



## Working memory training involves learning new skills

Susan E. Gathercole\*, Darren L. Dunning, Joni Holmes, Dennis Norris

MRC Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge CB2 7EF, England, United Kingdom



### ARTICLE INFO

#### Keywords:

Working memory  
Training  
Transfer  
Cognitive routine

### ABSTRACT

We present a new framework characterizing training-induced changes in WM as the acquisition of novel cognitive routines akin to learning a new skill. Predictions were tested in three studies analyzing the transfer between WM tasks following WM training. Study 1 reports a meta-analysis establishing substantial transfer when trained and untrained tasks shared either a serial recall, complex span or backward span paradigm. Transfer was weaker for serial recall of verbal than visuo-spatial material, suggesting that this paradigm is served by an existing verbal STM system and does not require a new routine. Re-analysis of published WM training data in Study 2 showed that transfer was restricted to tasks sharing properties proposed to require new routines. In a re-analysis of data from four studies, Study 3 demonstrated that transfer was greatest for children with higher fluid cognitive abilities. These findings suggest that development of new routines depends on general cognitive resources and that they can only be applied to other similarly-structured tasks.

Performance on many working memory (WM) tasks can be improved by training. However, the benefits of that training rarely transfer to other activities that also depend on WM. Why is this, and what conditions give rise to transfer? Here we present a new framework designed to explain both how and when the benefits of WM training will transfer from one task to another. Our claim is that training-induced transfer occurs only when we have learned a new complex cognitive skill in the course of training and when that skill can be applied to a novel task.

The potential of intensive training to expand on our intellectual capacities has long fascinated philosophers and psychologists. In recent years, many commercial training products have been developed for individuals keen to boost their cognitive skills (for reviews see: Bavelier, Green, Pouget, & Schrater, 2012; Simons et al., 2016; Strobach & Schubert, 2016). With extensive practice, performance on most trained tasks will improve, and gains are also reflected in changes in underlying brain systems. This kind of learning is often described as neuroplasticity. While there have been some important recent advances in understanding the impact of training on both the structure and functioning of neural networks (Astle et al., 2015; Barnes, Woolrich, Baker, Colclough, & Astle, 2016; Caeyenberghs, Metzler-Baddeley, Foley, & Jones, 2016; Salmi, Nyberg, & Laine, 2018), the field lacks detailed accounts of the cognitive changes that take place. A new framework presented here describes what these changes might be and how these both enable and constrain transfer to novel situations. We do this

in one of the most extensively investigated areas of cognitive training, WM.

Training has been investigated in many different areas of cognition ranging from rote learning, problem solving and WM through to expertise in highly specialized domains such as chess and academic learning (Simons et al., 2016 for recent review). Two broad conclusions have emerged. First, transfer is much more likely under conditions where trained and untrained activities share many features (near transfer) than few (far transfer, Barnett & Ceci, 2002; Noack, Lövdén, Schmiedek, & Lindenberger, 2009). Second, beyond this broad distinction, there is little understanding of the cognitive constraints on transfer (Shipstead, Redick, & Engle, 2010; Simons et al., 2016; Taatgen, 2013).

Our primary goal is to characterize the task features that engender transfer within WM. We propose that transfer occurs primarily when training leads to the acquisition of a new complex cognitive skill that can be applied to an untrained activity. This learning is conceptualized as the development of cognitive routines that coordinate the execution of the processes necessary to perform an unfamiliar task. For training activities supported by cognitive routines or mechanisms that are already firmly established, a new routine is not required. There is consequently less scope for transfer within these activities, even if the tasks are very similar.

Predictions derived from the framework about the conditions under which transfer to other WM tasks is expected to be strongest are tested

\* Corresponding author.

E-mail addresses: [susan.gathercole@mrc-cbu.cam.ac.uk](mailto:susan.gathercole@mrc-cbu.cam.ac.uk) (S.E. Gathercole), [Darren.dunning@mrc-cbu.cam.ac.uk](mailto:Darren.dunning@mrc-cbu.cam.ac.uk) (D.L. Dunning), [joni.holmes@mrc-cbu.cam.ac.uk](mailto:joni.holmes@mrc-cbu.cam.ac.uk) (J. Holmes), [dennis.norris@mrc-cbu.cam.ac.uk](mailto:dennis.norris@mrc-cbu.cam.ac.uk) (D. Norris).

<https://doi.org/10.1016/j.jml.2018.10.003>

Received 16 October 2017; Received in revised form 15 October 2018

Available online 01 December 2018

0749-596X/ © 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

in three studies. Study 1 provides a meta-analysis of published randomized controlled trials (RCTs) of WM training. The aim was to discover which features common to both trained and untrained tasks are associated with transfer, and to establish the magnitude of any transfer that does occur. Studies 2 and 3 re-analyze data from several published studies of Cogmed training in children to test whether transfer is indeed mediated by the development and application of new routines. Study 2 investigates transfer across WM tasks following training on a single WM program. Study 3 examines the sources of individual differences in transfer following WM training in a large sample of children with the aim of establishing whether transfer originates in the WM system itself or from more general cognitive resources.

### Transfer from WM training

WM is the cognitive system responsible for the temporary storage and manipulation of information (Baddeley & Hitch, 1974). Substantial and enduring benefits of intensive training for performance on similar untrained WM tests have been widely reported (e.g. Dunning, Holmes, & Gathercole, 2013; Jaeggi, Buschkuhl, Jonides, & Perrig, 2008; Klingberg et al., 2005). These gains have been found at all ages from the preschool years through to late adulthood, and for individuals with a range of developmental and acquired disorders of cognition (Karbach & Verhaeghen, 2014; Melby-Lervåg & Hulme, 2013; Sonuga-Barke et al., 2013; Wass, Scerif, & Johnson, 2012). Enhancements in cognitive performance following training are mirrored by changes in the neural activity and connectivity of the brain networks serving WM (Astle et al., 2015; Barnes et al., 2016; Caeyenberghs et al., 2016; Constantinidis & Klingberg, 2016; Dahlin, Neely, Larsson, Backman, & Nyberg, 2008; Kundu, Sutterer, Emrich, & Postle, 2013; Olesen, Westerberg, & Klingberg, 2004; Takeuchi & Kawashima, 2012; Takeuchi, Sekiguchi et al., 2010; Westerberg & Klingberg, 2007). These findings indicate that WM training may be capable of inducing fundamental and enduring improvements to a cognitive system critical for everyday functioning across the lifespan.

The therapeutic potential of WM training is far-reaching. WM is impaired in developmental disorders of attention (Holmes et al., 2014; Martinussen, Hayden, Hogg-Johnson, & Tannock, 2005), language (Hesketh & Conti-Ramsden, 2013; Montgomery, 2000; Montgomery & Evans, 2009; Newbury, Bishop, & Monaco, 2005; Pimperton & Nation, 2012; Ramus, Marshall, Rosen, & van der Lely, 2013; Schuchardt, Bockmann, Bornemann, & Maehler, 2013), reading (Swanson, Xinhua, & Jerman, 2009) and mathematics (McLean & Hitch, 1999; Swanson & Beebe-Frankenberger, 2004; Szucs, Devine, Soltesz, Nobes, & Gabriel, 2013; Szucs, Nobes, Devine, Gabriel, & Gebuis, 2013). It is also disrupted following brain injury (Cicerone & Giacino, 1992; Dunning, Westgate, & Adlam, 2016; Phillips, Parry, Mandalis, & Lah, 2017). In typically-developing populations, variation in WM skills is closely linked with practical abilities such as maintaining focused attention in everyday life and following instructions (Engle, Carullo, & Collins, 1991; Gathercole, Durling, Evans, Jeffcock, & Stone, 2008; Kane, Bleckley, Conway, & Engle, 2001).

To date the evidence that WM training can ameliorate these practical problems is not compelling. There have been reports that WM training yields benefits in selective attention (Klingberg et al., 2005), fluid intelligence (Jaeggi et al., 2008), reading (Loosli, Buschkuhl, Perrig, & Jaeggi, 2012), mathematical abilities (Holmes, Gathercole, & Dunning, 2009), and ratings of mental states and behaviors including mood (Åkerlund, Esbjörnsson, Sunnerhagen, & Björkdahl, 2013), inattentivity and hyperactivity/impulsivity (Klingberg et al., 2005), occupational satisfaction (Lundqvist, Grundstrom, Samuelsson, & Ronnberg, 2012) and behavioral self-regulation (Foy & Mann, 2014).

Positive outcomes such as these are outnumbered by null findings. Many systematic reviews and meta-analyses of WM training have been published in recent years and with some exceptions (Au et al., 2015; Spencer-Smith & Klingberg, 2015), the consensus is that training

benefits are largely confined to near transfer to other similar WM tasks (Cortese et al., 2015; Karbach & Verhaeghen, 2014; Melby-Lervåg & Hulme, 2013; Rapport, Orban, Kofler, & Friedman, 2013; Redick, Shipstead, Wiemers, Melby-Lervåg, & Hulme, 2015; Schwaighofer, Fischer, & Bühner, 2015; Shipstead, Redick, & Engle, 2012; Simons et al., 2016; Sonuga-Barke et al., 2013; Soveri, Antfolk, Karlsson, Salo, & Laine, 2017; Weicker, Villringer, & Thöne-Otto, 2016). Lack of far transfer is most evident in studies employing rigorous intervention methodologies such as randomized controlled trials (RCTs), active control conditions, and blind assessments before and after the intervention.

### Theories of transfer

#### Neuroplasticity

An appealing explanation for transfer is that the WM gains observed following adaptive training reflect cortical plasticity in the neural system underpinning WM. Klingberg (2010) speculated that WM training “might lead to durable neuronal changes in WM-related areas in the same way as perceptual training does for neurons of the visual cortex” (p. 318). Westerberg and Klingberg (2007) suggested that this could be mediated by changes in the response characteristics of single neurons, possibly reflecting plasticity in cellular components including synapses and dendrites. Takeuchi, Taki, and Kawashima (2010) proposed that training enhances the structure of the white matter tracts in the neural system underpinning WM. These plasticity-based accounts resonate with evidence for neural changes following intensive training in motor activities such as repetitive neural stimulation of fingers in primates (Xerri, Merzenich, Peterson, & Jenkins, 1998), perceptual discrimination learning in monkeys (Law & Gold, 2008), and in golf and juggling in humans (Bezzola, Mérillat, Gaser, & Jäncke, 2011; Draganski et al., 2004; May 2011).

The problem with this concept of neuroplasticity is that it fails to explain why training has so little benefit for aspects of everyday cognitive functions that are widely considered to depend on WM. If the neural efficiency of WM improves with training, its benefits should extend to these activities too. In fact, even transfer within WM is limited. Consider n-back and complex span, two common WM paradigms. In n-back, participants judge whether for each item in a lengthy sequence is the same as the item that appeared n positions back (1 item, 2 items, etc.). In complex span, an unrelated processing activity is interpolated between the presentation of successive memory items (Daneman & Carpenter, 1980; Turner and Engle, 1989). A recent meta-analysis of n-back training established that the magnitude of transfer to WM paradigms such as complex span is very small (Soveri et al., 2017).

#### Process-specific transfer

An alternative explanation is that rather than expanding the fundamental capacity of the system in an undifferentiated manner, WM training enhances the specific processes within WM that are engaged by particular tasks (Dahlin et al., 2008; Dunning & Holmes, 2014; Holmes et al., 2009; Minear et al., 2016; Shipstead et al., 2012; Soveri et al., 2017; Sprenger et al., 2013; von Bastian & Oberauer, 2013a). This approach accounts for the absence of transfer across WM paradigms by assuming that training originates in processes in WM such as updating, inhibitory function and short-term memory (STM) storage that are engaged by some but not all WM tasks (STM, Dahlin et al., 2008; Minear et al., 2016). Transfer should only be observed when training and transfer tasks both place demands on the same processes.

Participants in training studies often report using mnemonic strategies (Holmes et al., 2009; Minear et al., 2016), and these too are also potential sources of training-induced change. Strategy transfer will necessarily be limited by the ways in which the stimuli in the untrained stimuli can be represented. Training a mental imagery strategy to assist

recall of lists of concrete nouns would not, for example, be expected to benefit the recall of either abstract nouns or movements. Evidence for the content-specificity of mnemonic strategies is provided by Chase and Ericsson's (1981) study of an individual completing a lengthy period of digit span training. SF began with a typical digit span of seven items which had expanded to 79 items after two years of training. He reported that this was achieved by recoding digit sequences into long-distance running times that he was familiar with as a runner. Tellingly, his memory span for letter sequences over the same period did not change. The capacity of verbal STM *per se* was therefore unchanged. Similar conclusions were reached in a study of two adults who trained on digit span for a period of four months reached (Martin & Fernberger, 1929).

Strategies may be of limited value even when the stimuli are the same if the WM tasks change. Minear et al. (2016) asked participants completing either spatial n-back or verbal complex span training to describe the mnemonic strategies they had used. Although letters were the memoranda in both cases, participants reported different strategies. N-back trainees employed many different strategies, although a substantial minority reported using no strategy. Participants who underwent verbal complex span training described strategies involving chunking the letter sequences in some way, for example by forming associations between the letters and words and then forming sentences. A transfer measure of serial recall of letters was included in this study. The authors reasoned that if the gains on the trained complex span task reflected the development of such material-specific strategies, they should extend to this task too. No such benefits were found. It therefore appears that any letter-chunking strategy must have been tied in with the broader information processing demands of the paradigm that was originally trained, limiting its transfer (see also, von Bastian & Oberauer, 2013b).

A limitation of process-specific accounts of training to date is that they do not distinguish shared task features that will be sufficient for transfer from those that will not. Often, these accounts are advanced speculatively to explain unexpected transfer rather than generating specific hypotheses that are directly tested in new studies (Sprenger et al., 2013; von Bastian, Langer, Jancke, & Oberauer, 2013). One proposal is that the magnitude of transfer is related to the extent of task overlap, with highest levels of transfer for tasks with the greatest numbers of shared task features (Soveri et al., 2017). As we shall see later, the presence of shared task elements alone is not a sufficient explanation for either the presence or absence of its transfer, or its magnitude. One of the greatest challenges for existing theories of transfer is therefore not only to explain why transfer arises, but also why it does not.

### *Cognitive training as skill acquisition*

Here we present a new perspective on transfer following WM training. This broadly conceptualizes transfer as a consequence of acquiring complex cognitive skills that can then be applied to untrained tasks with similar demands. It has its origins in production system models that represent skilled behavior as sets of production rules incorporating specific knowledge (Anderson, 1982; Newell, 1991). Complex new activities are accomplished by combining these rules. The execution of the rules becomes increasingly automatic with practice, a developmental process characterized in Anderson's ACT-R model as progression from a declarative to a procedural stage. As learning progresses, the demand on limited resources diminishes and this leads to performance gains. Transfer arises when the production rules can be applied to new tasks (Singley & Anderson, 1989).

Taatgen (2013) incorporated new principles into a production system framework that provide more specific predictions about the conditions for transfer. His primitive elements theory of cognitive skills distinguishes individual low-level elements of production rules that are entirely specific to a particular task, from task-general skills that control the flow of information across the task independently of content.

Transfer occurs when the task-general skills are consistent, even if tasks differ in low-level task features. This approach was applied to model data showing transfer from complex WM span training to performance on a Stroop interference task (Chein & Morrison, 2010). Transfer was modelled as an increase in a process of proactive (executive) control that corresponds to a high-level executive state of planning. This enhanced both rehearsal in the WM task and the selection of the ink color in Stroop.

Primitive elements theory places no limits on the transferability of task-general skills across tasks: transfer may in principle occur for any tasks requiring common higher-order processes such as proactive control. We will see later that this is not necessarily the case for WM training, in which some paradigms are more trainable than others. Other models of skill acquisition do impose constraints on transfer, and these offer some insights as to why transfer within WM may not always occur even within paradigms. Fitts and Posner (1967) suggested that learning progresses through three stages: cognitive, associative, and autonomous. As an example, the cognitive stage of acquiring arithmetic skills might involve performing multiplication by explicit calculation. In the associative stage, the results of calculations would already have been stored in long-term memory, requiring only search followed by retrieval. In the autonomous stage, this process would operate automatically. A reasonable expectation is that for typical participants in WM training studies many of the basic processes of storage and retrieval within WM will already be fully established and will have reached the autonomous stage. This will leave relatively little scope for further refinement even with the extensive practice provided in a WM training program. Neither training nor transfer would therefore be expected for tasks supported by systems such as verbal STM that are already fully established.

Some of these principles are incorporated into the present framework. We propose that in many complex WM tasks, training cannot be accomplished by established configurations of processes within WM. Participants must therefore learn how to perform these unfamiliar tasks. This form of learning follows the conventional path to acquiring a new skill. It starts with a period in which execution and coordination of its components are highly demanding of cognitive resources. With experience the skill becomes more autonomous, improving performance. Once established, the new skill will transfer to other tasks with similar structures. For WM tasks that are already served by existing mechanisms there will be much less scope for training or transfer, because the configurations of processes needed to support them are already in place. The framework builds on these principles to explain the limits on WM training and its transfer to new tasks.

