores for each of the studies reviewed were located in Appendices C, E, F and G respectively.

Scores for individual studies were assigned as follows: 0=No/Limited Evidence, 1=Weak Evidence, 2=Promising Evidence and 3=Strong Evidence. Where several scores were averaged, the composite average score was assigned to a band: 0.00-1.09=No/Limited Evidence, 1.10-1.69=Weak Evidence, 1.70-2.29=Promising Evidence and 2.30-3.00=Strong Evidence.

3.3.1. Methodological Quality (WoE A)

The methodological quality of the studies was varied. One study (Green et al., 2012) was Weak, three were Promising (Chacko et al., 2014; Klingberg et al., 2005; van Dongen-Boomsma et al., 2014) and two were strong (Bigorra et al., 2016; Hovic et al., 2013).

The strongest areas in this section were General Design Characteristics, Research Methodology and Comparison Group. The weakest areas were Data Analysis, Follow-up Assessment and Site of Intervention. The former was explored further under WoE B (see section 3.3.2.1.) and Follow-Up Assessment considered more closely under WoE C (see section 3.3.3.1.). The CWMT intervention was conducted at home for five of the studies and at school for the sixth (Hovic et al., 2013), as described below.

3.3.1.1. General Design Characteristics, Research Methodology and Comparison Group

The constraints implicit in an RCT meant that the General Design Characteristics in terms of randomisation was Strong for all studies. Similarly, the computerised nature and manufacturer-prescribed protocols inherent in the Cogmed intervention rendered the Research Methodology and Comparison Group Strong overall (Appendix C, Table C), with Chacko et al. (2014) the only study which cited ITT instead of an intervention element placebo (Table 4).

3.3.1.2. Site of Intervention

A selling point of Cogmed was that the intervention could be implemented at the child's home supervised by a parent, hence reducing the need for on-site

professional supervision. However, this methodology reduced control over implementation in terms of the time spent on tasks, engagement with tasks, and the environment the tasks were conducted in, for example, in a quiet room or with siblings playing close by. Only one study implemented the program in school (Hovic et al., 2013), and was judged as Strong for this reason due to professional supervision.

3.3.2. Methodological Relevance (WoE B)

The methodological relevance of the studies was varied (Appendix E, Table E). One study (Klingberg et al., 2005) indicated Limited relevance, two studies indicated Weak relevance (Chacko et al., 2014; Green et al., 2012), one study indicated Promising relevance (Hovic et al., 2013), and one study indicated Strong relevance (Bigorra et al., 2016).

The areas further analysed in this section for methodological relevance were: the sample size and power calculations, which indicated Limited evidence of methodological relevance; strategies employed to recruit participants, which indicated Promising evidence; and support provided to participants during the intervention, which indicated Weak evidence.

3.3.2.1. Sample Size and Power Calculation

There were inconsistencies between studies, with only one Strong study where sampling procedures and participant characteristics, and a description of the intervention context, were clear (Bigorra et al., 2016). Sampling procedures and their description were poor in the weakest study, compounded by only 26 children being sampled, and the "treatment group included a significantly greater number of participants with actively prescribed

medication for ADHD (*n*=8) than the comparison group (*n*=2)" (Green et al., 2012, p643), rendering the groups unbalanced.

Only one study (Bigorra et al., 2016) included a power calculation, and three studies offered insufficient or no rationale for the sample size used (Green et al., 2012; Hovik et al., 2013; Klingberg et al., 2005). Furthermore, Chacko et al. (2014, p251) input missing data "with the assumption that the missing data are at least missing at random", which may have altered the outcomes, considering the impact of attrition of nine out of 85 children. This comprised 11 percent of the sample, and suggested that the study was under-powered.

3.3.2.2. Participant Recruitment

How participants with ADHD were sourced demonstrated considerable variation. This ranged from community advertisements (Chacko et al., 2014) as the weakest strategy, to promising strategies involving recruitment through psychological or educational organisations (Green et al., 2012; Klingberg et al., 2005; van Dongen-Boomsma et al., 2014). The strongest recruitment strategies involved recruiting clinically diagnosed children with ADHD from hospitals or therapeutic settings (Bigorra et al., 2016; Hovic et al., 2013).

3.3.2.3. Participant Support During the Intervention

Chacko et al. (2014) replicated previous studies (Green et al., 2012; Klingberg et al., 2005) to address methodological flaws concerning treatment groups (adaptive Cogmed) and intervention element placebo groups (non-adaptive Cogmed). The lack of challenge in the Cogmed non-adaptive placebo intervention was argued to have reduced the motivation and training time of control group participants, and the increased engagement between

parents and children in the treatment (adaptive) group was proposed to inflate parent ratings on questionnaires (Chacko et al., 2014).

Consequently, Chacko et al. (2014) focused on ensuring that the conditions of the non-adaptive group, such as equal time spent on training tasks, mirrored those of the adaptive group, and participants were provided with extra coaching support and intervention to modify training tasks according to their needs. However, in terms of Methodological Relevance (WoE B; Appendix E, Table E), Chacko et al. (2014) and Green et al. (2012) were both judged Promising, whilst Klingberg et al. (2005) was judged Weak, indicating that the anticipated improvements in Chacko et al.'s (2014) methodology did not manifest in terms of their relevance to the review question.

3.3.3. Study Relevance to the Review Question (WoE C)

The effectiveness of the studies to ascertain if CWMT improves behavioural outcomes for children with ADHD was varied (Appendix F, Table G). Bigorra et al. (2016) was the only Strong study in this area, and notably, the only study demonstrating an overall significant impact of CWMT on behavioural ADHD outcomes (Table 5). Two studies were judged Promising (Chacko et al., 2014; van Dongen-Boomsma et al., 2014) and two were judged as Weak; Green et al. (2012) and Klingberg et al. (2005). Hovic et al. (2013) measured WM but not far transfer effects such as behavioural outcomes, and so was judged as providing No/Limited evidence of study relevance to the review question.

The areas of focus for WoE C were to ascertain if CWMT improved the behavioural symptoms of children with ADHD, if the primary study outcomes effectively measured behavioural ADHD symptoms, and the effectiveness of ADHD diagnosis.

3.3.3.1. CWMT and Behavioural Symptoms of ADHD

Overall, there was Weak evidence to support claims that CWMT improved behavioural symptoms of children with ADHD. This was because only half of the studies included follow-up data to assess the longer-term impact of the intervention after six months (Bigorra et al., 2016; Klingberg et al., 2005) and eight months (Hovic et al., 2013), and sometimes primary outcomes did not align with each other. For example, parent ratings indicated significant effect, whereas teacher ratings were not significant (Klingberg et al., 2005), and children's off-task behaviour improved, but other behaviours such as fidgeting did not improve (Green et al., 2012). Where outcomes did align, with the exception of Bigorra et al. (2016), there was no overall significant effect of CWMT on behavioural symptoms of ADHD (Chacko et al., 2014; van Dongen-Boomsma et al., 2014).

3.3.3.2. Primary Outcomes Measures

Overall, primary outcome measures were Promising, although scores varied between studies. Strong studies provided reliability data for questionnaires and at least two outcome measures (Bigorra et al., 2016; Chacko et al., 2014). However, despite providing inter-rater reliability data for the observational methodology employed, Green et al. (2012) only used one outcome measure, hence why it was judged Weak.

3.3.3.3. Confirmation of ADHD Status

Overall, confirmation of ADHD status was Promising, with some variation between studies (Table 4). The Strong studies recruited participants from establishments where they were already diagnosed with ADHD (Bigorra et al., 2016; Hovic et al., 2013), so more was known about the children's condition. In the only Weak study, participants were recruited from the community and non-clinical "consensus diagnosis based on parent and teacher ratings" used to assign ADHD status (Chacko et al., 2014, p249).