### **The cognitive routine framework**

We make two assumptions about WM training. First, training on unfamiliar WM tasks will lead to the development of novel cognitive routines that control the sequence of cognitive processes required to perform the task. Second, these routines can be applied only to other tasks with common structures and only then will transfer occur. A cognitive routine is a structured specification of the coordinated sequence of processes that must be implemented to accomplish a mental activity. In the initial stages of performing a complex WM task, general cognitive resources are required to determine the optimal sequence of the processes, and to execute the routine. With practice, the execution of the routine will become more autonomous, mirroring changes seen in the acquisition of other cognitive skills (Tenison & Anderson, 2016).

A new routine is needed to martial and execute existing processes in a novel sequence when a task has complex and unfamiliar cognitive requirements. For complex WM tasks, it is envisaged that the routines will have a hierarchical structure composed in part of sub-routines repeatedly executed across the course of a trial. New routines and sub-routines may also support mnemonic strategies such as mental imagery and grouping items into larger meaningful chunks. These strategies

require sequences of processes integrating the to-be-remembered material with more permanent knowledge outside of WM. For example in using mental imagery, knowledge must be retrieved from semantic memory to generate visuo-spatial representations for temporary storage. To use a chunking strategy, items in WM must be supplemented by or linked with representations of stimuli from long-term memory bound into multi-item chunks (Cowan, Roudier, Blume, & Sauls, 2012; Miller, 1956). In these ways, strategies involve the coordination of processes in WM with systems in long-term memory. To do this, we propose that a new cognitive skill (or routine) has to be learned.

Current process-specific theories of transfer focus on the features common to both the training and transfer tasks. The framework takes two further steps, specifying both the conditions under which training itself will occur, and why. A core assumption is that transfer following WM training is restricted to cases where the training tasks require the establishment or refinement of a cognitive routine that is not already fully developed. If a routine is already well-established there will be little scope for either training or transfer even if the trained and untrained tasks both call upon the same routines and processes. One domain where the basic cognitive routines will already be well established is verbal short-term memory (STM). The encoding of item and order information and the engagement of a maintenance rehearsal process are core elements of this system which can readily account for many key verbal serial recall phenomena (Burgess & Hitch, 1992; Page & Norris, 1998). It is frequently engaged in everyday occasions such as remembering new words and names, following instructions, and remembering unfamiliar phone numbers, PIN codes and passwords. As the skills required to perform these tasks will have already been acquired there will be no need to develop new routines. If no new routines are developed, there will necessarily be no routine-mediated transfer.

A computer analogy is useful here to highlight the differences between this approach and concepts of plasticity that emphasize the malleability of the neural processes underpinning WM capacity. The framework represents a shift away from thinking of training as a way of modifying the hardware of WM towards viewing it as the generation of new software. This software controls both the operation of the hardware of WM and the interface between WM and other cognitive systems. The capacity of buffers in WM could be considered to be hardware whereas the control of rehearsal or of a chunking strategy would be controlled by software. As discussed above, there is little indication that the storage capacity of verbal STM (hardware) can be increased by training (Chase & Ericsson, 1981; Martin & Fernberger, 1929). However, the ability to rehearse (software) can be trained in non-rehearsing individuals (Broadley & MacDonald, 1993; Johnston, Johnson, & Gray, 1987).

Continuing with the computing analogy, we can see that there may be choices about the particular forms of software (routines) that will influence transfer. For instance, a function to reverse the order of a sequence might be written in a way that accepts only a list of spoken digits as an argument. Such a function would improve performance on the backward recall of spoken digits, but would be of no value for the backward recall of written letters. Alternatively, the function might be written in a more general fashion so that it could accept written and spoken words, digits, letters, and even visual objects as arguments. Either function will result in improved performance on a training task using digits, but only the more general function will be transferable to different materials. The basic capacity of the system would be unchanged but because its software differs, so too will the extent of transfer. It will also be influenced by the particular software solutions selected by the programmer.

To develop firm predictions about transfer we need to be able to say something about the detailed routines employed. Although we can speculate on grounds of principle alone, the best guide to the nature of these routines and the limits of their transferability is provided by hypothesis-driven experimental analysis. A good example of this is Chase and Ericsson's (1981) digit span training study of SF. The tenfold

increase in SF's digit span across two years of practice does not tell us whether he had developed a general-purpose serial recall routine or one specific to digits. However, his description of chunking digits in terms of running times led to the prediction that this strategy (or routine) would not transfer to letters. This was indeed found to be the case, confirming that the routines developed during training were indeed tied to a specific set of stimuli.

Up to this point, we have argued that training in complex tasks involving the novel coordination of existing processes or the development of new mnemonic strategies will lead to the construction of cognitive routines that are the source of transfer to other routine-compatible tasks. An important caveat is that training-induced changes also originate outside of routines in established processes. A wealth of evidence indicates that performance on almost any cognitive activity, including basic low-level visual discrimination of perceptual features, shows gradual improvement with training (for review see Bavelier et al., 2012). Performance on almost all speeded tasks also continues to improve to some degree, even after extensive practice, a ubiquitous phenomenon termed the "law of practice" (Newell & Rosenbloom, 1981). If components of a WM task can be performed faster, then this too should enhance performance (Barrouillet, Bernardin, & Camos, 2004).

It is therefore likely that extensive practice on all WM tasks will produce some fine-tuning in the efficiency of established processes. These subtle changes may generate relatively small degrees of transfer that cannot be reliably detected in the low- to moderately-powered studies that dominate WM training research. With larger sample sizes or more data-intensive psychophysical testing, however, they should be evident. The primary goal here is to understand the origins of the more substantial effect sizes that can be detected in the studies that typify the field of WM training research. We suggest that these moderate to large transfer effects are the hallmark of routine-mediated learning during training.

### A cognitive taxonomy of WM tasks

To generate predictions from the cognitive routine framework about transfer across WM tasks it is necessary to distinguish the tasks that need new routines from those that can be supported by existing processes. To do so requires the development of a cognitive taxonomy of WM tasks. Deriving such is not straightforward because there are many conceptually distinct theories and models of WM that also differ in the scope of the paradigms they address. Baddeley, Hitch and colleagues developed a highly influential modal model of WM that has framed much of the research in the field (Allen, Baddeley, & Hitch, 2006; Baddeley, 2000; Baddeley & Hitch, 1974). At its heart is a limited-capacity central executive sub-system supplemented by an episodic buffer that binds temporary representations both within and beyond WM. Further buffers provide limited and specialized storage for verbal and visuo-spatial material (Baddeley, 1986, 2000; Baddeley & Della Sala, 1996; Baddeley, Lewis, & Vallar, 1984). Cowan, Engle and others have conceptualized WM not as a separate storage medium but as long-term memory (LTM) representations temporarily boosted via a limited attentional resource (Cowan, 1998; Cowan & Morey, 2007; Engle, Tuholski, Laughlin, & Conway, 1999). Others have located WM within a broader framework of executive functions (von Bastian & Oberauer, 2013a) or as the combined product of two parallel memory systems (primary and secondary memory) with distinct temporal and organizational features (Shelton, Elliott, Matthews, Hill, & Gouvier, 2010; Unsworth & Engle, 2007). Even when models focus on the same paradigms there is little consensus about the nature of the component processes (e.g. Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012; Towse, Hitch, & Horton, 2007).

In the absence of theoretical convergence, the current taxonomy was generated from an evidence-based task analysis. The analysis was



confined to the paradigms required the serial recall of list items that dominate the current generation of WM training. In those paradigms in which evidence points to domain-specific differences, verbal and visuo-spatial tasks are analyzed separately.

### Verbal serial recall

Serial recall of verbal material is supported by a multi-component system of verbal STM. It consists of processes responsible for encoding item and order information, and for linking together the two sets of representations. Key phenomena including serial position and transposition functions have been successfully modelled as associations between temporary representations of each item and either temporally-evolving context or order signals that can be decoded to retrieve item position or order (e.g. Burgess & Hitch, 1992, 1999; Page & Norris, 1998).

Older children and adults spontaneously use a verbal rehearsal strategy to enhance serial recall (Gathercole & Hitch, 1993; Hitch et al., 1983). This has been suggested to involve the reactivation of phonological representations in STM as a means to offset time-based decay, possibly through a process of covert articulation (Baddeley et al., 1984; Baddeley, Thomson, & Buchanan, 1975). Rehearsal has been modelled as the re-presentation of the stored sequence back into verbal STM (Burgess & Hitch, 1992; Page & Norris, 1998). Once established at around seven years, rehearsal is a highly effective strategy for retaining information in verbal STM (Flavell et al., 1966).<sup>1</sup> What happens prior to this developmental milestone is not fully understood. Recent work indicates that rehearsal may be underestimated in children performing poorly on verbal STM as a consequence of low measurement sensitivity to experimental indicators of rehearsal (Jarrold & Citroën, 2013; Jarrold, Tam, Baddeley, & Harvey, 2010). But most importantly for the present purposes, it is reasonable to assume that the contribution of rehearsal to verbal serial recall develops over the early school years and becomes fully functional by seven years or so.

Does training on verbal STM training tasks transfer to similar untrained tasks? A key assumption of the framework is that established processes such as those involved in encoding verbal item and order information should not be amenable to further training. Verbal STM performance is therefore predicted to be relatively impervious to training once rehearsal has been established. Performance may, however, be enhanced by the adoption of new material-specific strategies under conditions of extensive and prolonged practice (Chase & Ericsson, 1981; Martin & Fernberger, 1929).

With sufficient practice and instruction, rehearsal can be induced in non-rehearsing children, leading to increases in memory span (Broadley & MacDonald, 1993; Johnston et al., 1987). Training gains can extend to serial recall for untrained verbal content: Comblain (1994) showed that training individuals with Down syndrome to rehearse word lists led to benefits in digit span. Thus, even without explicit strategy instruction, WM training programs that provide extensive practice in verbal serial recall for children who are not yet rehearsing may provide catalytic conditions for a new rehearsal routine (Holmes, Butterfield, Cormack, Loenhoud, Ruggero, Kashikar, & Gathercole, 2015). In older children and adults, the routine will already be well-established and hence not be amenable to further training.

When the number of items to be recalled is close to span, verbal memory depends primarily on phonological coding, although for supra-span sequences there may be a shift to non-phonological strategies (Gathercole & Baddeley, 1990; Salamé & Baddeley, 1982). Other non-phonological strategies such as semantic linkage and visual imagery can

also be beneficial in immediate memory tasks (McNamara & Scott, 2001; St Clair-Thompson, Stevens, Hunt, & Bolder, 2010; Turley-Ames & Whitfield, 2003). These recoding strategies may increase the depth of processing (Craik & Lockhart, 1972), permit the generation of multiple representations for each item (Paivio, 1990) and allow multiple items to be formed into single chunks (Cowan, Chen, & Rouder, 2004). We assume that for most individuals these strategies require the development of new routines that have the potential to transfer to other WM tasks with similar stimulus content.

In summary, two aspects of verbal STM may require the development of cognitive routines and hence to yield transfer to other verbal STM tasks. The first is subvocal rehearsal in pre-rehearsing children, and the second is the development of new mnemonic strategies. The basic mechanisms of encoding item and order information in verbal STM are already in place early in childhood and do not warrant new routines. Transfer of verbal STM training is therefore expected to be minimal for older children and adults unless novel mnemonic strategies are developed.

### Visuo-spatial serial recall

The representations and processes involved in visuo-spatial serial recall are much less well understood. Standard paradigms involve a variety of stimulus forms including spatial locations, continuous movements, static patterns, unfamiliar objects, and scenes. Hallmark experimental phenomena indicate that mechanisms encoding serial order may be similar in the verbal and spatial domains (for review see Hurlstone, Hitch, & Baddeley, 2014). However, verbal and visuo-spatial memory span are largely independent in both children and adults (Alloway, Gathercole, & Pickering, 2006; Baddeley, Papagno, & Vallar, 1988; Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999). Further evidence indicates that within this system, visual characteristics and spatial locations may be stored separately (Darling, Della Sala, Logie, & Cantagallo, 2006; Logie & Pearson, 1997; Pearson, Ball, & Smith, 2014; Pickering, Gathercole, Hall, & Lloyd, 2001), with serial spatial rehearsal providing a means of maintaining either visual or spatial representations (Logie, 1995). This appears to be accomplished through the covert control of eye movements (Awh, Vogel, & Oh, 2006; Logie & Pearson, 1997; Waters, Rochon, & Caplan, 1992).

While the wealth of evidence distinguishing between the STM processes involved in verbal and visuo-spatial STM is undisputed, the extent to which it reflects separate but analogous temporary storage for the two domains is a matter of current debate. For many years, the most prominent position has been that information is represented in terms of its visual or spatial characteristics in the relevant domain-specific store (STM) within WM (Baddeley, 2012; Logie, 1995). This conclusion has been strongly challenged in comprehensive analysis of experimental and neuropsychological evidence by Morey (2018). The issue here is not whether verbal and visuo-spatial STM can be distinguished, but whether STM for visuo-spatial material is domain-specific in nature. There is substantial evidence that it is not. Across many studies employing a wide range of paradigms, it has been shown that visuo-spatial tasks show a far greater reliance on general attentional resources than their verbal equivalents (Alloway et al., 2006; Kane et al., 2004; Morey & Miron, 2016; Pearson et al., 2014; Thompson et al., 2006). On this basis Morey concludes that “neither the neuropsychological evidence nor the dual-task literature provides strong support for a dedicated visual-spatial STM system” (p. 876).

One reason why a specialized STM system may not have evolved for this domain is that, unlike verbal material, recall of the order of visuo-spatial events is rarely required in everyday life. In our terms, this could mean that there is no established STM system to support this material. In order to make improvements during training, participants must therefore develop new cognitive routines that draw initially on domain-general cognitive resources. These routines may involve learning to exploit the unique configurations of visuo-spatial stimuli in particular

<sup>1</sup> On the basis of simulations Lewandowsky and Oberauer (2015) have argued that there is no evidence for the effectiveness of rehearsal. However, their simulations are based on the assumption that people make errors in rehearsal which they then further rehearse leading to an accumulation of errors.

tasks. For sequences involving temporally dynamic information such as Corsi block recall (De Renzi & Nichelli, 1975), this may involve encoding properties that can include multiple transitional features including the lengths of spatial paths, their crossing points and their angles (Parmentier, Elford, & Maybery, 2005). Neuroimaging data suggests that exploiting some of these spatial properties may impose significant attentional burdens: prefrontal activity associated with the multiple demand system increases when spatial sequences have properties that encourage their recoding into higher-order chunks (Bor, Duncan, Wiseman, & Owen, 2003).

In summary, verbal and visuo-spatial serial recall are both supported by domain-specific processes that encode and maintain item and order information. A difference that may turn out to be important for transfer is that visual-spatial STM shows a greater dependence on general cognitive resources than verbal STM. This may provide scope during training for the development of new cognitive routines (or the refinement of existing ones) tailored to meet the relatively unfamiliar task demands. If so, training on visuo-spatial STM should generate greater transfer to untrained tasks with similar recall demands than the corresponding verbal STM tasks.

### WM tasks

Most training programs include complex WM tasks that combine the temporary storage demands of simple serial recall tasks with additional processing requirements such as changing the order of items at recall, updating the contents of WM, or handling irrelevant distraction. They share the common feature of requiring participants to store material in highly unfamiliar and challenging cognitive conditions. To cope with these unfamiliar conditions it will be necessary to develop novel cognitive routines.

The cognitive routine framework predicts that these paradigms will not generate transfer unless the untrained task shares the same unfamiliar demands. Transfer will be determined by the fit of the coordinating structure of the routine (for example, executing a particular form of memory updating or reversing the sequence of items at recall) to an untrained task, and not simply by the overlap with individual processes embedded within subroutines. Significant mismatches in task structure will prevent transfer. We suggest that routines readily adapt to changes in lower-level features such as the modality of an input (e.g., auditory or visual) or a response (e.g., spoken or mouse click) that preserve the higher-level structure of the routine. Taatgen (2013) drew a similar distinction between low-level elements of production rules specific to a particular task and the task-general skills that alter the flow of information across the production system. In his conceptualization, transfer can only occur across individual elements in the production system if the same task-general skills can be applied across two tasks.

### Complex span

Complex span tasks differ from simple serial recall tasks through the interpolation of episodes of distractor processing between successive memory items. For example, in operation span (Turner and Engle, 1989), sequences of words are presented for serial recall and, after each word is presented, participants must read and verify an arithmetic problem such as “Is  $(4/2)-1 = 1$ ?”. After the final calculation the participant attempts serial recall of the word list. An example of a visuo-spatial complex span task is symmetry span. In this task, the distractor activity involves judging the symmetry of a pattern and the items to be remembered are the locations of squares presented successively in a matrix (Redick & Engle, 2011).

In these tasks, participants must work out how to protect the memory representations from the interference or decay that might be caused by the punctuating periods of distraction. Several ways in which this could be achieved have been proposed. One is through rapid switching between processing of the distractor events and rehearsal of the memory items (Towse, Hitch, & Hutton, 1998). An alternative

proposal is that participants might use attentional refreshing to revive decaying representations by switching between distractor processing and rapid serial reactivation of the encoded memory sequence (Barrouillet, Gavens, Vergauwe, Gaillard, & Camos, 2009). Oberauer et al. (2012) implemented a very different account of complex span in their SOB-CS (serial-order-in-a-box – complex span) neural network model. In this, interference is generated by the unwanted encoding of distractor items resulting from a novelty-gating mechanism. This is minimized by the active removal of distractor representations with the aim of restoring the quality of earlier memories. Note that in each one of these accounts, a novel set of cognitive processes (time-switching to permit rehearsal or attentional refreshing, or active removal of distractors) is required to meet the unusual needs of the particular complex span task. From our perspective these represent novel cognitive routines developed across the course of an extended training program that can then be applied to other similarly-structured tasks.

Beyond this point, it is not possible to make firm predictions about the limits on the transferability of a complex span routine to other untrained complex span tasks. We would certainly expect routines to adapt readily to changes in superficial task features such as the sensory modality of inputs or outputs that call on relatively peripheral and specialized processing systems, as these have few consequences for the higher-level structuring of task processes. Whether more profound mismatches such as changes in the interpolated distractor activities will be sufficient to prevent transfer to an untrained task is less clear. The overlap in the cognitive processes involved in performing distractor tasks such as verifying equations and judging whether letters are upright or mirror-reversed when mentally rotated (Harrison et al., 2013) is minimal. Indeed, the distractor activities employed in the trained and untrained complex span tasks differ in all relevant training studies familiar to the authors.

What the different distractor activities in complex span tasks do share is their functional position in the higher-level of the task structure: in each case, they are interpolated between memory items and disrupt stimulus maintenance and encoding processes that might otherwise take place. In most complex span tasks, the distractor activities are unrelated to the to-be-recalled stimulus items.<sup>2</sup> The different distractor activities may therefore be supported by substitutable subroutines within a broader routine common to multiple complex span tasks designed to minimize distraction and maintain stimulus representations. Could a match to this higher-order structure be sufficient to allow the routine to be applied to complex span tasks involving different distraction, or is the tolerance to deviance in task structures limited to more superficial tasks elements? The transfer data from the meta-analysis in Study 1 directly address this question.

### Re-sequencing

Another way in which complexity can be introduced into WM tasks is by changing the sequence in which memory items should be recalled. The most common re-sequencing task is backward span in which participants are instructed to recall lists in reverse sequence. Compared with forward recall, backward recall is generally slow (Anders & Lillyquist, 1971) and errorful (Isaacs and Vargha-Khadem, 1989). Some participants report doing this by engaging in successive forward retrievals in order to peel off the numbers backwards (Anders & Lillyquist, 1971; Conrad, 1965; Thomas, Milner, & Haberlandt, 2003). First, the whole list is run through and the final item reported. The process is then repeated, each time reporting what has now become the final unrecalled item (1,2,3,4 ...1,2,3..., 1,2... etc.).

Like any ad hoc strategy designed to solve the unusual problem of reversing an input sequence, this strategy requires a new routine for the

<sup>2</sup> An exception is listening span, in which the recall items are the final words in interpolated sentences that participants read aloud (Daneman & Carpenter, 1980)

recall phase. Existing cognitive processes must be coordinated in a novel way to make repeated forward covert retrieval attempts. For verbal stimuli, this is likely to involve control of the rehearsal process. As storage in verbal STM is phonological rather than semantic in nature (Baddeley et al., 1975; Gathercole, Frankish, Pickering, & Peaker, 1999), the same routine should be readily extended to backward span tasks employing different categories of verbal stimuli such as word, letters and digits. Transfer is therefore expected to any backward span tasks employing verbal material.

It is less clear whether training in backward span should transfer across verbal and visuo-spatial domains. Whereas backward recall leads to much lower memory span for verbal stimuli, its impact is minimal when the task involves recalling spatial sequences (Isaacs and Vargha-Khadem, 1989). This raises the possibility that the backward recall of spatial sequences may not involve the effortful and time-consuming peeling-off strategy that can be applied in backward digit span (Norris, Hall, & Gathercole, under review). If this is the case, the routines verbal and visuo-spatial will differ substantially and this will limit transfer.

Other re-sequencing tasks include recalling numbers in mixed lists in numerical order and recalling letters according to either alphabetical order (Wechsler, 2008) or semantic category (Sheslow & Adams, 1990). These tasks are also unfamiliar and highly challenging. They require use of stored knowledge such as numerical sequences, the alphabet, and object category to guide retrieval and output. Distinctive routines will be required to meet each specific task requirement. No transfer is therefore predicted across these different re-sequencing variants.

#### Updating

Many WM training programs use tasks that involve continuous updating of the to-be-remembered items in lists of unknown length. The two updating paradigms used most frequently in training studies are n-back and running span.

**N-back.** In n-back tasks, participants encounter a lengthy sequence of items and must judge, for each item, whether it matches the item presented n positions back. There are a number of ways in which n-back tasks might be performed. Most accounts of n-back have focused on how the potential set of memory items might be updated with successive presentations. Alternatively, participants might repeatedly break and then reconstruct bindings between item and order information as each new item is presented.

Chatham et al. (2011) developed a neurally inspired model of n-back. In line with the idea that n-back requires the development of a novel routine, it has to be trained explicitly to perform the task with a specific value of n. Juvina and Taatgen (2007) presented behavioral evidence indicating that n-back may be supported by at least two different spontaneous strategies. One group of participants showed decreases in recognition accuracy at later positions. This was interpreted and simulated in an ACT-R production system model as reflecting a strategy of maintaining an active rehearsal set of n items, in combination with inhibition of items that drop out of the window. The remaining participants showed a flatter serial position function simulated in an alternative model without rehearsal.

There are two key points here. The first is that, irrespective of the specific cognitive processes required to support n-back, the task demands are so unfamiliar and challenging that they cannot be met by ready-to-go mechanisms. We argue that this necessitates the development of a new routine specifying the set of coordinated processes that need to be performed. Second, there is more than one way to perform n-back and it looks as though individuals choose between alternative strategies, and hence routines. We speculate that this variability in how tasks are performed even without training may be a hallmark of the unfamiliar complex tasks that benefit from the acquisition of new routines during training. Similar findings have been reported both for running span, in which participants may adopt either a passive strategy or an active updating strategy (Hockey, 1973) and in backward span

(Norris et al., in preparation).

As with all complex WM tasks requiring novel routines, on-task performance is expected to improve with training, and benefits should extend to other similarly-structured tasks. Changes within stimulus domain (e.g., letters, digits), the input modality of the memory items, or the modality of response should not constrain transfer. In the case of these task deviations, the routine requires a minor modification that is unlikely to have repercussions on the higher-order structure of the routine. Whether transfer will extend across different representational domains (verbal, visuo-spatial) is harder to anticipate because it is unknown whether n-back is supported entirely by domain-general processes (in which case there should be cross-domain transfer) or in part by domain-specific processes such as rehearsal (in which case transfer may be limited to same-domain tasks). Transfer is not expected between n-back tasks and complex span tasks because of the mismatch in overall task structures and hence the routines they require. A similar point was made by Shipstead et al. (2012): “learning (e.g., practice, strategies) that occurs during n-back ... training may simply not apply to complex span tasks” (p. 646). Outcomes of a recent meta-analysis of n-back training studies indicate that this is indeed the case (Soveri et al., 2017).

**Running span.** In this paradigm, sequences of memory items of unpredictable length are presented and participants attempt to recall the last n items when the end of the list occurs (Jaeggi et al., 2008). Continuous updating of the memory items to be potentially recalled is a highly unfamiliar activity that can only be performed by developing a new routine. Although there is no agreed model of running span, there is consensus regarding the high cognitive demands of the task. Postle, Berger, Goldstein, Curtis, and D'Esposito (2001) proposed that it requires not only encoding, storage and rehearsal but also the discarding of previously encoded items and repositioning.

The cognitive routine framework and other process-specific accounts make different predictions regarding the transferability of training between n-back and running span. Despite the common updating requirement, no transfer is expected because the broad set of processes required to perform each paradigm are so different. Whereas n-back requires comparisons of each successive stimulus with the most recently presented item, running span requires full serial recall of the updated stimulus set following the unpredictable end of the sequence. Any updating that takes place will therefore occur in the context of completely different preceding and succeeding processes, yielding routines with distinct higher-order structures. In contrast, according to the process-specific account of Dahlin et al. (2008), both tasks employ a common updating process that can be enhanced by training and will therefore show transfer. The data they report are consistent with this position, with transfer to an n-back task following training on running span. However, this study lacked an active control condition for comparison with running span training, potentially over-estimating the specificity of any training effects to updating in particular. To our knowledge there have been no further tests of transfer of training across these paradigms.

#### Study 1

The aim of this study was to investigate whether transfer following WM training arises when specific task features are shared both by training activities and transfer tasks and to discover the magnitude of their transfer. We performed a meta-analysis of randomized controlled trials (RCTs) of WM training that reported data on transfer to other WM tasks. The analysis was restricted to studies including untrained tasks that required the ordered recall of memory items in the context of simple serial recall or more complex WM paradigms.

The extent to which transfer is mediated by common elements of trained and untrained tasks was examined for the following task features: stimulus input modality (auditory, visual), recall modality



(spoken, manual), stimulus category (words, letters, digits for verbal stimuli, objects and spatial locations for visuo-spatial stimuli), stimulus domain (verbal, visuo-spatial), and recall paradigm (serial recall, complex span, backward span). In each, transfer was measured under two conditions: i) when the feature was present in both the untrained task and at least one of the training activities (matched), and ii) when it was present in the untrained task but not during training (unmatched). If training induces changes in any of the cognitive processes associated with specific features of the trained activity, transfer should be greater when that feature is also present in the untrained WM task.

Analyzing the impact of feature overlap on transfer for multiple task features provides a systematic means of identifying the cognitive processes which undergo change during WM training. It also provides a way of testing predictions from the cognitive routine framework and, where predictions could not be made, of providing new information to inform its theoretical refinement. The framework predicts that training-related changes will be most substantial when participants have to perform unfamiliar WM activities that require the development of novel cognitive routines because these new routines can then be applied to untrained tasks sharing the same requirements. The more unusual the cognitive demands of a WM task and the less that they can be met by existing mechanisms, the greater the potential for transfer.

Transfer was not expected to be influenced by whether the input (memory items) or output (response) modalities were matched. Consider the case of linguistic material. Words can be presented and recalled through several modalities: we can recall verbal information by either speaking it aloud or writing it down, and this information may have initially been experienced in the form of spoken language, print, objects or images. In comprehension, production and reading, we readily translate between representations in these different modalities, so it should be unnecessary to develop new routines to enable transfer across modalities.

Predictions about whether transfer should depend on whether training and transfer tasks involve stimuli belonging to the same category are harder to derive. In the case of verbal stimuli there is substantial evidence that, irrespective of the input modality, verbal material is encoded in phonological form only in verbal STM (Salamé & Baddeley, 1982). We might therefore expect that transfer across verbal tasks will be independent of whether the semantic category of stimuli in trained and untrained tasks are matching, regardless of whether they are words, nonwords, digits, or letters. On the other hand, individuals engaging in extensive practice in verbal STM tasks clearly can develop highly material-specific recoding strategies that boost memory span if conditions permit (Chase & Ericsson, 1981). The gain in digit span was achieved by recoding the list items into familiar chunks in long-term memory (running times). We consider this strategy to be an example of a cognitive routine developed to exploit the recodable properties of just one category of stimuli. Category-specific transfer of this kind will necessarily be restricted to the trained stimulus category. However, it is not entirely clear whether the current generation of WM training programs provide the same opportunities for developing such strategies as these earlier studies of digit span training. In these, training typically took place over many months, the participants were small in number and highly motivated, and training was restricted to a single task or highly similar variants. In contrast, contemporary training programs usually involve multiple training activities with differing materials and involved fewer than 20 h of training in total. The strategies developed by individuals in complex WM tasks also appear to be relatively idiosyncratic (Minear et al., 2016; Norris et al., in preparation). For these reasons, the strength of transfer when the categories of verbal memory items are matched is hard to anticipate.

For the studies included in this meta-analysis, the majority of the visuo-spatial tasks involved recall of spatial locations. It was therefore not possible to test within-domain transfer across alternative forms of visuo-spatial stimuli. We can, however, ask whether transfer is mediated by stimulus domain. In serial recall, predictions differ for simple

verbal and visuo-spatial material. Serial recall for verbal items is not expected to require a new routine because it is fully served by the established system for encoding items and order information of verbal STM. A common verbal serial recall paradigm is therefore predicted to generate minimal transfer. For spatial recall, robust transfer is expected in light of the likely absence of an established visuo-spatial STM system (Morey, 2018; Morey & Miron, 2016). It is therefore proposed that in the course of training in the serial recall of spatial locations, participants develop new cognitive routines that allow them to improve performance and diminish reliance on more general cognitive resources. The detailed nature of these routines is not known but could, for example, involve refining a spatial rehearsal strategy or developing task-specific recoding strategies.

The final task feature is the WM paradigm itself. Transfer was analyzed for verbal and visuo-spatial variants of three WM paradigms: simple serial recall and two complex WM paradigms, complex span and backward span. N-back and running span were also coded, although for these paradigms it turned out that there were insufficient data for analysis. We expect that novel routines will be developed during training for visuo-spatial but not verbal serial recall, for complex span, and for backward span. Substantial routine-mediated transfer is therefore predicted across task sharing these paradigms.

## Method

### Literature search and inclusion criteria

A flow diagram summarizing the process of selecting studies for inclusion in this study is provided in Fig. 1. In March 2018 separate comprehensive literature searches of the electronic databases Psych Info and Google Scholar were carried out by two authors (DD, JH). Studies were identified from searches of keywords and titles that contained both *working memory* and *training*. The reference lists of studies and reviews were also checked for additional potentially relevant studies. The searches were then collated and, after duplicates were removed, the abstracts of the remaining studies were independently reviewed (SG, DD). If the abstract suggested that the study may be appropriate for inclusion in the meta-analysis then the full-text article of the study was evaluated against our inclusion criteria. These were:

- (i) publication in a peer-reviewed journal;
- (ii) randomized controlled trial of an adaptive WM training program;
- (iii) restriction of training activities to only one of the following complex WM paradigms with or without additional simple span tasks: complex span, backward span, updating;
- (iv) data for a minimum of 10 participants in the adaptive training condition;
- (v) inclusion of an active control training condition that involved either non-adaptive WM training or a form of adaptive training with a low WM load;
- (vi) outcome measures provided quantitative data from which effect sizes could be calculated for individual tasks. If not provided in the publications the data were requested from the authors and included if supplied;
- (vii) assessments of untrained WM tasks both before training and within 3 months of the completion of training.

Table 1 summarizes the characteristics of the selected studies. Training groups ranged in size from 14 to 62 participants, with a mean group size of 27 (median = 24) for adaptive WM training and 26 (median = 26) for the active control group. Participants were children or adolescents in 11 studies and adults aged 18 to 60 years in the remaining nine studies. Two of these included a group of older participants (60+). Most studies involved a single group of participants completing a single adaptive WM program. In four studies different groups completed different WM training programs: updating and Cogmed RM (Ang, Lee, Cheam, Poon, & Koh (2015); complex span and



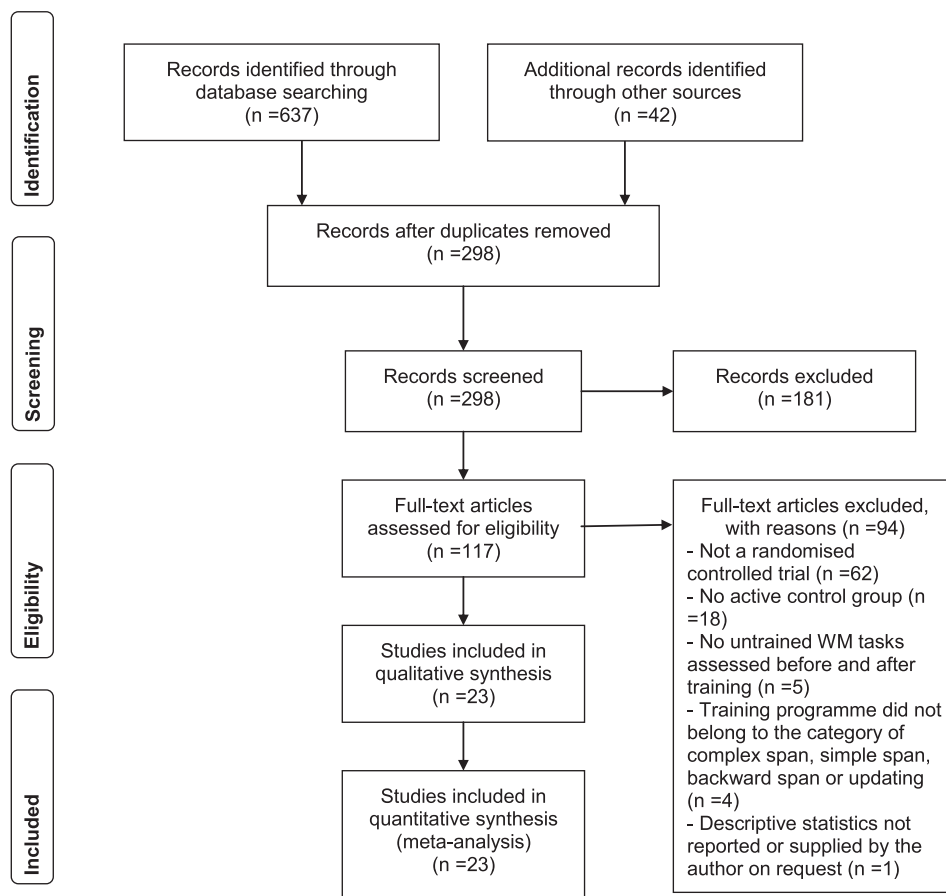


Fig. 1. PRISMA (Liberati et al., 2009) flow diagram for the selection of studies for the meta-analysis.