3.3.4. Overall Review Judgment (WoE D)

The review was judged Promising overall (Appendix G), however, no areas were judged Strong. Methodological Quality (WoE A) was Promising, highlighting that reviewing RCTs does not guarantee encountering high quality studies, although the structure embedded in RCTs and the Cogmed intervention may have inflated this score, masking genuine methodological weaknesses. WoE B was judged Weak, indicating that either the lack of quality RCTs resulted in outcomes that did not adequately answer the review question, or that RCTs were not the appropriate methodology to answer the review question. WoE C was at the lower end of the Promising category, indicating that the review question was only partially answered.

3.3.5. Summary of Review Findings

There was one study, and the only Strong study according to WoE D, which conclusively demonstrated that CWMT improved the behavioural symptoms of children with ADHD (Table 5; Bigorra et al., 2016). This referred to elements of real-world EF according to the Behaviour Rating Inventory of EF

(BRIEF; Gioia et al., 1996), with the caveat that only the significant results were reported and so an overall effect size could not be calculated. The two Weak studies according to WoE D generated mixed results with large (off-task behaviour) and small (fidgets) effect sizes rendering the overall result inconclusive (Green et al., 2012), and medium (hyperactivity/impulsivity) and small (inattention) effect sizes which were considered with caution due to the poor methodology employed (Klingberg et al., 2005). Moreover, the effect sizes calculated from the raw data published by Klingberg et al. (2005), using the Campbell Calculator (Wilson, 2022), were lower than those published in the article, which cited "corrected values" without a clear explanation of their rationale for correction (Klingberg et al., 2005, p181). Furthermore, it was noted that Klingberg et al. (2005), used one-tailed tests during statistical analyses, which reduced the statistical power required for significant outcomes, and justified this by stating that the study was based on previous research findings (Klingberg et al., 2002). Thus, the two Weak studies, plus a third study which did not measure behavioural outcomes (Hovic et al., 2013), contributed little to the overall review conclusions.

The remaining two studies were both judged Promising (WoE D) and both reported little or no effect of CWMT on behavioural symptoms of ADHD (Chacko et al., 2014; van Dongen-Boomsma et al., 2014). Interestingly, the latter used the BRIEF questionnaire (Gioia et al., 1996), as did Bigorra et al. (2016), but reported an effect size for the whole questionnaire rather than just significant elements. This may explain why one result supported CWMT (Bigorra et al., 2016) and the other did not (van Dongen-Boomsma et al., 2014) despite the same questionnaire being used. Alternatively, the contrast

may be due to the participants' age (5-7 years old verses 7-12 years old), sex (more boys than girls verses more girls than boys), or the computer program used (Cogmed Junior verses RoboMemo Cogmed), and highlighted challenges with the heterogeneity of the studies (Table 4).

Table 5

Summary of the review findings

Study	Sample Size	Primary Outcome Measure(s)	Post-Intervention and Follow-Up Effect Sizes	Effect Size Descriptor	Does CWMT Improve behavioural ADHD Symptom's in Children?	WoE D
1. Bigorra et al. (2016)	66 (Active n=36) (Control n=30)	[2]BRIEF-Parent T2 = 6-month follow up	Working Memory Subscale T2-T1 *d*= -0.86 T2-T0 *d*= -0.61 Plan/Organise T2-T1 *d*= -0.71 Metacognition Index T2-T1 *d*= -0.78	 Large Medium Medium Medium	Small to Large Effect though most effects were Medium	2.62 Strong
		[2]BRIEF-Teacher T2 = 6-month follow up	Initiate Subscale T1-T0 *d*= -0.55 T2-T0 *d*= -0.57 Working Memory T1-T0 *d*= -0.36 T2-T0 *d*= -0.84 Metacognitive Index T1-T0 *d*= -0.37 T2-T0 *d*= -0.81 Monitor Subscale T2-T1 *d*= -0.72 T2-T0 *d*= -0.79	 Medium Medium Small Large Small Large Medium Medium		

Study	Sample Size	Primary Outcome Measure(s)	Post-Intervention and Follow-Up Effect Sizes	Effect Size Descriptor	Does CWMT Improve behavioural ADHD Symptom's in Children?	WoE D
			Shift Subscale T2-T1 *d*= -0.39	 Small		
2. Chacko et al. (2014)	85 (Active n=44) (Control n=41)	Parent [3]DBD-IN Parent [4]DBD-HI Teacher [3]DBD-IN Teacher [4]DBD-HI	*d*= -0.24 *d*= -0.24 *d*= -0.03 *d*= 0.07	Small Small Negligible Negligible	Small to Little or No Effect	1.78 Promising
3. Green et al. (2012)	26 (Active n=12 (Control n=14)	[5]Restricted Academic Situations Task	Off-task behaviour [1]*d*= -1.10 Fidgets [1]*d*= -0.23	 Large Small	Inconclusive	1.52 Weak
4. Hovic et al. (2013)	75 (Active n=38 (Control n=37)	Visual WM Auditory WM Manipulation WM	*d*= 1.11 *d*= 0.47 *d*= 0.73	Large Medium Large	Did not measure behavioural outcomes (only WM outcomes)	1.79 Promising
5. Klingberg et al. (2005)	53 (Active n=26) (Control	T1 ADHD Parent [6]Inatt. T1 ADHD Parent [7]H/I T1 ADHD Teacher [6]Inatt. T1 ADHD Teacher [7]H/I.	[1]*d*= -0.36 [1]*d*= -0.37 [1]*d*= 0.29 [1]*d*= 0.49	Small Small Small Medium	Medium Effect for hyperactivity/ impulsivity.	1.59 Weak

Study	Sample Size	Primary Outcome Measure(s)	Post-Intervention and Follow-Up Effect Sizes	Effect Size Descriptor	Does CWMT Improve behavioural ADHD Symptom's in Children?	WoE D
	n=27)					
		6-month follow-up			Small Effect	
		T2 ADHD Parent [6]Inatt.	[1]*d*= -0.13	Negligible	for inattention.	
		T2 ADHD Parent [7]H/I	[1]*d*= -0.56	Medium		
		T2 ADHD Teacher [6]Inatt.	[1]*d*= 0.41	Small		
		T2 ADHD Teacher [7]H/I.	[1]*d*= 0.73	Medium		
6. Van Dongen-Boomsma et al. (2014)	51 (Active n=26) (Control n=21)	Investigator [8]ADHD Rating Scale IV	*d*= 0.28	Small	Small to Little to No Effect	1.89 Promising
		Teacher [8]ADHD Rating Scale IV	*d*= -0.01	Negligible		
		[2]BRIEF-Parent	*d*= -0.02	Negligible		
		[2]BRIEF-Teacher	*d*= -0.21	Small		

[1] Cohen's d effect size calculated using Campbell's Calculator (Wilson, 2022). Effect size descriptors: Small=0.2, Medium=0.5, Large=0.8 (Statistics How To https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/cohens-d/
[2]Behaviour Rating Inventory of Executive Function (BRIEF; Gioia et al., 1996).
[3]Disruptive Behaviour Disorders Rating Scale (DBD; Pelham et al., 1992). IN = Inattention.
[4]Disruptive Behaviour Disorders Rating Scale (DBD; Pelham et al., 1992). HI = Hyperactivity/Impulsivity.
[5]Restricted Academic Situations Task (no reference provided).
[6]ADHD DSM-IV Rating Scale (APA, 1994). Inatt = Inattention.
[7]ADHD DSM-IV Rating Scale (APA, 1994). H/I = Hyperactivity/Impulsivity.
[8]ADHD Rating Scale IV (Zhang et al., 2005).