Table 1

Characteristics of the selected studies in Study 1.

Study	Sample	Training program (n)	Active control (n)
Ang et al. (2015)	Children with low WM, 7–8 years	Updating (32), Cogmed RM (25)	No memory load updating (28)
Bergman Nutley et al. (2011)	Children, 4 years	Cogmed JM (24)	Non-adaptive Cogmed (26)
Bigorra et al. (2016)	Children with ADHD, 7–12 years	Cogmed RM (30)	Non-adaptive Cogmed (31)
Brehmer et al. (2012)	Younger (20–30 years) and older (60–70 years) adults	Cogmed QM (54)	Non-adaptive Cogmed (45)
Chacko et al. (2013)	Children with ADHD, 7–11 years	Cogmed RM: (44)	Non-adaptive Cogmed (41)
Chooi and Thompson (2012)	Adults	N-back (15)	Non-adaptive 1-back (26)
Dentz et al. (2017)	Adults with ADHD, 18–63 years	Cogmed QM (23)	Non-adaptive Cogmed (21)
Dunning and Holmes (2014)	Adults 18–21 years	Cogmed RM (15)	Non-adaptive Cogmed (15)
Foster et al. (2017)	Adults 18–35 years, low or high span	Complex span (40), running span (39)	Adaptive visual search (39)
Gray et al. (2012)	Adolescents with severe learning difficulties and ADHD, 12–17 years	Cogmed RM (32)	Adaptive mathematics (20)
Harrison et al. (2013)	Adults	Complex span (21), simple span (17)	Adaptive visual search (17)
Henry et al. (2013)	Children, 8–10 years	Complex span (18)	Processing speed (18)
Hitchcock and Westwell (2017)	Children, 10–13 years	Cogmed RM (50)	Non-adaptive Cogmed (44)
Karbach et al. (2015)	Children, 7–9 years	Complex span (14)	Non-adaptive Cogmed (14)
Kundu et al. (2015)	Adults, 18–35 years	N-back (15)	Adaptive computer game (15)
Lawler-Savage and Goghari (2016)	Adults, 30–60 years	N-back (27)	Adaptive processing speed (30)
Metzler-Baddeley et al. (2016)	Adults, mean age 26 years	Cogmed QM (20)	Non-adaptive Cogmed (20)
Minear et al. (2016)	Adults, mean age 19 years	N-back (31), complex span (32)	Adaptive video game (26)
Passolunghi and Costa (2016)	Children, 5 years	Complex span(15)	Adaptive numeracy training (15)
Redick et al. (2013)	Adults, 18–30 years	N-back (24)	Adaptive visual search (29)
Thompson et al. (2013)	Adults, 18–45 years	Updating (20)	Adaptive object tracking (19)
Van der Molen et al. (2010)	Adolescent with intellectual disabilities	Complex span (41)	Non-adaptive version same task (26)
von Bastian et al. (2013)	Adults 18–35 years & 61–77 years	Complex span (61)	Adaptive non-memory (62)

running span (Foster et al., 2017) complex span and simple span (Harrison et al. (2013); n-back and complex span (Minear et al., 2016). The nature of the active control conditions varied across studies. In 13 studies the control group received a non-adaptive version of the WM

training program that was fixed at a level of either low memory load or no memory load. The remaining studies either employed adaptive programs that either did not tax WM at all (e.g., visual search training) or other activities assumed to place only low demands on WM (e.g.,

mathematics training, video games).

### Feature coding

Each untrained WM task was paired with a single WM task in the training program and both tasks were then coded according to five categories of feature<sup>3</sup>: stimulus type (digits, letters, words, objects, spatial locations), stimulus domain (verbal, visuo-spatial), stimulus modality (auditory, visual), response modality (spoken, manual), and paradigm (serial recall, complex span, and backward span). The use of the serial recall category was restricted to simple serial recall tasks. The procedure for matching the trained task with each untrained task within each study was as follows.

1. Match on both paradigm and stimulus domain (e.g., verbal & complex span).
2. If 1 is not possible, match on paradigm alone (e.g., complex memory, or serial recall).
3. If 2 is not possible or there are multiple trained tasks for 2, match on the trained task with the greatest total number of other matched features.
4. If two or more training activities are equivalently matched according to the above criteria, select a single representative trained task for matching.

For some tasks, it was necessary to code multiple features within a single category. For example, in dual n-back tasks each stimulus item consists of both a verbal and visuo-spatial stimulus (e.g. Kundu et al., 2013). In total, 113 pairs of trained (T) and untrained (UT) WM tasks met the task selection criteria. Full feature coding for each of these pairs with details of the trained and untrained tasks and sample sizes is reported in Gathercole, Dunning, Holmes, and Norris (in press). For each task pair, each feature was coded as either not present (empty cell), present in the trained task only (T), present in the untrained task only (UT), or present in both tasks (T&UT). In the four studies in which different groups performed different WM training programs, each untrained task was matched with the closest task from each of the different training programs. For the Ang et al. (2015) study, for example, backward letter span performance was analyzed separately for each of the following combinations of groups: adaptive/ non-adaptive Cogmed WM training, and adaptive/ non-adaptive running span training.

### Meta-analytic procedure

The following data were recorded for each transfer task in each study: the number of participants in the adaptive WM training and the active control groups, the means and SDs for the two groups pre- and post-training. All analyses were conducted using version 3.3 of the Comprehensive Meta-Analysis program (Borenstein, Hedges, Higgins, & Rothstein, 2005). Confidence intervals were calculated for Cohen's *d* effect sizes (Cohen, 1988).

Due to variation in studies (e.g. type of training program used, age of sample, outcome measures used) a random effects model was chosen for all analyses. For each analysis, outcomes with a standard residual value greater than 2 were classified as outliers and excluded (Hedges & Olkin, 1985). The following effect sizes were excluded: Brehmer et al. (2012) – stimulus type (spatial location matched), stimulus domain (verbal matched, visuo-spatial matched), response modality (spoken unmatched, all), backward span (verbal matched, all); Chooi & Thompson (2012) – complex span (verbal unmatched, all). For serial recall, complex span, and backward span untrained tasks, data were analyzed separately for the conditions in which the stimuli were both verbal, both visuo-spatial, or crossed the two domains (verbal trained to visuo-spatial untrained and *vice versa*) if there were at least two effect

sizes in each case. A further set of analyses was performed for each paradigm summed across the different domain conditions.

### Analysis plan

For each feature separate analyses were conducted for the matched conditions, the unmatched feature conditions, and the summed comparisons across both categories ('all'). Cohen's *d* effect sizes were calculated for the pre- to post-training gain for the adaptive relative to the control group (difference between the gain scores for the two groups/summed SD), with confidence intervals, z-scores and *p* values for the effect sizes. The criterion for significance was set at .05. By convention, an effect size *d* of .2 is considered small, .5 moderate and .8 large. Measures of heterogeneity (*Q*, *p*, *I*<sup>2</sup>) and publication bias (Eggers) were also calculated. For *I*<sup>2</sup> estimates, a value of 0% equates to no heterogeneity, 25% to low heterogeneity, 50% to moderate heterogeneity and 75% to high heterogeneity (Higgins & Green, 2008).

Moderator analyses tested whether feature match had a significant impact on the magnitude of effect size. Feature match (matched, unmatched) was coded as a categorical moderator variable in each regression model and its influence on effect size was assessed for the summed matched and unmatched data for each feature. The critical outcomes of the moderator analyses are the *p* value and *R*<sup>2</sup>. The criterion for significance was set at *p* = .05. In order to test further whether the effect size (transfer) for matched conditions differs for verbal and visuo-spatial material, the matched feature data were summed across the two domains for each of the serial recall and complex span paradigms. The significance of domain as a moderator variable for matched effect sizes was tested for each paradigm.

### Results

The results of the analyses are summarized in Tables 2–4 (Supplementary materials). Across all 113 task pairs in the analysis, the mean effect size (*d*) was .42, *SD* = 0.54. The analyses assessed the statistical significance of the effect sizes according to each matched and unmatched feature, and of feature match condition as a moderator of transfer. The patterns of significance for each feature are summarized in Table 5. It should be noted that for three features the number of cases is less than 10, the recommended minimum number of cases for moderator analysis (Higgins & Green, 2008). These features are indicated by parentheses in Table 5. It should be noted that in each case the number of effect sizes included in the moderator analysis was greater than 18.

### Digits

The effect size was large for matched task pairs (*d* = 0.994, *p* < .001) and smaller but significant for unmatched pairs (*d* = 0.357, *p* < .01). Match was a significant moderator of transfer (*p* = .005).

### Letters

Effect sizes were small but significant both when the stimuli were matched across the trained and untrained tasks (*d* = 0.301, *p* < .05) and when they were not (*d* = 0.337, *p* = .046). Match was not a significant moderator of transfer (*p* > .05).

### Words/nonwords

The effect size was moderate and significant for matched pairs (*ES* = 0.568, *p* < .05) and nonsignificant for unmatched pairs (*d* = 0.294, *p* > .05). Match was not a significant moderator of transfer (*p* > .05).

### Objects

The effect sizes were moderate and comparable in magnitude both when the stimuli were matched (*d* = 0.575, *p* < .05) and when they were unmatched (*d* = 0.403, *p* < .05). Match was not a significant moderator of transfer (*p* > .05).

<sup>3</sup> Feature coding was conducted independently by SG and DD/ JH, with differences resolved by discussion.

**Table 2**  
Outcomes of the meta-analysis of matched features on transfer.

Feature category	Transfer feature	Matched feature at training										
		Intervention effects								Heterogeneity		
		<i>k</i>	<i>N</i>	<i>n</i>	<i>d</i>	<i>SE</i>	95% <i>CI</i>	<i>z</i>	<i>p</i>	<i>Q</i>	<i>I</i> <sup>2</sup>	<i>p</i>
<b>Stimulus type</b>												
	Digits	7	7	421	0.994	0.194	0.615–1.374	5.132	< .001	19.397	69.068	< .001
	Letters	5	10	245	0.301	0.142	0.023–0.578	2.123	0.034	4.750	15.791	0.314
	Words/ nonwords	3	4	115	0.568	0.273	0.033–1.103	2.080	0.038	3.952	49.399	0.139
	Objects	2	2	85	0.575	0.222	0.140–1.010	2.590	0.010	0.073	0.000	0.787
	Spatial location	14	26	774	0.551	0.097	0.362–0.740	7.088	< .001	18.770	0.130	0.130
<b>Stimulus domain</b>												
	Verbal	19	48	1058	0.455	0.069	0.319–0.591	6.546	< .001	22.691	16.268	0.251
	Visuo-spatial	14	40	681	0.470	0.078	0.316–0.624	0.470	< .001	45.542	69.259	< .001
<b>Stimulus modality</b>												
	Auditory	12	21	638	0.470	0.096	0.283–0.658	4.917	< .001	14.924	26.293	0.186
	Visual	21	82	1058	0.459	0.074	0.315–0.603	6.231	< .001	26.501	24.532	0.150
<b>Response modality</b>												
	Spoken	2	5	68	0.615	0.253	0.119–1.110	2.431	0.015	0.862	0.000	0.353
	Manual	18	78	1019	0.463	0.077	0.312–0.614	6.021	< .001	26.346	27.883	0.121
<b>Serial recall</b>												
	Verbal	9	9	509	0.508	0.095	0.322–0.694	5.355	< .001	8.575	6.709	0.379
	Visuo-spatial	10	10	517	1.054	0.205	0.652–1.456	5.142	< .001	39.318	77.110	< .001
	All	11	19	611	0.709	0.111	0.490–0.927	6.357	< .001	42.525	60.024	< .001
<b>Complex span</b>												
	Verbal	6	8	325	0.544	0.175	0.201–0.888	3.110	0.002	10.580	52.743	0.060
	Visuo-spatial	3	4	108	1.010	0.272	0.476–1.544	3.709	< .001	3.471	42.378	0.176
	Cross-domain	3	4	154	0.119	0.267	- 0.405–0.643	0.445	0.656	4.989	59.911	0.083
	All	7	16	423	0.540	0.153	0.241–0.840	3.540	< .001	33.025	66.692	0.001
<b>Backward span</b>												
	Verbal	7	7	381	0.778	0.120	.542–1.013	6.470	< .001	7.282	18.722	0.287
	Cross-domain	2	2	161	1.294	0.368	0.572–2.016	3.514	< .001	4.296	76.722	0.038
	All	7	9	381	0.901	0.143	.621–1.181	6.309	< .001	18.853	57.567	0.016

*k* = number of studies; *N* = number of effect sizes; *n* = number of participants

### Spatial location

The effect size was moderate for matched task pairs ( $d = 0.551$ ,  $p < .001$ ) and nonsignificant when the trained task for unmatched pairs ( $d = 0.360$ ,  $p > .05$ ). Match was not a significant moderator of transfer ( $p > .05$ ).

### Stimulus domain

#### Verbal

The effect size was moderate in magnitude and highly significant when both tasks were verbal ( $d = 0.455$ ,  $p < .001$ ) and non-significant when the domain did not match ( $d = 0.018$ ,  $p > .05$ ). Match was a significant moderator of transfer ( $p < .004$ ).

#### Visuo-spatial

With matched spatial material the effect size was significant and moderate ( $d = 0.470$ ,  $p < .001$ ). It remained moderate in magnitude but was nonsignificant when the stimuli were not matched across tasks ( $d = 0.643$ ,  $p > .05$ ). Match was not a significant moderator of transfer ( $p > .05$ ).

### Stimulus modality

#### Auditory

The effect size was large when both tasks employed auditory presentation ( $d = 0.470$ ,  $p < .001$ ) and was nonsignificant when they did not ( $d = 0.128$ ,  $p > .05$ ). Match was not a significant moderator of transfer ( $p > .05$ ).

#### Visual

The effect size was large and significant when both tasks employed visual presentation ( $d = 0.459$ ,  $p < .001$ ). There were no data for cross-modal task pairs.

### Response modality

#### Spoken

The effect size was moderate in magnitude both when both tasks employed spoken recall and when they did not ( $d = 0.615$  and  $.401$ , respectively,  $p < .001$ ). Match was not a significant moderator of transfer ( $p > .05$ ).

#### Manual

The effect size was significant and moderate in magnitude when spoken recall at transfer was combined with manual recall at training ( $d = 0.463$ ,  $p < .001$ ). No data are available for cross-modal task pairs.

### Serial recall

#### Verbal

The effect size was moderate and significant for matched pairs ( $d = 0.508$ ,  $p < .001$ ) and nonsignificant for unmatched pairs ( $d = 0.173$ ,  $p > .05$ ). The match significantly moderated transfer ( $p < .05$ ).

#### Visuo-spatial

The effect size was very large and significant for matched pairs ( $d = 1.054$ ,  $p < .001$ ) and nonsignificant for unmatched pairs

**Table 3**  
Outcomes of the meta-analysis of unmatched features on transfer.

Feature category	Transfer feature	Unmatched feature at training										
		Intervention effects								Heterogeneity		
		<i>k</i>	<i>N</i>	<i>n</i>	<i>d</i>	<i>SE</i>	95% CI	<i>z</i>	<i>p</i>	<i>Q</i>	<i>I</i> <sup>2</sup>	<i>p</i>
Stimulus type												
	Digits	10	15	564	0.357	0.136	0.091–0.623	2.631	0.009	21.545	58.228	0.010
	Letters	3	9	233	0.337	0.169	0.006–0.668	1.998	0.046	3.019	33.752	0.221
	Words/nonwords	8	18	482	0.294	0.159	–0.017 to 0.606	1.855	0.064	19.026	63.209	0.008
	Objects	3	22	136	0.403	0.174	0.062–0.744	2.316	0.021	0.571	0.000	0.752
	Spatial location	5	10	286	0.360	0.411	–0.446 to 1.166	0.875	0.382	42.027	90.482	< .001
Stimulus domain												
	Verbal	5	14	246	0.018	0.136	–0.248 to 0.283	0.130	0.897	4.382	8.711	0.357
	Visuo-spatial	4	11	257	0.643	0.423	–0.185 to 1.472	1.522	0.128	29.480	89.824	< .001
Stimulus modality												
	Auditory	2	6	121	0.128	0.288	–0.437 to 0.693	0.443	0.658	2.394	58.236	0.658
	Visual											
Response modality												
	Spoken	12	26	672	0.401	0.103	0.199–0.602	3.899	0.000	18.192	39.533	0.077
	Manual											
Serial recall												
	Verbal	6	11	305	0.173	0.150	–0.122 to 0.468	1.151	0.250	8.219	39.167	0.145
	Visuo-spatial	6	10	279	0.138	0.121	–0.990 to 0.376	1.140	0.254	4.810	0.000	0.440
	All	10	21	490	0.158	0.077	0.008–0.308	2.064	0.039	21.254	20.014	0.215
Complex span												
	Verbal	8	11	416	0.046	0.103	–0.155 to 0.248	0.452	0.652	7.340	4.636	0.394
	Visuo-spatial	7	16	379	0.183	0.103	–0.019 to 0.386	1.776	0.076	1.467	0.000	0.962
	Cross-domain											
	All	12	27	649	0.112	0.072	–0.029 to 0.253	1.562	0.118	9.728	0.000	0.782
Backward span												
	Verbal	3	3	178	0.144	0.153	–0.315 to 0.644	0.673	0.501	5.113	60.885	0.078
	Cross-domain											
	All	4	4	239	0.354	0.258	–0.152 to 0.860	1.372	0.170	11.298	73.446	0.010

*k* = number of studies; *N* = number of effect sizes; *n* = number of participants.