4.0. Conclusion and Recommendations

The findings of the strongest study in the sample indicated that CWMT significantly improves elements of EF (Bigorra et al., 2016). However, there were little or no far-effects on wider behavioural symptoms of ADHD in children. This may be because EF is only one of several endophenotypes of ADHD (Sonuga-Barke et al., 2014), EF deficits are not consistently represented across ADHD populations (Barkley & Murphy, 2010), and because ADHD is a heterogenous condition with many manifestations (Wahlstedt et al., 2009). Therefore, a multi-faceted, multi-modal approach may be more effective at reducing the behavioural symptoms of ADHD in children (Murray et al., 2008).

Including CMWT in a multi-modal approach raises the ethical issue of cost-verses-benefit (Guyatt et al., 2000), which would not work in CWMT's favour, as the costs are high. Furthermore, one of the selling points of CWMT is its usability in the home, reducing the need for on-site professional input. Chacko et al. (2014) raised legitimate concerns regarding the efficacy of home intervention implementation, which could be addressed via more intensive coaching support, but more effectively by CWMT being conducted in a school or clinical environment. However, that again would increase the cost of the intervention in terms of both time and professional personnel, rendering it very resource intensive.

These issues of limited effectiveness on a range of ADHD manifestations, cost implications and availability of professional personnel are what EPs would need to consider before recommending CWMT to clients.

4.1. Limitations to the Research

There were a number of limitations to this review which generally emanated from the overall poor quality of the studies, as evidenced by only one Strong study out of six, and two Weak studies overall (WoE D). Therefore, one should not assume that an RCT guarantees high quality research.

Limitations included under-powered studies, heterogeneity in the ADHD samples due to different recruitment strategies sourcing children with diverse ADHD status, home intervention limiting the amount of control the researcher had over intervention implementation, questionable alignment between primary outcome measures and ADHD behavioural symptoms, and there was no pupil voice or feedback on the processes involved from participants. This may have yielded insights into how the intervention was perceived by those involved and how the methodology could have been improved.

4.2. Recommendations

In order to gain a better understanding of how WM could improve behavioural symptoms of children with ADHD, a priority for further research would be to either improve RCT methodologies and address the limitations cited in this review, or adopt a different approach involving qualitative research and mixed methodologies. This would ask more process rather than an effectiveness questions (Petticrew & Roberts, 2003), in order to gain a deeper understanding of the phenomena.

As EF deficits are not universal amongst children with ADHD, it may also be beneficial to target more universal or multiple ADHD endophenotypes, consistent with the heterogenous nature of the ADHD condition. Possible

suggestions for consideration are motivational or energetic endophenotypes (Sonuga-Barke, 2014). These interventions may prove more successful, especially if they are used synergistically to reduce the behavioural symptoms of children with ADHD.

5.0. References

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Medical Disorders* (4th ed.). APA.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Medical Disorders* (4th ed., text rev.). APA.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Medical Disorders* (5th ed.). APA.

Baddeley, A. D., Bressi, S., Sala, D., Logie, R., & Spinnler, H. (1991). The decline of working memory in Alzheimer's disease. *Brain, 114*, 2521-2542. https://doi.org/10.1093/BRAIN/114.6.2521

Barkley, R. A., & Murphy, K. R. (2010). Impairment in occupational functioning and adult ADHD: The predictive utility of executive function (EF) ratings versus EF tests. *Archives of Clinical Neuropsychology*, *25*, 157-173. https://doi.org/10.1093/arclin/acq014

Biederman, J., Monuteaux, M. C., Mick, E., Spencer, T., Wilens, T. E., Silva, J. M., Snyder, L. E., & Faraone, S. V. (2006). Young adult outcome of attention deficit hyperactivity disorder: A controlled 10-year follow-up study. *Psychological Medicine*, *36*, 167-179. https://doi.org/10.1017/S0033291705006410

Bigorra, A., Garolera, M., Guijarro, S., & Hervas, A. (2016). Long-term far-transfer effects of working memory training in children with ADHD: A randomised controlled trial. *European Child and Adolescent Psychiatry, 25*, 853-867. https://doi.org/10.1007/s00787-015-0804-3

BPS (2018). *Code of Ethics and Conduct*. BPS. https://www.bps.org.uk/guideline/code-ethics-and-conduct

CAMHS Advisory Group (2018). *Delivering Effective Services for Children and Young People with ADHD*. CAMHS Advisory Group. https://www.england.nhs.uk/north-west/wp-content/uploads/sites/48/2019/03/GM-wide-ADHD-guidance.pdf

Castellanos, F. X., & Tannock, R. (2002). Neuroscience of attention-deficit/hyperactivity disorder: The search for endophenotypes. *Neuroscience, 3*, 617-628. https://doi.org/10.1038/NRN896

Cortese, S., Ferrin, M., Brandeis, D., Buitelaar, J., Daley, D., Dittman, R. W., Holtmann, M., Santosh, P., Stevenson, J., Stringaris, A., Zuddas, A., & Sonuga-Barke, E. J. S. (2015). Cognitive training for attention-deficit/hyperactivity disorder: Meta-analysis of clinical and neuropsychological outcomes from randomized controlled trials. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*(3), 164-174. https://doi.org/10.1016/J.JAAC.2014.12.010

Defeyter, M. A., Russo, R., & McPartlin, P. L. (2009). The picture superiority effect in recognition memory: A developmental study using the response signal procedure. *Cognitive Development, 24*, 265-273. https://doi.org/10.1016/J.COGDEV.2009.05.002

Dobrakowski, P., & Lebecka, G. (2020). Individualised neurofeedback training may help achieve long-term improvement of working memory in children with ADHD. *Clinical EEG and Neuroscience, 51*(2), 94-101. https://doi.org/10.1177/1550059419879020

Fabiano, G. A., Pelham, W. E., Gnagy, E. M., Kipp, H., Lahey, B. B.,Burrows-MacLean, L., & Chronis, A. M. (2006). The reliability and validity of the Children's Impairment Rating Scale: A practical measure of impairment

in children with ADHD. *Journal of Clinical Child and Adolescent Psychology, 35*, 369-385.

Faraone, S. V., & Buitelaar, J. (2010). Comparing the efficacy of stimulants for ADHD in children and adolescents using meta-analysis. *European Child and Adolescent Psychiatry, 19*, 353-364. https://doi.org/10.1007/s00787-009-0054-3

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A-M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology, 13*, 65-77. https://doi.org/10.1002/(SICI)1099-0720(199902)13:1<65::AID-ACP548>3.0.CO;2-O

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (1996). *Behaviour Rating Inventory of Executive Function: Professional Manual.* Psychological Assessment Resources Inc.

Green, C. T., Long, D. L., Green, D., Losif, A-M., Dixon, J. F., Miller, M. R., Fassbender, C., & Schweitzer, J. B. (2012). Will working memory training generalize to improve off-task behaviour in children with Attention-Deficit/Hyperactivity Disorder? *Neurotherapeutics*, *9*(3), 639–648. https://doi.org/10.1007/s13311-012-0124-y

Groenman, A. P., Oosterlaan, J., Rommelse, N. N. J., Franke, B., Greven, C. U., Hoeksra, P. J., Hartman, C. A., Luman, M., Roeyers, H., Oades, R. D., Sergeant, J. A., Buitelaar, J. K., & Faraone, S. V. (2013). Stimulant treatment for attention-deficit hyperactivity disorder and risk of developing substance use disorder. *The British Journal of Psychiatry, 203*, 112-119. https://doi.org/10.1192/bjp.bp.112.124784

Guyatt, G. H., Haynes, R. B., Jaeschke, R. Z., Cook, D. J., Green, L., Naylor, C. D., Wilson, M. C., & Richardson, W. S. (2000). Users' guides to the medical literature XXV. Evidence-based medicine: Principles for applying the users' guides to patient care. *Journal of American Medical Association, 284*(10), 1290-1296. https://doi.org/10.1001/jama.284.10.1290

Guyatt, G. H., Sackett, D. L., Sinclair, J. C., Hayward, R., Cook, D. J., & Cook, R. J. (1995). Users' guides to the medical literature IX. A method for grading health care recommendations. *Journal of American Medical Association, 274*(22), 1800-1804.