( $d = 0.138$ ,  $p > .05$ ). The match was a significant moderator of transfer ( $p < .001$ ). As there was only one cross-domain matched effect size for serial recall, this condition could not be analyzed.

In order to test whether transfer differed significantly across verbal and spatial tasks, a further moderator analysis was performed on the matched feature data across all serial recall tasks with domain entered as a categorical moderator variable. Domain was a significant moderator of effect size,  $Q = 41.473$ ,  $I^2 = 61.688$ ,  $p < .001$ , demonstrating greater transfer for matched pairs for spatial than verbal serial recall.

#### All

The effect size was moderate and significant for matched pairs ( $d = 0.709$ ,  $p < .001$ ) and nonsignificant for unmatched pairs ( $d = 0.158$ ,  $p > .05$ ). Match was a significant moderator of transfer ( $p < .001$ ).

#### Complex span

##### Verbal complex span

The effect size was moderate and significant for matched pairs ( $d = 0.544$ ,  $p = .002$ ) and nonsignificant for unmatched pairs ( $d = 0.046$ ,  $p > .05$ ). Match was a significant moderator of transfer ( $p < .005$ ).

##### Visuo-spatial complex span

The effect size was moderate and significant for matched pairs ( $d = 1.010$ ,  $p < .001$ ) and nonsignificant for unmatched pairs ( $d = 0.183$ ,  $p > .05$ ). Match was a significant moderator of transfer ( $p = .003$ ).

In order to test whether transfer differed significantly across verbal

and visuo-spatial spatial tasks, a further moderator analysis was performed on the matched feature data with domain entered as a categorical moderator variable. Domain was not a significant moderator of transfer,  $Q = 17.706$ ,  $I^2 = 58.487$ ,  $p > .05$ .

#### Cross-domain

The effect size was nonsignificant ( $d = 0.119$ ,  $p > .05$ ).

#### All

The effect size was moderate for matched pairs ( $d = 0.540$ ,  $p < .001$ ) and very small for unmatched pairs ( $d = 0.112$ ,  $p > .05$ ). The match was a significant moderator of transfer ( $p < .001$ ).

#### Backward span

##### Verbal

The effect size was moderate and significant for matched pairs ( $d = 0.778$ ,  $p < .001$ ) and nonsignificant for unmatched pairs ( $d = 0.144$ ,  $p > .05$ ). Feature match was a significant moderator of transfer ( $p < .05$ ).

##### Spatial

There were insufficient data for analysis in this condition.

#### Cross-domain

The effect size was large and significant for matched pairs ( $d = 1.294$ ,  $p < .001$ ).

#### All

The effect size was moderate and significant for matched pairs



**Table 4**  
Outcomes of the meta-analysis collapsed across feature match condition on transfer and the feature moderator analysis.

Feature category	Transfer feature	All features									Feature moderator analysis			
		Intervention effects							Heterogeneity			$R^2$	$p$	
		$k$	$N$	$n$	$d$	$SE$	95% CI	$z$	$p$	$Q$	$I^2$			$p$
Stimulus type	Digits	13	22	985	0.619	0.139	0.347–0.892	4.459	< .001	67.416	76.267	< .001	0.490	0.003
	Letters	7	19	478	0.306	0.099	0.113–0.500	3.111	0.002	7.774	9.953	0.353	0.000	0.163
	Words/nonwords	11	22	597	0.361	0.134	0.098–0.623	2.660	0.007	24.006	58.344	0.008	0.090	0.362
	Objects	4	24	221	0.468	0.137	0.200–0.737	3.420	0.001	1.015	0.000	0.908	0.000	0.543
	Spatial location	16 <sup>1</sup>	36	810	0.516	0.125	0.271–0.761	4.122	< .001	61.205	70.591	< .001	0.020	0.462
Stimulus domain	Verbal	23	62	1185	0.370	0.070	0.232–0.508	5.253	< .001	36.101	33.521	0.054	0.690	0.004
	Visuo-spatial	16	51	912	0.531	0.111	0.314–0.748	4.798	< .001	44.425	61.733	< .001	0.090	0.401
Stimulus modality	Auditory	13	27	759	0.416	0.094	0.231–0.601	4.408	< .001	20.478	36.517	0.084	0.400	0.134
	Visual													
Response modality	Spoken	14	31	738	0.422	0.094	0.238–0.607	4.486	< .001	19.732	34.118	0.102	0.080	0.482
Serial recall	Verbal	14	20	814	0.371	0.093	0.189–0.554	3.985	< .001	22.742	38.439	< .001	0.830	0.044
	Visuo-spatial	14	20	826	0.732	0.18	0.380–1.084	4.076	< .001	82.633	81.847	< .001	0.530	< .001
	All	16	40	910	0.445	0.081	0.286–0.604	5.471	< .001	95.801	63.466	< .001	0.560	< .001
Complex span	Verbal	12	19	407	0.263	0.117	.034–0.492	2.253	0.024	29.541	55.994	0.005	0.870	0.002
	Visuo-spatial	9	20	447	0.390	0.132	0.130–0.650	2.942	0.003	17.892	49.698	0.036	1.000	< .001
	All	15	43	859	0.320	0.083	0.157–0.483	3.854	< .001	60.968	55.715	< .001	0.290	0.021
Backward span	Verbal	10	10	559	0.579	0.146	0.294–0.865	3.973	< .001	24.252	62.889	0.004	0.800	0.025
	All	11	13	620	0.724	0.146	0.438–1.010	4.961	< .001	44.09	72.783	< .001	0.41	0.026

$k$  = number of studies;  $N$  = number of effect sizes;  $n$  = number of participants.

( $d = 0.901$ ,  $p < .001$ ) and nonsignificant for unmatched pairs (see verbal backward span above). Feature match was a significant moderator of transfer ( $p < .05$ ).

One limitation of the coding method is that although individual features are coded and analyzed independently of one another, matched features are often highly or indeed perfectly correlated with one another. For example, a pair of tasks involving spatial stimuli will necessarily also be coded as visuo-spatial. As a consequence, it is not always possible to identify which matched task features are critical to transfer. We therefore examined whether any of the matched features found to be significantly associated with substantial transfer in the moderator analysis could be explained in terms of another correlated feature that could potentially be the origin of transfer.

There were two such features. One is digits. Matched digits yielded a large transfer effect ( $d = 0.990$ ). On closer examination it was found that all seven of the task pairs with matched digits employed a backward span paradigm which was also associated with a high level of transfer effect ( $d = 0.901$ ). A second confounded feature is verbal material. For this feature the transfer effect was smaller, but still highly significant ( $d = 0.455$ ). This feature too was associated with recall paradigm: of the 50 matched verbal pairs, 27 also had shared the same paradigm, whereas none of the 14 unmatched verbal untrained tasks did. Thus both for digits and for verbal material more generally, it is possible that the significant levels of transfer observed were consequences of matched paradigms rather than the stimulus items.

## Discussion

The meta-analysis evaluated the features associated with transfer within WM in RCTs of adaptive WM training with active control training conditions. Across the 24 studies furnishing data for 113 pairs of trained and WM untrained tasks, the strength of transfer was small to moderate ( $d = 0.42$ ). The magnitude of transfer was associated with

some matched features but not with others. Transfer was high when the tasks employed the same paradigm - either serial recall, complex span, or backward span paradigms. Paradigms differed in the impact of domain on transfer. For complex span, transfer was substantial when both the untrained and trained tasks employed material in the same domain (either verbal or visuo-spatial), but was absent when the domains differed across the task pairs. For serial recall, transfer was very large for spatial material, and reduced, although still significant, for verbal material. In the case of backward span, transfer was large both for verbal tasks and for task pairs that crossed domains. In contrast, the following more specific features did not generate transfer: stimulus category (letters, words/ nonwords, objects, locations), the stimulus domain (verbal, visuo-spatial), input modality (auditory, visual), and output modality (manual, spoken). Transfer was observed when the memory items were either digits or verbal material more generally, although this could reflect the confounding influences of shared paradigms.

The findings provide broad support for the predictions of the cognitive routine framework. By this account, substantial transfer following WM training occurs only when both the trained and untrained activities impose the same unfamiliar task demands that are not supported by existing WM sub-systems. In complex span, distractor activities are interpolated between the presentation of items required for subsequent serial recall. This presents a major challenge for the maintenance and retrieval of stimulus presentations in the face of near-continuous distraction that must be addressed by engaging additional cognitive processes either to prevent decay (Barrouillet et al., 2009) or to minimize interference caused by encoding distractor items (Oberauer et al., 2012). It is this novel schedule that we suggest constitutes a cognitive routine, and its construction and refinement represents the process of acquiring a new cognitive skill.

The present findings tell us two important things about the structure and generalizability of a complex span routine. First, high levels of

**Table 5**  
Summary of significance of effect sizes (transfer) as a function of feature match.

Feature category	Transfer feature	Feature match		Feature as a moderator
		Matched	Unmatched	
Stimulus type				
	Digits	***	**	**
	Letters	*	*	(-)
	Words/nonwords	*	-	-
	Objects	*	*	(-)
	Spatial location	***	-	-
Stimulus domain				
	Verbal	***	-	**
	Visuo-spatial	***	-	-
Stimulus modality				
	Auditory	***	-	-
	Visual	***	-	-
Response modality				
	Spoken	*	***	-
	Manual	***	-	-
Serial recall				
	Verbal	***	-	*
	Visuo-spatial	***	-	***
	All	***	*	***
Complex span				
	Verbal	**	-	**
	Visuo-spatial	***	-	(**)
	Cross-domain	-	-	-
	All	***	-	*
Backward span				
	Verbal	***	-	*
	Cross-domain	***	-	-
	All	***	-	*

Cases < 10 for symbols shown in parentheses (see text for details).

- $p > .05$ .

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

transfer between complex span tasks indicate that the routine can be readily adapted to accommodate novel interpolated activities. In every case included in the meta-analysis, these activities differed between the trained and untrained tasks. Combinations included processing the meaning of sentences and mental arithmetic (Harrison et al., 2013), sentence verification and counting (Henry, Messer, & Nash, 2014), symmetry and orientation judgments (Harrison et al., 2013), lexical decision and arithmetic calculations (Minear et al., 2016), and vowel/consonant and odd/even judgments (von Bastian & Oberauer, 2013a). The processes required to accomplish these activities are highly specific and have relatively little in common. Transfer across these distractor activities indicates that if the high-level structure of alternating stimulus presentation and distractor activity is preserved, distractor sub-routines can be substituted one for another with relative ease.

Second, the representational domain of the memory items limits the generalizability of the routine. There was no transfer for complex span task pairs that crossed stimulus domain ( $d = 0.12$ ) although it was substantial when the task pairs shared the stimulus domain ( $d = 0.66$ ). This suggests that domain-specific encoding and/or maintenance processes cannot readily be adapted to fit a complex span task drawn from another domain. This contrasts with the subroutines for handling distractor activities, which do show a high degree of content generality. The implication is that the domain influences more than just the local elements (subroutines) of the routine, preventing modular substitutions of one for another. This may not be too surprising given the different nature of stimulus maintenance processes for verbal and visuo-spatial

material. Rehearsal depends on the covert control of the relevant domain-specific action production systems - of articulation in the case of verbal rehearsal (Caplan, Rochon, & Waters, 1992; Waters et al., 1992) and of eye movements for spatial rehearsal (Pearson et al., 2014; Pearson & Sahraie, 2003). A consequence of the distinctiveness of these underpinning systems may be that the two forms of rehearsal cannot be readily substituted for one another. Rehearsal may also be interleaved with the periods of distractor activity throughout the trial (Barrouillet et al., 2009), influencing multiple points during the execution of the routine. In this way, informational domain may directly shape the broad structure of the routine, limiting its transfer to complex span tasks embedded in the same domain.

Backward span is another complex WM task that is not readily served by existing mechanisms in STM, as a consequence the forward-going representation of serial order in STM (Hurlstone et al., 2014). One of the ways in which the task can be accomplished is by peeling off the last item across successive forward retrievals of diminishing length (Conrad, 1965; Anders & Lillyquist, 1971; Norris et al., in preparation). Such a strategy would require the establishment of a new routine to coordinate the multiple processes required to achieve the end goal. The present findings demonstrate high levels of transfer across backward span tasks, both for verbal stimuli and for cross-domain task pairs. It is tempting to conclude that this demonstrates the application of a backward recall routine to a novel task on this basis. However, in the present dataset the degree of generalization captured by the task pairs in the verbal condition is very limited. In the six of the seven task pairs, the trained and untrained tasks were identical (with digits as memory items), differing only in the modality of response (manual responding during training, spoken recall at transfer). The large degree of cross-modal transfer in backward span suggests that higher-order features of the routine can be adapted to fit new stimuli, although here the data are restricted to just two task pairs. The extent to which the high-order structure of the backward span routine can be generalized across materials has therefore yet to be fully established.

Although tests of visuo-spatial serial recall such as Corsi blocks and spatial span are considered to tap STM, the conclusion of the task analysis was that there is not a highly developed STM system for storing visuo-spatial material (Alloway et al., 2006; Kane et al., 2004; Pearson et al., 2014; Thompson et al., 2006). We therefore speculate that routines may be developed and refined for tasks involving the recall of spatial locations. The present findings of substantial transfer across visuo-spatial serial recall tasks are consistent with this position. As the great majority of matched task pairs employed spatial locations as the memory items, the potential generalizability of this routine to other visuo-spatial characteristics is unknown.

The cognitive routine framework contrasts with other process-specific accounts of WM training in generating predictions not only about the features that generate transfer but also those that will not. One prediction is that verbal serial recall tasks will not require a new routine because they are already supported directly by specialized system of verbal STM. Transfer was indeed significantly lower for verbal than visuo-spatial recall ( $d = 0.51$  vs  $.84$ ). However, it extended across different stimuli (letters, digits, words), suggesting that it does not originate from material-specific mnemonic strategies. So, where does it come from? Not, we suggest, from new routines. The training benefits may instead reflect subtle adjustments and re-calibration of the order-based encoding system already in place in verbal STM. These could have two possible sources. One is the optimization of the routines to task-specific characteristics likely to generalize to untrained tasks in the same study. An example is the temporal properties of the task such as the stimulus presentation rate and time allowed for recall. Alternatively, the extended practice may lead to subtle fine-tuning of the core mechanisms themselves.

Minimal levels of transfer were also predicted for local features of tasks such as the stimulus presentation modality and the response format as these are likely to be handled by specialized and highly

modularized systems for the processing of inputs and outputs. The present findings provide strong support for this position: neither a common input nor output modality had any impact on transfer. It was predicted that within the verbal domain at least, the specific category of stimuli (letters, digits, etc.) will not modulate transfer because information in verbal STM is represented in a phonological rather than semantic form (Salamé & Baddeley, 1982). There was partial support for this prediction: for letters and words/nonwords a category match across tasks did not generate transfer. For digits, transfer was high. This could be because participants developed digit-specific recoding strategies during training, as reported in previous studies of lengthy intensive training in digit span (Ericsson & Simon, 1981). However, in each of the task pairs included in this analysis, both tasks also employed a backward span paradigm and this is an alternative and possibly more plausible source of routine-mediated transfer. Further data are needed to resolve this issue.

## Study 2

The broad set of training paradigms included in the meta-analysis provide a more comprehensive analysis of the boundary conditions to transfer within WM than any single study could provide. However, some of the strengths of the approach are also limitations. One of these is the heterogeneity of the studies included in the meta-analysis: there was variation in the training activities, training regimens, the transfer tests, the active control conditions, the participants, and sample sizes. While this forms a sound basis for the generalization needed for broad theoretical analysis, it cannot provide the more detailed level of analysis needed to test the cognitive routine framework as an explanation for the patterns of transfer observed in individual studies.

A second limitation is the relatively coarse level of analysis yielded by the independent coding and analysis of individual features of the pairs of trained and untrained tasks selected to be most closely matched. We have already seen that this can generate potentially misleading results: in all cases in which both tasks in the meta-analysis employed digits as the memory items, they also happened to be backward span tasks, but the two features were analyzed independently. Whether the robust levels of transfer observed in these data is mediated by the common stimulus category, the complex WM paradigm, or the combined impact of both is therefore indeterminate. Other kinds of interactions between matched and unmatched features that influence transfer could also go undetected in this method. For example, we cannot using this method determine whether cross-paradigm transfer occurs only under conditions of matched stimulus features such as domain.

One way of addressing these limitations is to examine whether the cognitive routine framework can successfully account for the transfer found with a single WM training program. This allows a more fine-grained analysis of the impact of combinations of features on transfer. We adopted this complementary approach in the present training study. In this, we re-analyzed a dataset from two studies of children ( $n = 106$ ) who had completed either a single adaptive or non-adaptive WM training program and four WM transfer tasks (Dunning et al., 2013; Holmes et al., 2009). The children were selected on the basis of low scores on complex WM tasks. These studies were not included in the meta-analysis and therefore provide an independent evaluation of the cognitive routine framework. Participants in the Holmes et al. study were excluded from Study 1 because they were not randomly allocated to the adaptive and non-adaptive training conditions; instead, they were recruited sequentially (adaptive first, then non-adaptive) using identical recruitment criteria. Dunning et al. reported analyses only for composite scores that combined pairs of transfer tests (e.g., digit and word span for verbal STM). We chose not to incorporate previously unreported descriptive statistics for these measures in Study 1, retaining the data instead for the present independent test of transfer patterns.

Participants completed either the standard adaptive or non-adaptive form of Cogmed RM Working Memory Training (<http://www.cogmed.com>). It provides training on eight WM activities drawn from a larger set of 12 activities on each daily session for at least 20 daily sessions. Further information on Cogmed RM training activities is provided in the Appendix. The following transfer tests were completed before and after training. The WM measures were taken from the Automated Working Memory Assessment (Alloway, 2007) and consisted of tests of verbal and visuo-spatial serial recall (digit span and dot matrix), backward digit span and Mr. X, a test of visuo-spatial complex span. The Wechsler Abbreviated Scales of Intelligence (WASI; Wechsler, 1999) was also administered. This consists of four subtests: Similarities and Vocabulary (from which a verbal IQ composite is formed), and Matrix Reasoning and Block Design (performance IQ). Analyses were based on standard scores for each measure.

The cognitive routine framework was used to generate predictions based on analysis of the profile of matched and unmatched features in this study, guided by the coding protocol employed in the meta-analysis. Consider first dot matrix, a test of serial recall of successively highlighted spatial locations in cells in a  $4 \times 4$  grid. This task differs only in superficial aspects of the spatial layout multiple from spatial STM activities trained daily in Cogmed RM including Data Link, in which spatial elements such as colored panels illuminate in sequence. The coded features for dot matrix and Data Link were identical. In the meta-analysis, combinations of spatial recall tasks such as these generated very high levels of transfer ( $d = 1.12$ ). We propose that this reflects the development and refinement of a routine for the serial recall of spatial material. A high degree of transfer was therefore expected for the dot matrix task.

The digit span transfer task involved the spoken serial recall tasks of spoken items. Its closest match, the Decoder task, was completed on each of the first five days of training. Sequences of letters were presented auditorily with recall by mouse selection of each item from a choice of three items at each position. The two tasks therefore differed on two features: stimulus category (digits vs letters) and response modality (spoken vs mouse click). These differences would not be expected to have any impact on ease of the phonological encoding of items and their order in verbal STM, representing input and output conditions that are widely used in studies of verbal serial recall. Findings from the meta-analysis reported in Study 1 indicate transfer across verbal serial recall tasks that although moderate in magnitude and highly significant ( $d = 0.51$ ), is markedly smaller than that for visuo-spatial serial recall. We have argued that this transfer is not mediated by the development of a new routine as verbal STM is already well-established but instead by fine-tuning of STM processes within this system, possibly to optimize task fit. In the present study, a small to moderate transfer effect is therefore predicted for digit span.

The untrained backward digit span task involved auditory presentation and spoken recall. Two daily Cogmed RM tasks also require backward recall of digits. In Input Module with Lid the response pad is not displayed until recall, and in Input Module without Lid it was present during presentation. The trained and untrained tasks differed only in response modality (spoken vs mouse selection). Transfer levels were found to be high for verbal backward span tasks ( $d = 0.78$ ) in the meta-analysis. The framework predicts high levels of transfer across backward span tasks because the unusual nature of the task demands will require the development of a new routine. We note that because in this study the trained and untrained tasks are distinguished only by the mode of recall, transfer across tasks would represent very near transfer only. In this respect, as in the majority of previous studies, this does not provide a particularly strong test of routine-mediated transfer across backward span tasks.

Mr. X is a complex span task in which participants judge whether pairs of figures are holding a ball in the same hand as one another for each of a series of displays, and then remember the location of the ball held by the figure on the right which can appear at one of six compass

positions. At the end of a trial, participants are required to recall the successive locations of the ball. There was no visuo-spatial complex span task in the Cogmed RM training program, and its closest match was a visuo-spatial serial recall activity, Visual Data Link. This shares the requirements for the serial recall of spatial locations in Mr. X, but not the interpolated distractor activity characteristic of complex span tasks. The cognitive routine framework does not generate any strong predictions for this transfer task because the trained and untrained paradigms do not squarely match. However, it is at least plausible that any improvements in encoding, maintaining and recalling spatial locations resulting from extensive training will enhance the same stimulus encoding and maintenance activities in subroutines embedded within the more complex structure of a complex span routine. We would though expect transfer to be diminished to Mr. X relative to the dot matrix, for which there is a precise match in the entire paradigm.

Finally, measures of verbal and performance IQ provided tests of the specificity of transfer. Individual differences studies have established close links between WM and problem-solving skills (Engle et al., 1999; Kyllonen & Christal, 1990; Wiley, Jarosz, Cushen, & Colflesh, 2011). However, the cognitive routines developed to accomplish complex WM tasks would be expected to have very little overlap with the specific structures problem-solving tasks using unfamiliar spatial designs used to index performance IQ. These typically involve simultaneous presentation of multiple response alternatives as well as the application of rules to transform the spatial form of one stimulus for another (Cattell, 1963; Raven, 2003). The same applies to the vocabulary-based assessment of verbal IQ, which relies on access to stored knowledge based on prior learning with little or no reliance on either the basic processes of WM or the more specific routines that develop them further to address more specific needs. Transfer of WM training benefits to performance on tests of either verbal or nonverbal IQ is therefore not expected.

## Method

The participants were children aged 7 to 12 years with low WM scores identified through routine classroom-wide screening on two WM tests. In the adaptive training group there were 32 boys and 24 girls ranging in age from 7 years 1 month to 12 years 0 months ( $M = 9$  years 1 month,  $SD = 14.42$  months), and in the non-adaptive group there were 29 boys and 21 girls with ages ranging from 7 years 6 months to 11 years 6 months ( $M = 8$  years 10 month2,  $SD = 9.82$  months). The children participated in either standard Cogmed RM training ( $n = 56$ ) or the non-adaptive version of the same program in which the difficulty level was fixed at a span length of two ( $n = 50$ ). Sixty-four participants (34 adaptive, 30 non-adaptive) who participated in the Dunning et al. (2013) study were selected on the basis of standard scores below 86 on the backward digit span and Mr. X tasks from the AWMA (Alloway, 2007). A further 42 children from the Holmes et al. (2009) study (22 adaptive, 20 non-adaptive) had standard scores below 86 for the listening recall and backward digit span tasks from the AWMA.

Written parental consent for participation was provided for all children. Further details of participant recruitment and methods are supplied in the original publications. A post hoc power analysis with a total  $N$  of 106 yielded power of .99 to detect a large effect size,  $f^2 = 0.35$ , with linear regression at  $p = .05$ . The power to detect a medium effect size of  $f^2 = 0.15$  was .95, and .23 for a small effect size,  $f^2 = 0.02$ .

## Analysis plan

The effect size Cohen's  $d$  was computed as the difference between the gain scores (pre- to post-training) for the two groups divided by their pooled  $SD$  (Weisz & Hawley, 2001). Univariate analyses were performed on baseline scores to identify any group differences prior to training. For baseline measures without a significant group effect, a general linear model (GLM) was run with post-training scores as the

dependent variable and age and baseline scores as independent variables. For baseline measures with significant group differences, both centered baseline scores and centered baseline score  $\times$  group product terms were also included in the regression models. Product terms were derived by the product of the centered scores (individual score minus the group mean) and the grouping variable. Where product terms were not significantly associated with post-training scores, GLMs were re-run omitting the product term or centered scores in the final analyses (i.e. with post-training scores as the dependent variable and group and pre-training scores as independent variables). In all cases, the critical term of interest was the group effect in the GLM and the criterion for significance was set at  $p = .05$ . Separate regression models were tested with each of the six transfer tests. Corresponding analyses were also performed on verbal and performance IQ scores in order to test the specificity of any transfer effects to WM.

In addition to traditional null hypothesis significance testing (NHST), Bayesian analysis was used to test support for the null hypothesis relative to the alternative hypothesis that training has a genuine effect. Bayesian linear regressions were conducted using JASP (Love et al., 2015). Regression models with baseline scores and intervention condition as independent variables were computed individually for each post-test dependent variable. Bayes factors are reported. By convention (Jeffreys, 1961), values are interpreted as follows: 1–3 (anecdotal evidence for the alternative hypothesis, in this case that there is an effect of training), 3–10 (substantial evidence), 10–30 (strong evidence), 30–100 (very strong evidence), 100+ (decisive evidence). The corresponding values in support of the null hypothesis are the inverse values: 0.33–1.0 (anecdotal evidence for the null hypothesis), 0.10–.33 (substantial evidence), and so on.

## Results

Descriptive statistics and statistical outcomes are shown in Table 6. Bayesian analysis provided decisive evidence in favor of the alternative hypothesis for dot matrix and backward digits. The effect sizes for transfer were 0.92 and 0.77, respectively. For Mr. X there was substantial support for the alternative hypothesis and a transfer effect size of .58. For digit span, the transfer effect size was 0.29 and the evidence did not favor either the null or alternative hypothesis (effect size of 0.29). For both verbal and performance IQ, the evidence provided substantial evidence in support of the null hypothesis of no transfer.

The same pattern of outcomes was reflected in the GLM analyses. Significant group differences were found at baseline on the dot matrix and Mr. X tests. For these variables the group product term (centered baseline score  $\times$  group) was entered along with centered baseline scores and group. In the GLMs for dot matrix and Mr. X, the product terms were nonsignificant, indicating that baseline differences did not affect the outcomes. The GLMs were subsequently run without the product terms and centered baseline scores. Highly significant group differences arising from increased performance following adaptive training were found for backward digit recall, dot matrix, and Mr. X ( $p < .001$  in each case). For digit recall, the training group effect was significant ( $p = .035$ ). It should be noted that this term reflects in part a reduction in post-training scores for the non-adaptive comparison group (.09), with a mean increase in standard scores in the adaptive group of only 2.6 points. The effect of training condition on verbal IQ was nonsignificant ( $p > .05$ ) for verbal IQ but significant for performance IQ ( $p < .05$ ). This arose from a greater post-training improvement for the non-adaptive than the adaptive group and so does not reflect a positive transfer effect.

## Discussion

The transfer patterns observed in this re-analysis of transfer data from two published studies are consistent with predictions derived by applying the cognitive routine framework to the profile of overlapping



**Table 6**  
Descriptive statistics and statistical outcomes for Study 2.

Measure	Time	Adaptive		Non-adaptive		Effect size <i>d</i>	Group effects	
		Mean	<i>SD</i>	Mean	<i>SD</i>		<i>GLM p</i>	<i>BF</i> <sub>10</sub>
Digit span	pre	92.21	15.07	87.16	14.62	0.30	0.035	1.050
	post	94.88	16.92	86.22	14.82			
Backward digit span	pre	76.11	8.59	78.58	7.62	0.77	< .001	186.170
	post	97.96	14.13	88.34	11.05			
Dot matrix	pre	84.45	12.19	90.84	14.01	0.92	< .001	1283.100
	post	106.88	19.92	96.00	13.14			
Mr. X	pre	80.38	7.68	84.02	9.31	0.58	< .001	3.870
	post	97.09	15.35	92.00	13.51			
Verbal IQ	pre	90.00	10.83	93.61	15.64	0.31	0.249	0.173
	post	90.95	12.93	91.59	13.75			
Performance IQ	pre	90.70	13.35	88.78	10.18	−0.46	0.029	0.481
	post	91.29	13.17	92.91	11.20			

and distinct features of the training program and the individual transfer tasks. They also largely uphold the conclusions regarding the features governing transfer from the meta-analysis.

The strength of transfer varied considerably across the untrained WM tasks. Highest levels were found for the two untrained tasks with WM paradigms that were also employed during training and which are hypothesized to require the development of new routines that can be transferred to similarly-structured tasks. These tasks involve the recall of spatial locations (dot matrix) and the recall of digits in reverse sequence. The same paradigms also generated large transfer effects in the meta-analysis.

An intermediate degree of transfer was found for a visuo-spatial complex span task (Mr. X), even though there was no corresponding paradigm in the Cogmed training program. We propose that this is driven by routine-mediated improvements in the recall sequences of spatial locations of multiple daily training activities. This finding is important as it extends the range of the critical conditions beyond the scope of the meta-analysis, which was not designed to quantify cross-paradigm transfer. Transfer was very weak for the untrained verbal serial recall task, even though participants had trained on another serial recall task employing different verbal stimuli (letters rather than digits) during training. This outcome is entirely consistent with the proposal that cognitive routines are only necessary and only generate transfer when they cannot be readily accomplished with existing WM processes and mechanisms. The verbal STM system amply meets the needs of verbal serial recall tasks and will therefore not lead to transfer. In the meta-analysis, tasks that shared this paradigm showed levels of transfer that were moderate but significantly smaller than for visuo-spatial STM tasks. Finally, we found no evidence that WM training influenced performance on either vocabulary or nonverbal reasoning tests. This too we expected: although WM may play a role in nonverbal reasoning, the paradigm-specific routines developed in the course of WM training will have minimal overlap with the cognitive demands of such assessments.

One limitation of the study is that the participants were all children who performed poorly at screening on tests on complex WM. The reason for this is that the motivation of the original studies was to investigate whether their relatively common WM problems (they represented the 16% of children with the lowest WM scores in the school population) could be ameliorated through training. The extent to which these results generalize to other populations of individuals with either typical or atypical WM abilities or of different ages therefore cannot be determined. What we can conclude is that in accordance with the predictions of the cognitive routine framework, training-induced transfer is greatest when the trained and untrained tasks tap aspects of WM other than verbal STM in children with weak WM skills. Moreover, the patterns of transfer observed in this particular population provide a close fit to the broader analysis of the factors driving transfer from the

meta-analysis in Study 1 of training studies employing many different populations.

### Study 3

One issue we have not yet addressed is how new routines might be created. These are only needed when existing processes and mechanisms are not sufficient to satisfy the demands of the task. The process of constructing a routine must therefore draw in part at least on cognitive resources that fall outside of WM. This position stands in opposition to more general plasticity-based accounts of training, according to which training-induced changes reflect increases in the fundamental capacity of the underlying WM system itself.

One way of conceptualizing the establishment of new routines is as a form of problem-solving behavior involving the decomposition of a complex task into its constituent cognitive parts. When assembled, these parts enable the individual to meet task goals. This might, for example, be to protect memory representations from interference or decay caused by distraction in complex span tasks, or to reverse the input sequence at output in backward span. This kind of problem-solving capacity has been most widely investigated in the context of nonverbal reasoning where it has been widely considered to be supported by limited general cognitive resources labelled either *g* or fluid intelligence (Cattell, 1963; Duncan, Emslie, Williams, Johnson, & Freer, 1996). It has been linked to the fronto-parietal brain networks that respond flexibly to multiple kinds of cognitively challenging activities (Duncan, Burgess, & Emslie, 1995; Duncan & Owen, 2000). It has been proposed that the same flexible resources may be responsible for the cognitive segmentation of complex reasoning tasks into their component parts (Duncan, 2013; Duncan, Chylinski, Mitchell, & Bhandari, 2017). Although the concept of a cognitive routine that we use is broadly influenced by the production systems approach to the acquisition of complex skills (Anderson, 1982; Taatgen, 2013), it also has much in common with Duncan's notion of cognitive segmentation. Perhaps, then, cognitive routines depend on the same general attentional resources believed to be critical for the restructuring process in the context of nonverbal reasoning tasks.

To date, investigations of individual differences in transfer following WM training have focused primarily on links with WM performance prior to training. There is some evidence from n-back training that trainees who begin training with relatively high WM scores fare better on the training tasks and, to some extent, on measures of transfer (Au et al., 2015; Jaeggi et al., 2008). In a study of complex span training, participants with higher WM showed greater performance benefits on the trained tasks as the number of sessions increased, but there was little evidence that the magnitude of transfer to other WM tasks varied as a function of pre-training WM ability (Foster et al.,

**Table 7**  
Descriptive statistics and statistical outcomes for Study 3.

Measure	Time	Mean	SD	Effect size	F	p	BF <sub>10</sub>
Digit span	pre	93.70	15.12	0.37	18.60	< .001	473.943
	post	99.70	17.60				
Backward digit span	pre	85.72	15.14	0.96	74.34	< .001	3.60E + 11
	post	100.25	15.04				
Dot matrix	pre	88.74	14.29	1.27	131.67	< .001	2.81E + 18
	post	110.46	20.01				
Mr. X	pre	87.14	15.00	0.75	51.31	< .001	9.76E + 07
	post	99.21	17.20				
Verbal IQ	pre	92.06	11.89	0.14	4.218	0.042	1.022
	post	94.02	15.4				
Performance IQ	pre	91.10	13.32	0.14	5.052	0.027	1.526
	post	92.91	12.451				

2017). The authors speculated that these findings may reflect greater ability to generate and implement new strategies during training rather than a fundamental increase in WM capacity.

Borella, Carbone, Pastore, De Beni, and Carretti (2017) conducted a detailed analysis of individual differences in baseline measures of vocabulary, age and WM the magnitude of transfer across STM and WM tasks following training on a verbal complex span task. Participants monitored lists of words in a sequence of lists for the presence of animal words and recalled the sentence-final words. In tasks judged to require active processing (which included dot matrix and backward digit span), higher-performing participants benefitted more from training. The converse pattern of greater gains for individuals with lower initial levels of performance was found for digit span. These findings show the same differentiation observed in the previous WM training study between digit span on the one hand (yielding little or no transfer) and dot matrix and backward span on the other (with very substantial transfer). In this study, the baseline predictors of post-training scores did not include fluid intelligence, the issue of particular interest of the present study.