Gwernan-Jones, R., Moore, D. A., Cooper, P., Russell, A. E., Richardson, M., Rogers, M., Thompson-Coon, J., Stein, K., Ford, T. J., & Garside, R. (2016). A systematic review and synthesis of qualitative research: the influence of school context on symptoms of attention deficit hyperactivity disorder. *Emotional and Behavioural Difficulties, 21*(1), 83-100. https://doi.org/10.1080/13632752.2015.1120055

Hague, C., Holland, A. A., & Gottleib, M. (2020). Selected ethical issues in computerized cognitive training. *Clinical Practice in Pediatric Psychology, 8*(4), 402-408. https://doi.org/10.1037/cpp0000311

Harden, A., & Gough, D. (2012). Quality and Relevance Appraisal. In D. Gough, S. Oliver, and J. Thomas (Eds.), *An Introduction to Systematic Reviews* (pp.153-178). Sage Publications Ltd.

HCPC. (2016). *Standards of Conduct, Performance and Ethics.* HCPC. https://www.hcpc-uk.org/globalassets/resources/standards/standards-of-conduct-performance-and-ethics.pdf

Hovik, K. T., Saunes, B-K., Aarlien, A. K., & Egeland, J. (2013). RCT of working memory training in ADHD: Long-term near-transfer effects. *PLOS One, 8*(12), 1-9. https://doi.org/10.1371/journal.pone.0080561

Jackson, E., Leitao, S., & Claessen, M. (2016). The relationship between phonological short-term memory, receptive vocabulary, and fast mapping in children with specific language impairment. *International Journal of Language Communication Disorder, 51*(1), 61-73. https://doi.org/10.1111/1460-6984.12185

Kasper, L. J., Alderson, R. M., & Hudec, K. L. (2012). Moderators of working memory deficits in children with attention-deficit/hyperactivity disorder (ADHD): A meta-analytic review. *Clinical Psychological Review, 32,* 605-617. https://doi.org/10.1016/j.cpr.2012.07.001

Kaufman, J., Birmaher, B., Brent, D., Rao, U., & Ryan, N. (1996). *Kiddie-SADS present and lifetime version (K-SADS-PL)*. University of Pittsburgh.

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

Klingberg, T., Fernell, E., Pernille, J. O., Johnson, M., Gustafsson, P., Dahlstrom, K., Gillberg, C. G., Forssberg, H., & Westerberg, H. (2005). Computerized training of working memory in children with ADHD – A randomized, controlled trial. *Journal of American Academy of Child and Adolescent Psychiatry, 44*(2), 177-186. https://doi.org/10.1097/00004583-200502000-00010

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Psychology, 24*(6), 781-791. http://dx.doi.org/10.1076/jcen.24.6.781.8395

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

Lieberman, D. A. (2012). *Human Learning & Memory*. Cambridge University Press.

Moore, D. A., Russell, A. E., Arnell, S., & Ford, T. J. (2017). Educator's experiences of managing students with ADHD: A qualitative study. *Child: Care, Health and Development, 43*(4), 489-498. https://doi.org/10.1111/cch.12448

Murray, D. W., Arnold, E., Swanson, J., Wells, K., Burns, K., Jensen, P., Hechtman, L., Paykina, N., Legato, L., & Strauss. (2008). A clinical review of outcomes of the multimodal treatment study of children with attention-deficit/hyperactivity disorder (MTA). *Current Psychiatry Reports, 10*(5), 424-431. https://doi.org/10.1007/s11920-008-0068-4

NHS. (2021, December 24). *Overview: Attention deficit hyperactivity disorder*. NHS. https://www.nhs.uk/conditions/attention-deficit-hyperactivity-disorder-adhd/

Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience, 7*(1), 75-79. https://doi.org/10.1038/nn1165

Petticrew, M., & Roberts, H. (2009). Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiology and Community, 57*(7), 527-529. http://dx.doi.org/10.1136/jech.57.7.527

Roche, J. D., & Johnson, B. D. (2014). Cogmed working memory training product review. *Journal of Attention Disorders, 18*(4), 379-384. http://dx.doi.org/10.1177/1087054714524275

Silk, T. J., Vance, A., Rinehart, N., Bradshaw, J. L, & Cunnington, R. (2008). Dysfunction in the fronto-parietal network in attention deficit hyperactivity disorder (ADHD): An fMRI study. *Brain Imaging and Behavior, 2*, 123-131. https://doi.org/10.1007/s11682-008-9021-8

Simons, D. J., Boot, W. R., Charness, N., Gathercole S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do "Brain-Training" programs work? *Psychological Science in the Public Interest, 17*(3), 103-186. https://doi.org/10.1177/1529100616661983

Sonuga-Barke, E. J. S., Brandeis, D., Cortese, S., Daley, D., Ferrin, M., Holtmann, M., Stevenson, J., Danckaerts, M., van der Oord, S., Dopfner, M., Dittmann, R. W., Simonoff, E., Zuddas, A., Banaschewski, T., Buitelaar, J., Coghill, D., Hollis, C., Konofal, E., Lecendreux, M., …Sergeant, J. (2013). Nonpharmacological interventions for ADHD: Systematic review and meta-analyses of randomized controlled trials of dietary and psychological treatments. *American Journal of Psychiatry, 170*(3), 275-289. https://doi.org/10.1176/appi.ajp.2012.12070991

Sonuga-Barke, E., Brandeis, D., Holtman, M., & Cortese, S. (2014). Computer-based cognitive training for ADHD: A review of current

evidence. *Child Adolescent Psychiatric Clinics of North America, 23*, 807-824. http://dx.doi.org/10.1016/j.chc.2014.05.009

Steiner, N. J., Frenette, E. C., Rene, K. M., Brennan, R. T., & Perrin, E. C. (2014). In-school neurofeedback training for ADHD: Sustained improvements from a randomized control trial. *Pediatrics, 133*(3), 483-492. http://dx.doi.org/10.1542/peds.2013-2059

Taylor, E., Dopfner, M., Sergeant, J., Asherson, P., Banaschewski, T., Buitelaar, J., Coghill, D., Danckaerts, M., Rothenberger, A., Sonuga-Barke, E., Steinhausen, H-C, & Zuddas, A. (2004). European clinical guidelines for hyperkinetic disorder – first upgrade. *European Child and Adolescent Psychiatry, 13*(1), 7-30. https://doi.org/10.1007/s00787-004-1002-x

Van de Loos-Neus, G. H. H., Rommelse, N., & Buitelaar, J. K. (2011). To stop or not to stop? How long should medication treatment of attention-deficit hyperactivity disorder be extended? *European Neuropsychopharmacology, 21*, 584-599. https://doi.org/10.1016/j.euroneuro.2011.03.008

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, 55*(8), 886-896. https://doi.org/10.1111/jcpp.12218

Wahlstedt, C., Thorell, L. B., & Bohlin, G. (2009). Heterogeneity in ADHD: Neuropsychological pathways, comorbidity and symptom domains.

Journal of Abnormal Child Psychology, 37, 551-564. http://dx.doi.org/10.1007/s10802-008-9286-9

Wilson, D. B. (2022, July 20). Practical Meta-Analysis Effect Size Calculator [Online calculator]. Campbell Collaboration. https://campbellcollaboration.org/research-resources/effect-size-calculator.html

Zhang, S., Faries, D. E., Vowles, M., & Michelson, D. (2005). ADHD rating scale IV: Psychometric properties from a multinational study as a clinician-administered instrument. *International Journal of Methods in Psychiatric Research, 14*, 186-201. http://dx.doi.org/10.1002/mpr.7

6.0. Appendices

Appendix A: Studies Excluded at Full Text Screening with Rationale

Table A

To Show Studies Excluded at the Full Text Screening Stage of the Literature Search

Referenced Article	Exclusion Criteria	Rationale
Bikic et al. (2017)	Feasibility study – not a fully-fledged RCT. Only include RCTs.	Feasibility studies are underpowered and explore if a full RCT is justified (Abbott, 2014).
Klingberg et al. (2002).	Not using commercial software.	Difficult to establish the reliability of the software. Published software can be reviewed and the study replicated.
Prins et al. (2011).	Not using commercial software.	Difficult to establish the reliability of the software. Published software can be reviewed and the study replicated.
Woltering et al. (2021).	Children and young people must be 19 years old or under.	The age group was 18-35 years old, but this was not stated in the abstract.

References

Abbott, J. H. (2014). The distinction between randomized clinical trials (RCTs) and preliminary feasibility and pilot studies: What are they and what are they not. *Journal of Orthopaedic & Sport Physical Therapy, 44*(8), 555-558. http://dx.doi.org/10.2519/jospt.2014.0110

Bikic, A., Christensen, T, O., Leckman, J. F., Bilenberg, N., & Dalsgaard, S. (2017). A double-blind randomized pilot trial comparing computerized cognitive exercises to Tetris in adolescents with attention-deficit/hyperactivity disorder. *Nordic Journal of Psychiatry, 71*(6), 455-464. https://doi.org/10.1080/08039488.2017.1328070

Prins, P. J. M., Dovis, S., Ponsioen, A. P., ten Brink, E., & van der Oord, S. (2011). Does computerised working memory training with game elements enhance motivation and training efficacy in children with ADHD? *Cyberpsychology, Behaviour and Social Networking, 14*(3), 115-122. https://doi.org/10.1089/cyber.2009.0206

Woltering, S., Gu, C., Liu, Z-X., & Tannock, R. (2021). Visuospatial working memory capacity in the brain after working memory training in college students with ADHD: A randomized controlled trial. *Journal of Attention Disorders, 25*(7), 1010-1020. http://dx.doi.org/10.1177/1087054719879487

Appendix B: WoE A Adaptations to the Kratochwill (2003) Coding Protocol

Weight of Evidence (WoE) ratings were based on the Harden & Gough (2012) framework which provides systematic and objective ways to evaluate studies. WoE A assesses the Methodological Quality of a study. This was achieved using an adapted version of the Procedural Manual of the Task Force on Evidence-Based Interventions in School Psychology, American Psychology Association (Kratochwill, 2003). The Kratochwill (2003) protocol was chosen to assess WoE A because it was a robust protocol suited to the design of RCTs and all of the studies reviewed were RCTs.

Table B

To Show how the Kratochwill (2003) Coding Protocol was Adapted for the Review

Section of Protocol Excluded	Rationale for Exclusion
Domain	This was reduced to 'School- and community-based intervention programmes for social and behavioural problems' and the other options deleted. This was because the intervention was for social and behavioural problems and the inclusion criteria stated that interventions should be practical enough for use at school and home. Thus, this box would be ticked for every study reviewed.
Part I General Study Characteristics	
A1 to A5	The study characteristics were described in the Mapping the Field table (Table 5), so this was considered duplication.
C3 Effect Size	Effect sizes for each study were calculated elsewhere (Table 6) and so this was adapted to state the researcher's justification for the sample size used.

Section of Protocol Excluded	Rationale for Exclusion
C7 and C8	These refer to qualitative studies and the studies reviewed were all quantitative.
D Type of Program	All the studies were intervention studies, so this seemed superfluous.
E and F stage of the program and historical status of the program	Not considered relevant as all the studies used a similar CWMT program as an intervention and historical use would be unknown.
Part II	
A1 and A2 Characteristics of the data collector and participants	The characteristics of the data collector and 'what culture means' to the participants were considered beyond the scope of this review.
B4 Extent of Engagement	How the researcher conducted data collection was considered irrelevant as primary data was collected via an internet-based computer programme.
B6 Cultural appropriateness of measures	Considered beyond the remit of the review.
D1-6 Primary/secondary outcomes statistically significant	Outcomes were represented in Table 6, including effect sizes, and explored under WoE C (Appendix F). Power was addressed in Part I B Data Analysis. Therefore, this was removed.
E Cultural significance	Considered beyond the scope of this review.
F Educational/clinical significance	The impact of intervention outcomes on pupils with ADHD were explored in WoE C and so this was considered duplication.
G1 External validity – sampling procedures	'Specify' for G1.2, G1.3 was not included as the information was not considered to assist discriminating between the studies. Sampling in terms of recruitment was also explored in WoE B. G1.7 was excluded as similarity to school practice was not relevant.
G2 External validity – participant characteristics	Ethnic identity, race, acculturation, functional descriptors were excluded as the included characteristics were considered sufficient.
G3, G4, G6 External validity	Considered beyond the remit of or not relevant to the review.

Section of Protocol Excluded	Rationale for Exclusion
H1.4, H2, H3, H4 Follow-up	Excluded as not considered to add to the findings. Information in H1.1 to H1.3 was considered sufficient to discriminate between studies.
I Identifiable intervention components	Excluded - not considered relevant as the intervention was a very specific computer-based working memory training programme and all studies used the same programme.
J1 Intervention Fidelity	Not considered it would add to differentiating between studies as J2 – manualisation – gave sufficient information.
J4.1 to J4.4 Implementation Context	These aspects were not considered relevant to the review and would be difficult to assess.
J4.7 Implementation Context Dosage response	Not considered relevant as dosage was not part of the study.
J4.9 Intervention style	Not considered to add to the review as the same computer programme was used for the intervention in all of the studies reviewed.
J4.10 Cost analysis data	Considered beyond the remit of this study.

Appendix C: WoE A Ratings for Methodological Quality

Weight of Evidence A Ratings

The WoE A judgements for each study reviewed were summarised in Table C.

The scores represented the following judgements:

- 3 = Strong Evidence
- 2 = Promising Evidence
- 1 = Weak Evidence
- 0 = No/Limited Evidence.

The scores for each indicator represented the average score of the sub-sections of each indicator, hence the decimalised scores.

The following key was used to describe the average scores calculated:

Key:

Range used to assign descriptors to mean scores:

0.00 – 1.09 = No/Limited Evidence

1.10 – 1.69 = Weak

1.70 – 2.29 = Promising

2.30 – 3.00 = Strong

Table C

To show a summary of WoE A scores for Methodological Quality

	Indicator	Overall Evidence Rating for Each Study Reviewed (0-3 Score)						Mean	Description of Evidence: Strong Promising Weak No/Limited Evidence
		Bigorra et al. (2016)	Chacko et al. (2014)	Green et al. (2012)	Hovic et al. (2013)	Klingberg et al. (2005)	Van Dongen-Boomsma et al. (2014)		
	Study Number	1	2	3	4	5	6		
Part I									
A	General Design Characteristics	3.0	3.0	3.0	3.0	3.0	3.0	3.00	Strong
B	Data Analysis	3.0	1.5	0.5	1.5	1.0	2.5	1.67	Weak
Part II									
A	Research Methodology	3.0	2.0	2.0	2.0	3.0	3.0	2.50	Strong
B	Measurement	2.0	3.0	1.0	2.0	1.0	2.0	1.83	Promising
C	Comparison Group	3.0	3.0	3.0	3.0	3.0	3.0	3.00	Strong
D	External Validity Indicators	3.0	2.0	1.5	2.0	2.0	2.5	2.17	Promising
E	Follow-up Assessment	3.0	0.0	0.0	3.0	3.0	0.0	1.50	Weak
F	Implementation Fidelity	1.7	2.7	2.0	2.0	2.0	1.7	1.95	Promising
G	Site of Intervention	1.0	1.0	1.0	3.0	1.0	1.0	1.33	Weak
Overall Indicator Mean		2.52	2.02	1.55	2.38	2.11	2.01	2.11	Promising
Overall Descriptor		Strong	Promising	Weak	Strong	Promising	Promising	Promising	

Appendix D: WoE A Scoring Criteria

The scores in Table C were assigned according to criteria for each indicator. The criteria are described in Table D.