In Study 3 we examined whether the magnitude of transfer following adaptive WM training is modulated by the fluid cognitive abilities indexed by nonverbal reasoning. If the development of new routines during training and hence their transferability depends on this ability, the strength of routine-mediated transfer should be more strongly predicted by pre-training measures of fluid intelligence than by the WM measures themselves. This association should be present in the three tasks showing substantial transfer following Cogmed WM training in the previous WM training study: dot matrix, backward span and Mr. X. No link is predicted for post-training digit span, for which transfer was insubstantial.

To achieve the statistical power and heterogeneity required to examine individual differences in transfer following training, we re-analyzed data from four WM training studies in which measures of IQ as well as WM were obtained before and after training. The sample of 108 children were a mixed group of individuals who were either typically developing, had low WM but no other recognized developmental impairments, or had a diagnosis of ADHD. They completed the four tests of WM from the AWMA (digit span, backward digit span, the dot matrix test of visuo-spatial STM, and Mr. X, a visuo-spatial test of interpolated span) and the verbal and performance IQ subtests of the WASI (WASI, Wechsler, 1999).

#### Method

The sample was composed of 34 children with low WM from Dunning et al. (2013), 22 children with low WM from Holmes et al. (2009), 25 children with a diagnosis of ADHD from Holmes et al. (2014), and 27 children (12 with low language abilities and 15 children

matched for nonverbal abilities) from Holmes et al. (2015). The participants from the Dunning et al. and Holmes et al. (2009) studies comprised the adaptive training group in WM Training Study 1. All children completed standard adaptive Cogmed RM training. Further information regarding recruitment and methods are provided in the individual publications. Written parental consent for participation was provided for all children. The mean age of the children was 9 years 3 months ranging from 8 to 11 years, and there were 68 boys and 40 girls.

#### Analysis plan

The effect size  $d$  calculated for this study is the mean training gain (average of difference between pre- and post- training scores) divided by the pooled  $SD$  (average of pre and post  $SD$ s). It should be noted that because all of the participants completed adaptive training, the effect sizes are not directly comparable with those reported in Study 1 and 2 which compared the gain scores for the adaptive and control training conditions. Univariate ANOVAs were conducted on scores on each of the four WM transfer tests as a function of time (pre-training and post-training). Corresponding Bayesian ANOVAs were also performed. In order to examine individual differences in transfer, separate GLMs and Bayesian regression analyses were performed for each set of post-training WM scores. The criterion for significance was set at  $p = .05$ . In these, the post-training measure was the dependent variable and baseline WM scores, verbal and performance IQ scores were the independent variables. Outcomes of these analyses establish the strength of unique associations between each independent and dependent variable.

#### Results

Descriptive statistics for the six tests administered before and after training and the outcome of analyses of these data are shown in Table 7. Transfer increased significantly after training on all four WM tests ( $BF_{10} > 100$ ,  $p < .001$ , in each case). The effect sizes were large for both dot matrix (1.27) and backward digit span (0.96), moderate for Mr. X (0.75) and small for digit span (0.37). For verbal and performance IQ, they were very small (0.14 in both cases).

Table 8 summarizes the outcomes of the individual differences analyses examining associations between baseline (pre-training) measures and post-training scores on the WM tests. Pre-training scores significantly predicted post-training scores for all four WM tests. In the Bayesian analyses the evidence in favor of a positive effect was decisive for digit span, strong for Mr. X and substantial for both backward span and dot matrix. NHST outcomes were consistent with this pattern, with  $p < .001$  for digit span and  $p < .05$  for all other tasks. The corresponding beta weights were between 2.2 and 2.9 times greater for digit span than each of the other measures.

Pre-training performance IQ significantly predicted dot matrix and

**Table 8**  
Outcomes of regression analyses of post-training WM scores in Study 3.

Dependent variable	Independent variable at pre-training	Regression statistics			
		Stand. $\beta$	$t$	$p$	BF <sub>10</sub>
Digit span	Digit span	0.572	6.48	< .001	3.368e <sup>6</sup>
	Verbal IQ	0.046	0.481	0.631	0.200
	Performance IQ	0.088	1.028	0.306	0.298
Backward digit span	Backward digit span	0.231	2.551	0.012	4.663
	Verbal IQ	0.197	1.983	0.050	1.489
	Performance IQ	0.209	2.141	0.035	1.996
Dot matrix	Dot matrix	0.196	2.051	0.043	1.574
	Verbal IQ	0.168	1.692	0.094	0.852
	Performance IQ	0.291	3.004	0.003	13.256
Mr. X	Mr. X	0.26	2.875	0.005	9.255
	Verbal IQ	0.15	1.575	0.118	0.681
	Performance IQ	0.295	3.138	0.002	18.381

Mr. X scores ( $p < .005$  in both cases) and, less strongly, backward digit span ( $p < .05$ ). It was not significantly associated with digit span ( $p > .05$ ). Bayesian analysis provided substantial evidence for the null hypothesis for digit span, anecdotal evidence for the alternative hypothesis for backward span and strong support for the alternative hypothesis for both dot matrix and Mr. X. Baseline verbal IQ scores were not significantly associated with any of the four WM measures ( $p > .05$  in each case). There was substantial evidence only for digit span, where the null hypothesis of difference was favored.

### Discussion

The baseline predictors of post-training performance differed markedly across the four WM transfer tests. Digit span was very strongly related to digit span prior to training but was not linked with either verbal or performance IQ. Dot matrix, backward digit span and Mr. X were also related to their own baseline scores, although here the strength of association was much weaker than for digit span. Each of these WM tasks was significantly associated with baseline performance IQ but not verbal IQ. This association was strongest for the two visuo-spatial tasks - dot matrix and Mr. X.

Performance IQ was therefore linked with post-training scores only on the same tasks that showed substantial transfer with adaptive training in the previous study: dot matrix, backward span and Mr. X. This is consistent with the suggestion that the flexible cognitive capacities indexed by this measure are involved in the construction and refinement of new routines. It is not obvious why these general resources would be critical if training simply results in increases in the fundamental efficiency of established processes. If so, post-training performance should be best predicted by pre-training scores on the same task, as an index of the baseline capacity on which training experience could be built. This was not the case.

We note that the composition of the sample included in this study is unusual. It combines data from studies of typically-developing children and two samples with compromised WM skills: children selected on the basis of low WM, and children with ADHD. The cognitive heterogeneity of this large sample is ideal for exploring the cognitive origins of transfer through individual differences analyses. However, generalization across other populations of trainees will require further data.

### General discussion

We have proposed that substantial transfer from WM training is a consequence of the development of new routines that are applied to new tasks. In learning a new task, individuals must develop a new cognitive routine that specifies the precise sequence of cognitive processes necessary to perform the task. It is a form of learning that follows

well-established principles of the acquisition of complex skills (Anderson, 1982; Fitts & Posner, 1967; Taatgen, 2013). Once established, new routines can be applied to other similar new tasks. For activities that are already highly practiced before training commences, no new routine will be needed. There will therefore be little scope for either improvement on the trained activity or for transfer.

The approach differs from previous conceptualizations of WM training and transfer in several respects. In contrast to accounts of training with a primary focus on the neuroplasticity (Klingberg, 2010; Takeuchi, Sekiguchi et al., 2010), it does not characterize training-induced change as an undifferentiated process that simply and automatically propagates itself across any activity relying on WM. Instead, transfer is considered to be the direct consequence of the cognitive routines developed during training and, critically, how they can be adapted to fit new tasks at the point of potential transfer. Unlike process- and task-specific accounts of WM transfer (Dahlin et al., 2008; Dunning & Holmes, 2014; Holmes et al., 2009; Minear et al., 2016; Shipstead et al., 2012; Soveri et al., 2017; Sprenger et al., 2013; von Bastian & Oberauer, 2013a), the framework generates predictions about which overlapping task properties both will and will not lead to transfer, and specifies the cognitive conditions necessary for transfer. It does so by building both on a detailed task analysis of common WM tasks and on a skill acquisition framework. Finally, the approach is clearly differentiated from theories of cognitive training which assume that benefits might arise from “learning to learn” (Bavelier et al., 2012; Harlow, 1949). The learning that we believe drives WM training extends only to tasks that can benefit directly from common routines. Learning to learn implies that there are higher-order routines that can be developed in the course of experience with different learning situations which can then be applied to new learning situations. In WM training we believe there is little opportunity to develop sufficiently higher-order routines that can be applied to other paradigms.

The framework was tested in two steps. First, a task analysis of the cognitive processes involved in common WM tasks was performed. This was used to guide the classification of whether or not each task meets the criteria for requiring a new cognitive routine. Framework predictions were then tested in three studies investigating the task features influencing transfer. The first two studies examined the shared characteristics of trained and untrained WM activities associated with transfer. Study 1 reported a meta-analysis of the features associated with transfer in RCTs of WM training. The studies varied widely in the nature of the training activities, transfer tasks, control conditions, participants, statistical power and the methods of analysis. Study 2 investigated the task features associated with transfer for a single WM training program with a small set of WM transfer tests in children with low WM. Study 3 adopted an individual differences approach to investigate the hypothesis that routine-mediated transfer depends primarily not on expansion of existing WM processes, but instead on the recruitment of more general cognitive resources that fall outside of the WM system.

The findings were broadly consistent with the predictions. The task analysis led to the conclusion that the following paradigms require new routines: visuo-spatial serial recall, complex span, backward span, and the updating tasks of n-back and running span. When these paradigms are shared by trained and untrained tasks, there should be transfer. This was indeed found to be the case: transfer was strongly linked with the serial recall of spatial locations (Studies 1 and 2), complex span (Study 1), and backward span (Studies 1 and 2). Data were insufficient for corresponding analysis of the two updating paradigms.

A further prediction was that there should be little transfer across verbal serial recall tasks because they do not need a new routine. This is because mechanisms for encoding, maintaining and retrieving item and order information are served by an existing and highly-practised system of verbal STM. This prediction was partially supported. Transfer was found across verbal serial recall tasks in the meta-analysis (Study 1). However, its magnitude was significantly lower than for visuo-spatial

recall in Study 1 and in the training study (Study 2), it was very weak. We therefore conclude that there may be subtle but genuine training-related improvements in the efficiency of established processes within verbal STM. These gains are small in magnitude and may be reliably detected only under conditions of higher statistical power than the standard WM training study. We speculate that they reflect either the fine-tuning of basic mechanisms or recalibration to optimize task parameters is evident even in low-level perceptual discrimination paradigms (Bavelier et al., 2012). What is clear is that transfer within verbal STM does not have the same robust character associated with more complex WM paradigms that require new routines.

The meta-analysis established that more basic features of tasks such as the input modality (auditory or visual) and output modality (spoken or manual) are not critical for transfer. For example, there were high levels of transfer across backward span tasks that differed only in whether the responses involved spoken recall or mouse-based selection of response alternatives. We had predicted that this would indeed be the case: changes in the sensory modality of either stimulus items or responses should require adjustment only to specialized input and output processing systems employed as low-level elements of routines, and shifting from one to another (from auditory to visual presentation, or spoken to manual responses) should not modify the higher-order structures of the task. It should therefore be possible for routines to be rapidly modified to adapt to these peripheral changes.

A common stimulus category in itself does not appear to be a sufficient basis for transfer (see Minear et al., 2016; von Bastian & Oberauer, 2013a). The meta-analysis established that for letters, words/nonwords and spatial locations, transfer was not influenced by whether the stimulus category was matched except when digits were the memory items in both tasks. Although this may reflect a recoding strategy (or routine) developed during training (Chase & Ericsson, 1981), it could also be due to a confound in the shared features of the relevant task pairs. In all of the cases in this dataset, the trained and untrained tasks shared not only digits but also a backward span paradigm. This was itself predicted to generate high levels of routine-mediated transfer.

The goal of the new framework was to define the boundary conditions on transfer within WM. Our starting position was that routines are hierarchically organized, consisting of subroutines that control the execution of processes necessary to accomplish different task components (e.g., encoding, maintenance, distractor processing, retrieval). At the most peripheral level, these processes include task-specific input and output settings such as the modality of the stimulus inputs or the responses. Switching between these settings should not have an impact on other elements and subroutines: borrowing Taatgen's (2013) terminology, they should not alter the flow of information processing across the system. There should therefore be transfer across low-level settings, as was indeed found in Study 1. In contrast, mismatches in the high-level structure of the subroutines corresponding to the paradigm were expected to limit the adaptability of a routine to a new task. The meta-analysis established that this was the case.

What was much less clear to us in formulating the framework were the limits on transfer when tasks differed at intermediate levels of a routine. The meta-analysis generated new information on this issue. In every pair of complex span tasks included in the analysis, the distractor activity was different. The strength and consistency of transfer found for both verbal and visuo-spatial matched complex span task pairs establishes that transfer does indeed occur when the higher-order structure [(stimulus, distractor processing)<sup>rep</sup>, retrieval] of the routine can be preserved but individual subroutines differ. However, this was true only when the stimulus domain was also matched across the two tasks (verbal or verbal, or visuo-spatial to visuo-spatial): transfer did not extend across complex span tasks in which the domains differed. This suggests that although one distractor subroutine can be substituted for another, the intermediate level of organization is so strongly framed by domain-specific encoding and maintenance processes that the routine

cannot be adapted to fit complex span tasks with stimuli drawn from a different domain. This interpretation clearly requires further systematic experimental analysis.

Study 3 investigated the cognitive skills that might support the development of new routines during training. We speculated that the flexible cognitive resources believed to play a critical role in the process of decomposing the unfamiliar tasks into their constituent cognitive parts (Duncan et al., 2017) may also play a critical role in constructing the bespoke specification of cognitive processes to be executed in a routine. This hypothesis was explored in the re-analysis of training data from a large sample of children completing a single WM training program in Study 3. Higher performance IQ scores were uniquely associated with the magnitude of transfer to the three WM tests (visuo-spatial serial recall, visuo-spatial complex span and backward span) hypothesized to require new routines. In contrast, post-training performance on the verbal serial task (digit span) was very strongly associated only with baseline levels of performance on the same task but not with nonverbal reasoning. This suggests that when WM tasks do not need new routines, training-induced changes reflect modifications in the capacity of existing systems. In contrast when routines must be constructed, they depend on general attentional resources.

Other theoretical accounts of WM training are less successful in accounting for both the presence and absence of transfer with different shared task features in the present studies. Accounts that attribute training and transfer to plasticity in the undifferentiated neural substrate of WM (Klingberg, 2010; Westerberg & Klingberg, 2007) lack the specificity to accommodate limited transfer across WM tasks. On the other hand, process-specific explanations of transfer as reflecting increases in the efficiency of individual processes within WM (Dahlin et al., 2008; Minear et al., 2016; Sprenger et al., 2013) struggle to explain why some shared task features and processes are sufficient for transfer whereas others are not.

These challenges disappear if we move away from the idea that training improves existing processes within WM and consider instead that WM training simply involves learning how to perform an unusual task. This learning is conceived as the construction of new cognitive routines and transfer is only expected for tasks that might be expected to share routines. From this we conclude that there is little prospect that WM training with a small number of tasks will ever have a substantial impact on real-life skills such as those required to enhance educational achievement. Such real-life skills are likely to rely on an extensive array of cognitive routines; too many, probably, to be trained with anything other than real-life experience.

#### Declaration of interests

None.

#### Data statement

The data reported in this article are available in the CBU Open Data Repository <http://www.mrc-cbu.cam.ac.uk/publications/opendata/>.

#### Acknowledgments

Gathercole led the conception and design of the work. Gathercole and Norris developed the theoretical framework and took primary responsibility for drafting the manuscript. Dunning conducted the meta-analysis reported in Study 1, and Dunning and Holmes contributed to the feature coding and commented on drafts. Correspondence concerning this article should be sent to Susan E. Gathercole ([susan.gathercole@mrc-cbu.cam.ac.uk](mailto:susan.gathercole@mrc-cbu.cam.ac.uk)). This research was supported by the Medical Research Council of Great Britain, the University of Cambridge, the Economic and Social Research Council (RES-000-23-0979), the Leverhulme Trust (F00/224/AI), and the Wellcome Trust (104908/Z/14/Z).