Table D

Summarising the criteria used to assign scores to assess the studies for WoE

A.

		Criteria used to assign scores of …		
	Indicator	**Weak 1**	**Promising 2**	**Strong 3**
Part I				
A	General Design Characteristics	Randomisation stated but not explained. Moderate confidence in judgement.	Randomisation explicitly stated. High confidence in judgement.	Randomisation explicitly stated and explained. Very high confidence in judgement.
B	Data Analysis	One out of appropriate analysis, statistical test and familywise error controlled.	Two out of appropriate analysis, statistical test and familywise error controlled.	Appropriate analysis, statistical test and familywise error controlled.
Part II				
A	Research Methodology	All items in this section have individual scores which are averaged to calculate an overall score.		
B	Measurement	Average scores on B4, B6 and B7. At least 1 'Yes'	Average scores on B4, B6 and B7. At least 2 'Yes'	Average scores on B4, B6 and B7. At least 3 'Yes'
C	Type of Comparison Group	Typical Contact, No Intervention, Wait List. Moderate confidence in judgement	Minimal contact. High confidence in judgement	Placebo, Alternative Intervention. Very High confidence in judgement
	Change Agent	Post hoc test for group equivalence.	Post hoc matched set. Post or follow-up analysis with low attrition	Random assignment/ statistical matching. Post and follow-up analysis with low attrition

		Criteria used to assign scores of …		
	Indicator	**Weak 1**	**Promising 2**	**Strong 3**
D	External Validity Indicators			
	Sampling Procedures	D1.1 to D1.7 At least 4 'Yes'	D1.1 to D1.5 At least 3 'Yes' D1.6 or D1.7 'Yes'	D1.1 to D1.5 At least 4 'Yes' D1.6 and D1.7 'Yes'
	Participant Characteristics	At least 2/8 characteristics	At least 4/8 characteristics	At least 6/8 characteristics
E	Follow-Up Assessment	1 'Yes'	2 'Yes'	3 'Yes' and information specified
F	Implementation Fidelity			
	Manualisation and Adaptation	F1.4	F1.1 or F1.2 or F1.3 or F1.4 and Adaptation	F1.1 or F1.2 and Adaptation
	Length of Intervention	One section completed	Two sections completed	All three sections completed
	Implementer Training and Support	Parents, college students, peers. Little or no training. Ongoing technical support	Research staff, school speciality staff, teachers, education assistants. Simple orientation. Program materials	Certified coach. Training workshops. Special facilities
G	Site of Implementation	Home	School	Hospital/clinic/health centre

Appendix E: WoE B Ratings for Methodological Relevance

WoE B focused on:

1. Data Analysis: Is the sample size adequately powered to justify statistical analyses?
2. External Validity Indicators: Were the strategies employed appropriate to recruit participants with ADHD?
3. Implementation Fidelity: Was adequate support provided for implementers to deliver the intervention effectively?

Key:

Range used to assign descriptors to mean scores:

0.00 – 1.09 = No/Limited Evidence

1.10 – 1.69 = Weak

1.70 – 2.29 = Promising

2.30 – 3.00 = Strong.

Table E

To show a summary of WoE B scores for Methodological Relevance

Study	Criterion 1: Statistical Power	Criterion 2: Recruitment Strategy	Criterion 3: Support for Implementers	Overall WoE B	Descriptor
1. Bigorra et al. (2016)	3	3	1	2.33	Strong
2. Chacko et al. (2014)	1	1	2	1.33	Weak
3. Green et al. (2012)	0	2	2	1.33	Weak
4. Hovic et al. (2013)	0	3	3	2.00	Promising
5. Klingberg et al. (2005)	0	2	1	1.00	No/Limited Evidence
6. Van Dongen-Boomsma et al. (2014)	2	2	1	1.67	Weak
Mean	1.00	2.16	1.67	1.61	
Descriptor	No/Limited Evidence	Promising	Weak	Weak	

**NB: A score of zero indicates that one or none of the criteria is met or there is insufficient evidence provided.*

Table F

To show WoE B criteria for Methodological Relevance

	Criteria	*Ratings	Rationale
1	Is the sample size adequately powered to justify statistical analyses?	3. Power calculation conducted and adequate number of participants recruited accordingly. Twenty percent more participants recruited to account for potential attrition. 2. Researcher provided a rationale to justify the number of participants recruited. 1. Researcher commented on the number of participants recruited.	All of the studies used statistical analyses. The number of participants determines which statistical tests are applicable. More robust statistical tests need to be adequately powered and potential attrition factored into recruitment to facilitate this.
2	Were the strategies employed appropriate to recruit participants with ADHD?	3. Institutions which provide specific support for children with ADHD and their families were targeted. 2. Combination of criteria 1 and 3. 1. Children with ADHD and their families from schools and the general community were targeted.	Recruitment strategies which increase the homogeneity of a sample are central to methodological quality. Therefore, recruitment strategies need to maximise and increase the likelihood of obtaining a homogenous sample that represents the construct being investigated.

	Criteria	*Ratings	Rationale
3	Was adequate support provided for implementers to deliver the intervention effectively?	3. The intervention was delivered by certified professionals. 2. Training and ongoing support were adapted so that implementers conducting the intervention at home could deliver the intervention to a consistent standard. Systems for monitoring implementation were employed and responded to with agility. 1. Some training and ongoing support was provided so that implementers conducting the intervention at home could deliver the intervention. Systems for monitoring implementation were employed.	The intervention should be implemented consistently across participants, and this is more likely when certified professionals are employed. Interventions conducted at home are more difficult to control and require stringent training, support and monitoring, and an agile response to queries or lack of engagement. This is to increase consistency of intervention implementation.

**NB: A score of zero indicates that one or none of the criteria is met or there is insufficient evidence provided.*

Appendix F: WoE C Ratings for Study Relevance to the Review Question

WoE C focused on:

1. Follow-Up: Does the study conclude whether or not CWMT improves behavioural symptoms of children with ADHD?
2. Measurement: Did the primary outcome measures effectively measure children's behavioural ADHD symptoms?
3. External Validity Indicators: Was children's ADHD status confirmed effectively?

Key:

Range used to assign descriptors to mean scores:

0.00 – 1.09 = No/Limited Evidence

1.10 – 1.69 = Weak

1.70 – 2.29 = Promising

2.30 – 3.00 = Strong

Table G

To show a summary of WoE C scores for Study Relevance to the Review Question

Study	Criterion 1: Conclusive Results	Criterion 2: Outcome Measures	Criterion 3: ADHD Status	Overall WoE C	Descriptor
1. Bigorra et al. (2016)	3	3	3	3.00	Strong
2. Chacko et al. (2014)	2	3	1	2.00	Promising
3. Green et al. (2012)	2	1	2	1.67	Weak
4. Hovic et al. (2013)	0	0	3	1.00	No/Limited Evidence
5. Klingberg et al. (2005)	1	2	2	1.67	Weak
6. Van Dongen-Boomsma et al. (2014)	2	2	2	2.00	Promising
Mean	1.67	1.83	2.16	1.89	Promising
Description	Weak	Promising	Promising	Promising	

**NB: A score of zero indicates that one or none of the criteria is met or there is insufficient evidence provided.*

Table H

To show WoE C criteria for Study Relevance to the Review Question

	Criteria	*Ratings	Rationale
1	Does the study conclude whether or not CWMT improves the behavioural symptoms of children with ADHD?	3. Most primary outcomes, including follow-up outcomes, align, confirming whether or not CWMT improves the behavioural symptoms of ADHD in children. 2. Most outcome measures align, confirming whether or not CWMT improves the behavioural symptoms of ADHD in children, but there is no follow-up data. 1. Outcome measures are inconsistent such that whether or not CWMT improves the behavioural symptoms of ADHD in children with ADHD is inconclusive.	The main challenge for children with ADHD is conforming to the behavioural expectations of others. Therefore, it is important that CWMT improves children's behaviour, and that this is a key outcome for consideration when assessing the impact of CWMT on the behavioural symptoms of children with ADHD.
2	Did the primary outcome measures effectively measure children's behavioural ADHD symptoms?	3. At least two primary outcome measures measure a behavioural symptom of ADHD in children, with reliability data. 2. At least two primary outcome measures measure a behavioural symptom of ADHD in children. 1. At least one primary outcome measure measures a behavioural symptom of ADHD in children.	In order to assess if CWMT improves the behaviour of children with ADHD, it is important that outcome measures actually measure aspects of children's behaviour effectively, and are reliable. It is also important that more than one source of primary data is generated, so that results can be triangulated.

	Criteria	*Ratings	Rationale
3	Was children's ADHD status confirmed effectively?	3. ADHD status confirmed by clinical assessment prior to screening. 2. ADHD status confirmed by clinical assessment during screening. 1. Non-clinical screening for ADHD status.	The ADHD status of child participants should be clear and diagnosis should be according to standardised clinical criteria. This is so that results can be generalised to a specific population.

**NB: A score of zero indicates that one or none of the criteria is met or there is insufficient evidence provided.*

Appendix G: Summary of WoE A to C to Calculate WoE D

Weight of Evidence D (WoE D) is used to summarise the judgements for methodological quality, methodological relevance and relevance of the study to the review question into a single measure (Table I). The overall judgement was Promising (score 1.87).

Key:

Range used to assign descriptors to mean scores:

0.00 – 1.09 = No/ limited evidence

1.10 – 1.69 = Weak

1.70 – 2.29 = Promising

2.30 – 3.00 = Strong

Table I

To show how WoE A, B and C were used to calculate WoE D

Study	WoE A	WoE B	WoE C	WoE D
1. Bigorra et al. (2016)	2.52 Strong	2.33 Strong	3.00 Strong	2.62 Strong
2. Chacko et al. (2014)	2.02 Promising	1.33 Weak	2.00 Promising	1.78 Promising
3. Green et al. (2012)	1.55 Weak	1.33 Weak	1.67 Weak	1.52 Weak
4. Hovic et al. (2013)	2.38 Strong	2.00 Promising	1.00 No/Limited Evidence	1.79 Promising
5. Klingberg et al. (2005)	2.11 Promising	1.00 No/Limited Evidence	1.67 Promising	1.59 Weak
6. Van Dongen-Boomsma et al. (2014)	2.01 Promising	1.67 Weak	2.00 Promising	1.89 Promising
Mean	2.10	1.61	1.89	1.87
Description	Promising	Weak	Promising	Promising

**NB: A score of zero indicates that one or none of the criteria is met or there is insufficient evidence provided.*

Appendix H: WoE A Coding Protocols

[Adapted from the Procedural Manual of the Task Force on Evidence-Based Interventions in School Psychology, American Psychology Association, Kratochwill, T.R. (2003)]

Coding Protocol: Group-Based Design

Domain: School- and community-based interventions for social and behavioural problems.

Name of Coder: TEP	**Date:** 17/06/2022
Full Study Reference in APA format: Bigorra, A., Garolera, M., Guijarro, S., & Hervas, A. (2016). Long-term far-transfer effects of working memory training in children with ADHD: A randomised controlled trial. *European Child and Adolescent Psychiatry, 25,* 853-867.	
Intervention Name (from description of study): Working memory computer training	
Study ID Number (Unique identifier): 01	

Type of Publication: Journal article

Part I

A. General Design Characteristics

A1. 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).

A2. Overall confidence of judgment on how participants were assigned (select one 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.

A. General Design Characteristics rating

☒ 3 – Strong Evidence ☐ 2 – Promising Evidence

☐ 1 – Weak Evidence ☐ 0 - No Evidence.

B. Data Analysis

B1. Appropriate unit of analysis? ✓ Yes ❒ No

Statistical test: ___ General linear regression

B2. Familywise error rate controlled? ✓ Yes ❒ No ❒ Not applicable

B1. and B2. Statistical Analysis rating

☒ 3 – Strong Evidence ☐ 2 – Promising Evidence
☐ 1 – Weak Evidence ☐ 0 - No Evidence.

B3. Total size of sample (start of study): _66

B4. Intervention group sample size: _36

B5. Control group sample size: _30

B6. Sufficient power (adapted and added to protocol):

☒ 3 Sufficiently large N justified (Power >= 80%).
α level = 5%
Statistical power (1-*β*) = 95%
Over N recruited to account for attrition ✓ Yes ❒ No
Additional Information _____
☐ 2 Clear rationale given for sample size.
☐ 1 Some rationale given for sample size.
☐ 0 Insufficient or no rationale given for sample size.

B3. to B6. Power rating

☒ 3 – Strong Evidence ☐ 2 – Promising Evidence
☐ 1 – Weak Evidence ☐ 0 - No Evidence.

Part II

A. Research Methodology

A1. Sample appropriate to research methods. Research methods guide sampling procedures.

☒ 3 Clear links established between constructs and methods, and sampling is *appropriate* to research methods.

☐ 2 Vague or no links established between research methods and sampling, but sampling is *appropriate* to the research methods.
☐ 1 Links established between research method and sampling, but sampling is *inappropriate* to the research methods.
☐ 0 No links are established and sampling is inappropriate to research methods.

A2. Operationalisation. Specifying the link between key abstract constructs (variables)

☒ 3 Clear links established between constructs and methods, **and** all key constructs are clearly operationalised.
☐ 2 Some, but not all, key constructs are clearly operationalised.
☐ 1 Vague reference to link between constructs and methods.
☐ 0 No evidence that key constructs are operationalised.

A3. Integration of data from multiple sources, methods and investigators.

☒ 3 Used multiple sources, methods, and investigations.
☐ 2 Used two of the following: multiple sources, multiple methods, multiple investigators.
☐ 1 Used one of the following: multiple sources, multiple methods, multiple investigators.
☐ 0 No evidence of multiple sources, methods, or investigators.

A. Overall rating for Research Methodology

☒ 3 – Strong Evidence ☐ 2 – Promising Evidence
☐ 1 – Weak Evidence ☐ 0 - No Evidence.

B. Measurement

B1. Use of outcome measures that produce reliable scores for the majority of the primary 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.

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.

☐ 3 Provided evidence for high level of engagement to ensure deep and accurate representation.
☒ 2 Provided evidence for some level of engagement to ensure deep and accurate representation.
☐ 1 Provided minimal level of engagement to ensure deep and accurate representation.
☐ 0 Provided no evidence for level of engagement to ensure deep and accurate representation.

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.

B6. Cultural appropriateness of the measures. In rating this item, consider the following dimensions: meaning, language, dialect, and response format

☐ 3 Developed measure for use with target group in study on the basis of empirical evidence (conducted formative research and developed measure).
☐ 2 Adapted existing measure for use with target group on the basis of formative research and/or empirical evidence with target group.
☒ 1 Developed or adapted measures for use with target group based on empirical evidence with similar or related populations.
☐ 0 Measure not tailored specifically for target group.

B7. Measures of key outcomes are linked to the conceptual model

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

B. Overall Rating for Measurement

☐ 3 – Strong Evidence
☒ 2 – Promising Evidence
☐ 1 – Weak Evidence
☐ 0 - No Evidence.

C. Comparison Group

C1. Type of Comparison Group (Select one of the following)

☐ Typical contact
☐ Typical contact (other) specify: _____.
☐ Attention placebo.
☒ Intervention element placebo.
☐ Alternative intervention.
☐ Pharmacotherapy.
☐ No intervention.
☐ Wait list/delayed intervention.
☐ Minimal contact.
☐ Unable to identify type of comparison.

C2. Overall confidence rating in judgement of type of comparison group

☐ Very low (little basis).
☐ Low (guess).
☐ Moderate (weak inference).
☐ High (strong inference).
☒ Very high (explicitly stated).
☐ Unknown/unable to code.

C1. and C2. Rating for Type of Comparison Group

☒ 3 – Strong Evidence
☐ 2 – Promising Evidence
☐ 1 – Weak Evidence
☐ 0 - No Evidence.

C3. Counterbalancing of change agents

☐ By change agent.
☐ Statistical.
☒ Other.

C4. Group equivalence established

☒ Random assignment.
☐ Post hoc 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 _____.

C3. to C5. Rating for Confidence in Judgement, Counterbalancing, Group Equivalence and Mortality of Comparison Group

☒ 3 – Strong Evidence ☐ 2 – Promising Evidence
☐ 1 – Weak Evidence ☐ 0 - No Evidence.

D. External Validity Indicators

D1. Sampling procedures

D1.1 Sampling procedures described in detail

☒ Yes ☐ No (incomplete or no evidence)

D1.2. Rationale for sample selection specified

☒ Yes ☐ No (incomplete or no evidence)

D1.3. Rationale for sample size specified

☒ Yes ☐ No (incomplete or no evidence)

D1.4. Evidence that sample represents target population

☒ Yes ☐ No (incomplete or no evidence)

D1.5. Recruitment procedures congruent with target cultural group. Researcher used culturally appropriate ways/methods to contact, recruit, inform and maintain participation.

☒ Yes ☐ No (inadequate description or no evidence)

D1.6. Inclusion/exclusion criteria specified ☒ Yes ☐ No

D1.7. Specified criteria related to concern ☒ Yes ☐ No

D1. Overall rating on Sampling Procedures

☒ 3 – Strong Evidence ☐ 2 – Promising Evidence
☐ 1 – Weak Evidence ☐ 0 - No Evidence.

D2. Participant Characteristics specified for treatment and control groups (select all that apply).

☒ Grade/Age ☒ Gender
☒ Ethnicity or multi-ethnic ☒ Primary language
☒ Socioeconomic status ☒ Disability
☐ Location ☒ Family structure.

D3 Transferability of the intervention

☒ 3 Complete and detailed description of the context within which the intervention occurs.
☐ 2 Detailed description of some but not all contextual components.
☐ 1 Provides overview of contextual components but lack details.
☐ 0 No description of context.

D2 and D3 Rating for Participant Characteristics and Transferability of the Intervention

☒ 3 – Strong Evidence ☐ 2 – Promising Evidence
☐ 1 – Weak Evidence ☐ 0 - No Evidence.

E. Follow-up Assessment

E1. Timing of follow-up assessment

☒ Yes ☐ No ☐ Unclear

Specify _1-2 weeks post-training and 6 months post-training

E2. Number of participants included in the follow up assessment

☒ Yes ☐ No ☐ Unclear

Specify __55/66.

E3. Consistency of assessment method used

☒ Yes ☐ No ☐ Unclear

Specify _____6 months, Time 2 tests.

E. Overall Rating for Follow-up Assessment

☒ 3 – Strong Evidence ☐ 2 – Promising Evidence
☐ 1 – Weak Evidence ☐ 0 - No Evidence.

F. Implementation Fidelity

F1 Manualisation

F1.1 ❒ Written material involving a detailed account of the exact procedures and the sequence in which they are to be used.

F1.2. ❒ Formal training session that includes a detailed account of the exact procedures and the sequence in which they are to be used.

F1.3 ❒ Written material involving an overview of broad principles and a description of the intervention phases.

F1.4 ✓ Formal or informal training session involving an overview of broad principles and a description of the intervention phases.

F2 Adaptation procedures are specified ☐ Yes ☐ No ☒ Unknown

F1. and F2. Rating for Manualisation and Adaptation

☐ 3 – Strong Evidence
☒ 1 – Weak Evidence
☐ 2 – Promising Evidence
☐ 0 - No Evidence.

F3 Conditions of intervention

F3.1. Length of intervention

❒ Unknown/Insufficient information provided

✓ Information provided (if information provided, specify the following:)

☐ Weeks _5
☐ Months _____
☐ Years _____
☐ Other _____

F3.2. Intensity/dosage of intervention

☐ Frequency of intervention session _5 sessions per week (25 total)
☐ Length of intervention session _35-45mins/ 90 trials
☐ Unknown/insufficient information provided _____

F3.1 and F3.2. Rating length of intervention and intensity

☒ 3 – Strong Evidence
☐ 1 – Weak Evidence
☐ 2 – Promising Evidence
☐ 0 - No Evidence.

F3.3. Programme implementer (select all that apply)

☒ Research staff.
☐ School specialty staff.
☐ Certified coach.
☐ Teachers.
☐ Educational assistants.
☒ Parents.
☐ College students.
☐ Peers.
☐ Other.
☐ Unknown/Insufficient information provided.

F3.4 Training and support resources (select all that apply)

☒ Simple orientation given to change agents.
☐ Training workshops conducted.
☐ Little or no training specified.

Number of workshops provided _____.
Average length of training _____.
Who conducted the training (select all that apply)?
☐ Project director.
☐ Graduate/project assistants.
☐ Other (please specify).
☒ Unknown.

☒ Ongoing technical support.
☐ Program materials obtained.
☐ Special facilities.
☐ Other (specify).

F3.5 Level of difficulty in training intervention agents (select one of the following)

☐ High.
☐ Moderate.
☐ Low.
☒ Unknown.

F3.3 to F3.5 Rating for Implementer, Training and Support

☐ 3 – Strong Evidence
☒ 1 – Weak Evidence
☐ 2 – Promising Evidence
☐ 0 - No Evidence.

G. Site of implementation

☐ School.
☒ Home.
☐ Other.
☐ Hospital/clinic/health centre.
☐ University.
☐ Unknown/ not specified.

G. Rating for Site of Intervention

☐ 3 – Strong Evidence
☒ 1 – Weak Evidence
☐ 2 – Promising Evidence
☐ 0 - No Evidence.

Summary of Evidence

	Indicator	Overall Evidence Rating 0-3	Description of Evidence: Strong, Promising Weak, No/Limited Evidence
	Part I		
A.	General Design Characteristics	3	Strong
B.	Data Analysis		
1-2	Statistical Analysis	3	Strong
3-6	Power Rating	3	Strong
	Part II		
A.	Research Methodology	3	Strong
B.	Measurement	2	Promising
C.	Comparison Group		
1-2	Type of Comparison Group	3	Strong
3-5	Counterbalancing, Group Equivalence and Mortality of Comparison Group	3	Strong
D.	External Validity Indicators		
1	Sampling Procedures	3	Strong
2,3	Participant Characteristics and Transferability of Intervention	3	Strong
E.	Follow-up Assessment	3	Strong
F.	Implementation Fidelity		
1,2	Manualisation and Adaptation	1	Weak
3	Conditions of the Intervention		
3.1-3.2	Length of intervention and intensity.	3	Strong
3.3-3.5	Implementer, Training and Support, and Difficulty in Training Agents.	1	Weak
G.	Site of intervention.	1	Weak