## A. Cogmed RM tasks

Measure	Paradigm	Stimulus presentation	Stimulus domain	Stimulus category	Stimulus modality	Response	Training sessions
Visual data link	Visuo-spatial serial recall	Sequence of spatial location of illuminated lamps in 4 × 4 grid	Visuo-spatial	Spatial locations	Visual	Mouse selection of lamp locations in sequence	1–25
Data room	Visuo-spatial serial recall	Sequence of lamps illuminated in a 2D-representation of a 3D grid of 20 lamps	Visuo-spatial	Spatial locations	Visual	Mouse selection of lamp locations in sequence	1–10, 12, 14, 16, 18, 20, 22, 24
Space whack	Visuo-spatial serial recall	Sequence of puffs of smoke appearing from craters	Visuo-spatial	Spatial locations	Visual	Mouse selection of puff locations in sequence	16–25
Input module with lid	Backward span	Sequence of spoken digits with response grid not visible	Verbal	Digits	Auditory	Mouse selection of digits on a 3 × 3 grid in sequence	1–25
Input module without lid	Backward span	Sequence of spoken digits with response grid visible	Verbal	Digits	Auditory	Mouse selection of digits on a 3 × 3 grid in sequence	1–25
Decoder	Verbal serial recall	Sequence of spoken letters	Verbal	Letters	Auditory	Mouse selection of each letter from 3 alternatives presented visually	1–5
Random letters	Verbal serial recall and visuo-spatial recall	Sequence of spoken letters each accompanied by the illumination of a lamp in a particular spatial location	Verbal & visuo-spatial	Letters and spatial locations	Auditory & visual	Mouse selection of spatial locations corresponding to position illuminated as each letter was presented	1–11, 13, 15–21, 23, 25
Numbered grid	Resequencing	Sequence of spoken letters each accompanied by the illumination of a lamp in a particular spatial location	Verbal & visuo-spatial	Digits and spatial locations	Auditory & visual	Mouse selection of spatial locations corresponding to the location of digits in ascending order	6–15, 17, 19, 20, 22, 24
Asteroids	Visuo-spatial serial recall	Slowly-moving asteroids illuminated in sequence	Visuo-spatial	Spatial locations	Visual	Mouse selection of asteroid locations in sequence	11–16, 18, 20, 22–25
Rotating data link	Visuo-spatial serial recall + spatial rotation	Sequence of spatial location of illuminated lamps in 4 × 4 grid	Visuo-spatial	Spatial locations	Visual	Mouse selection of lamp locations in sequence on a grid rotated by 90 degrees	1–10, 12, 14, 16, 18, 21, 23, 25
Rotating dots	Visuo-spatial serial recall + spatial rotation	Sequence of spatial location of illuminated lamps on a rotating dial	Visuo-spatial	Spatial locations	Visual	Mouse selection of dot locations in sequence on rotating dial	1–11, 13, 15, 17, 19, 21–22, 24
Space cube	Visuo-spatial serial recall	Sequence of panels illuminated on a 2D representation of a 3D cube	Visuo-spatial	Spatial locations	Visual	Mouse selection of the panel locations in sequence	11–15, 17, 19, 21–25

## B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.jml.2018.10.003>.

## References

- Åkerlund, E., Esbjörnsson, E., Sunnerhagen, K. S., & Björkdahl, A. (2013). Can computerized working memory training improve impaired working memory, cognition and psychological health? *Brain Injury*, 27(13–14), 1649–1657.
- Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2006). Is the binding of visual features in working memory resource-demanding? *Journal of Experimental Psychology: General*, 135(2), 298–313. <https://doi.org/10.1037/0096-3445.135.2.298>.
- Alloway, T. P. (2007). *Automated working memory assessment*. Pearson Publishing.
- Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development*, 77(6), 1698–1716. <https://doi.org/10.1111/j.1467-8624.2006.00968.x>.
- Anders, T. R., & Lillyquist, T. D. (1971). Retrieval time in forward and backward recall. *Psychonomic Science*, 22(4), 205–206.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89(4), 369.
- Ang, S. Y., Lee, K., Cheam, F., Poon, K., & Koh, J. (2015). Updating and working memory training: Immediate improvement, long-term maintenance, and generalisability to non-trained tasks. *Journal of Applied Research in Memory and Cognition*, 4(2), 121–128.
- Astle, D. E., Barnes, J. J., Baker, K., Colclough, G. L., & Woolrich, M. W. (2015). Cognitive training enhances intrinsic brain connectivity in childhood. *Journal of Neuroscience*, 35(16).
- Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuhl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review*, 22(2), 366–377.
- Awh, E., Vogel, E., & Oh, S.-H. (2006). Interactions between attention and working memory. *Neuroscience*, 139(1), 201–208.
- Baddeley, A. D. (1986). *Working memory Oxford psychology series, No. 11*. New York: Clarendon Press: Oxford University Press.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? Retrieved from *Trends in Cognitive Sciences*, 4(11), 417–423. <<http://www.ncbi.nlm.nih.gov/pubmed/11058819>>.
- Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63, 1–29. <https://doi.org/10.1146/annurev-psych-120710-100422>.
- Baddeley, A. D., & Della Sala, S. (1996). Working memory and executive control. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 351(1346), 1397–1403. <https://doi.org/10.1098/rstb.1996.0123> discussion 1403–1394.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Vol. Ed.), *The psychology of learning and motivation: Vol. V111*, (pp. 47–90). New York: Academic Press.
- Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology*, 36(2), 233–252.
- Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, 27(5), 586–595.
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 575–589.
- Barnes, J. J., Woolrich, M. W., Baker, K., Colclough, G. L., & Astle, D. E. (2016). Electrophysiological measures of resting state functional connectivity and their relationship with working memory capacity in childhood. *Developmental Science*, 19(1), 19–31.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612.
- Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, 133(1), 83.
- Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3), 570.
- Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: A time-based resource-sharing model account. *Developmental Psychology*, 45(2), 477.
- Bavelier, D., Green, C. S., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: Learning to learn and action video games. *Annual Review of Neuroscience*, 35, 391–416.
- Bezzola, L., Mérillat, S., Gaser, C., & Jäncke, L. (2011). Training-induced neural plasticity in golf novices. *Journal of Neuroscience*, 31(35), 12444–12448.
- Bor, D., Duncan, J., Wiseman, R. J., & Owen, A. M. (2003). Encoding strategies dissociate prefrontal activity from working memory demand. *Neuron*, 37(2), 361–367.
- Borella, E., Carbone, E., Pastore, M., De Beni, R., & Carretti, B. (2017). Working memory training for healthy older adults: The role of individual characteristics in explaining short- and long-term gains. *Frontiers in Human Neuroscience*, 11, 99.
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis 2.0*. Englewood, NJ: Biostat.
- Broadley, I., & MacDonald, J. (1993). Teaching short-term memory skills to children with

- Down syndrome. *Down Syndrome Research and Practice*, 1(2), 56–62.
- Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language*, 31(4), 429–460.
- Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, 106(3), 551.
- Caeyenberghs, K., Metzler-Baddeley, C., Foley, S., & Jones, D. K. (2016). Dynamics of the human structural connectome underlying working memory training. *Journal of Neuroscience*, 36(14), 4056–4066.
- Caplan, D., Rochon, E., & Waters, G. S. (1992). Articulatory and phonological determinants of word length effects in span tasks. *The Quarterly Journal of Experimental Psychology*, 45(2), 177–192.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1), 1.
- Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. *Cognitive Skills and their Acquisition*, 141–189.
- Chatham, C. H., Herd, S. A., Brant, A. M., Hazy, T. E., Miyake, A., O'Reilly, R., & Friedman, N. P. (2011). From an executive network to executive control: A computational model of the n-back task. *Journal of Cognitive Neuroscience*, 23(11), 3598–3619.
- Chen, J. M., & Morrison, A. B. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. *Psychonomic Bulletin & Review*, 17(2), 193–199. <https://doi.org/10.3758/PBR.17.2.193>.
- Cicerone, K., & Giacino, J. (1992). Remediation of executive function deficits after traumatic brain injury. *Neurorehabilitation*, 3, 12–22.
- Cohen, J. (1988). The effect size. *Statistical Power Analysis for the Behavioral Sciences*, 77–83.
- Comblain, A. (1994). Working memory in Down syndrome: Training the rehearsal strategy. *Down Syndrome Research and Practice*, 2(3), 123–126.
- Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior*, 4, 161–169.
- Constantinidis, C., & Klingberg, T. (2016). The neuroscience of working memory capacity and training. *Nature Reviews Neuroscience*, 17(7), 438.
- Cortese, S., Ferrin, M., Brandeis, D., Buitelaar, J., Daley, D., Dittmann, R. W., & Stringaris, A. (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.
- Cowan, N. (1998). Visual and auditory working memory capacity. Retrieved from *Trends in Cognitive Sciences*, 2(3), 77. <<http://www.ncbi.nlm.nih.gov/pubmed/21227076>>.
- Cowan, N., Chen, Z., & Roudier, J. N. (2004). Constant capacity in an immediate serial-recall task: A logical sequel to Miller (1956). *Psychological Science*, 15(9), 634–640.
- Cowan, N., & Morey, C. C. (2007). How can dual-task working memory retention limits be investigated? *Psychological Science*, 18(8), 686–688. <https://doi.org/10.1111/j.1467-9280.2007.01960.x>.
- Cowan, N., Roudier, J. N., Blume, C. L., & Saults, J. S. (2012). Models of verbal working memory capacity: What does it take to make them work? *Psychological Review*, 119(3), 480–499. <https://doi.org/10.1037/a0027791>.
- Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684.
- Dahlin, E., Neely, A. S., Larsson, A., Backman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science*, 320(5882), 1510–1512. <https://doi.org/10.1126/science.1155466>.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.
- Darling, S., Della Sala, S., Logie, R. H., & Cantagallo, A. (2006). Neuropsychological evidence for separating components of visuo-spatial working memory. *Journal of Neurology*, 253(2), 176–180.
- De Renzi, E., & Nichelli, P. (1975). Verbal and non-verbal short-term memory impairment following hemispheric damage. *Cortex*, 11(4), 341–354.
- Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unbending visuo-spatial memory. *Neuropsychologia*, 37(10), 1189–1199.
- Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: Changes in grey matter induced by training. *Nature*, 427(6972), 311–312.
- Duncan, J. (2013). The structure of cognition: Attentional episodes in mind and brain. *Neuron*, 80(1), 35–50.
- Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia*, 33(3), 261–268.
- Duncan, J., Chylinski, D., Mitchell, D. J., & Bhandari, A. (2017). Complexity and compositionality in fluid intelligence. *Proceedings of the National Academy of Sciences*, 114(20), 5295–5299.
- Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology*, 30(3), 257–303.
- Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, 23(10), 475–483.
- Dunning, D. L., Westgate, B., & Adlam, A.-L. R. (2016). A meta-analysis of working memory impairments in survivors of moderate-to-severe traumatic brain injury.
- Dunning, D. L., & Holmes, J. (2014). Does working memory training promote the use of strategies on untrained working memory tasks? *Memory and Cognition*. <https://doi.org/10.3758/s13421-014-0410-5>.
- Dunning, D. L., Holmes, J., & Gathercole, S. E. (2013). Does working memory training lead to generalized improvements in children with low working memory? A randomized controlled trial. *Developmental Science*, 16(6), 915–925. <https://doi.org/10.1111/desc.12068>.
- Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *Journal of Educational Psychology*, 84(5), 253–262.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. Retrieved from *Journal of Experimental Psychology: General*, 128(3), 309–331. <<http://www.ncbi.nlm.nih.gov/pubmed/10513398>>.
- Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Belmont, Calif.: Brooks/Cole Publishing Co.
- Flavell, J. H., Beach, D. R., & Chinsky, J. M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development*, 37, 283–299.
- Foster, J. L., Harrison, T. L., Hicks, K. L., Draheim, C., Redick, T. S., & Engle, R. W. (2017). Do the effects of working memory training depend on baseline ability level?
- Foy, J. G., & Mann, V. A. (2014). Adaptive cognitive training enhances executive control and visuospatial and children's performance in laboratory analogues of classroom activities. *International Education Research*, 2(2), 19–43.
- Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language-disordered children: Is there a causal connection? *Journal of Memory and Language*, 29(3), 336–360.
- Gathercole, S. E., Dunning, D. L., Holmes, J., & Norris, D. G. (in press). Feature coding of trained and untrained tasks from randomized controlled trials of working memory training. Data in Brief. <https://dx.doi.org/10.1016/j.dib.2018.11.040>.
- Gathercole, S. E., Durling, E., Evans, M., Jeffcock, S., & Stone, S. (2008). Working memory abilities and children's performance in laboratory analogues of classroom activities. *Applied Cognitive Psychology*, 22, 1019–1037.
- Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(1), 84.
- Gathercole, S. E., & Hitch, G. J. (1993). Developmental changes in short-term memory: A revised working memory perspective. In M. A. Conway, S. E. Gathercole, & C. Corcolli (Vol. Eds.), *Theories of memory: Vol. 1*, (pp. 189–209). Hove: Psychology Press.
- Harlow, H. F. (1949). The formation of learning sets. *Psychological Review*, 56(1), 51.
- Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science*. <https://doi.org/10.1177/0956797613492984>.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- Henry, L. A., Messer, D. J., & Nash, G. (2014). Testing for near and far transfer effects with a short, face-to-face adaptive working memory training intervention in typical children. *Infant and Child Development*, 23(1), 84–103.
- Hesketh, A., & Conti-Ramsden, G. (2013). Memory and language in middle childhood in individuals with a history of specific language impairment. *Plos One*, 8(2), e56314.
- Higgins, J., & Green, S. (2008). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0*. The Cochrane Collaboration.
- Hitch, G., Halliday, M., Hulme, C., Le Voi, M., Routh, D., & Conway, A. (1983). Working memory in children. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 325–340.
- Hockey, R. (1973). Rate of presentation in running memory and direct manipulation of input-processing strategies. *Quarterly Journal of Experimental Psychology*, 25(1), 104–111.
- Holmes, J., Butterfield, S., Cormack, F., Loenhoud, A., Ruggero, L., Kashikar, L., & Gathercole, S. E. (2015). Improving working memory in children with low language abilities. *Frontiers in Psychology*, 6, 519.
- Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, 12(4), F9–F15. <https://doi.org/10.1111/j.1467-7687.2009.00848.x>.
- Holmes, J., Hilton, K. A., Place, M., Alloway, T. P., Elliott, J. G., & Gathercole, S. E. (2014). Children with low working memory and children with ADHD: Same or different? *Frontiers in Human Neuroscience*, 8, 976. <https://doi.org/10.3389/fnhum.2014.00976>.
- Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin*, 140(2), 339–373. <https://doi.org/10.1037/a0034221>.
- Isaacs, E. B., & Vargha-Khadem, F. (1989). Differential course of development of spatial and verbal memory span: A normative study. *British Journal of Developmental Psychology*, 7(4), 377–380.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 105(19), 6829–6833. <https://doi.org/10.1073/pnas.0801268105>.
- Jarrold, C., & Citroën, R. (2013). Reevaluating key evidence for the development of rehearsal: Phonological similarity effects in children are subject to proportional scaling artifacts. *Developmental Psychology*, 49(5), 837.
- Jarrold, C., Tam, H., Baddeley, A. D., & Harvey, C. E. (2010). The nature and position of processing determines why forgetting occurs in working memory tasks. *Psychonomic Bulletin & Review*, 17(6), 772–777. <https://doi.org/10.3758/PBR.17.6.772>.
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Clarendon Press.
- Johnston, R. S., Johnson, C., & Gray, C. (1987). The emergence of the word length effect in young children: The effects of overt and covert rehearsal. *British Journal of Developmental Psychology*, 5(3), 243–248.
- Juvina, I., & Taatgen, N. A. (2007). *Modeling control strategies in the n-back task. Paper presented at the proceedings of the 8th international conference on cognitive modeling*.
- Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. Retrieved from *Journal of Experimental Psychology: General*, 130(2), 169–183. <<http://www.ncbi.nlm.nih.gov/pubmed/11409097>>.

- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133(2), 189–217. <https://doi.org/10.1037/0096-3445.133.2.189>.
- Karbach, J., & Verhaeghen, P. (2014). Making working memory work: A meta-analysis of executive-control and working memory training in older adults. *Psychological Science*, 25(11), 2027–2037.
- Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, 14(7), 317–324. <https://doi.org/10.1016/j.tics.2010.05.002>.
- Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlstrom, K., & Westerberg, H. (2005). Computerized training of working memory in children with ADHD—a randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44(2), 177–186. <https://doi.org/10.1097/00004583-200502000-00010>.
- Kundu, B., Sutterer, D. W., Emrich, S. M., & Postle, B. R. (2013). Strengthened effective connectivity underlies transfer of working memory training to tests of short-term memory and attention. *Journal of Neuroscience*, 33(20), 8705–8715.
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?!. *Intelligence*, 14(4), 389–433.
- Law, C.-T., & Gold, J. I. (2008). Neural correlates of perceptual learning in a sensory-motor but not a sensory cortical area. *Nature Neuroscience*, 11(4), 505.
- Lewandowsky, S., & Oberauer, K. (2015). Rehearsal in serial recall: An unworkable solution to the nonexistent problem of decay. *Psychological Review*, 122(4), 674.
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, 6(7), e1000100.
- Logie, R. H. (1995). *Visuo-spatial working memory*. Hove, Sussex: Lawrence Erlbaum.
- Logie, R. H., & Pearson, D. G. (1997). The inner eye and the inner scribe of visuo-spatial working memory: Evidence from developmental fractionation. *European Journal of Cognitive Psychology*, 9(3), 241–257.
- Loosli, S. V., Buschkuhl, M., Perrig, W. J., & Jaeggi, S. M. (2012). Working memory training improves reading processes in typically developing children. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, 18(1), 62–78. <https://doi.org/10.1080/09297049.2011.575772>.
- Love, J., Selker, R., Marsman, M., Jamil, T., Verhagen, A. J., Ly, A., & al., e. (2015). JASP (Version 0.6.6).
- Lundqvist, A., Grundstrom, K., Samuelsson, K., & Ronnberg, J. (2012). Computerized training of working memory in a group of patients suffering from acquired brain injury. *Brain Injury*, 26(4–5), 423–424 Retrieved from < Go to ISI >://000304104600214.
- Martin, P. R., & Fernberger, S. W. (1929). Improvement in memory span. *The American Journal of Psychology*, 41(1), 91–94.
- Martiniussen, R., Hayden, J., Hogg-Johnson, S., & Tannock, R. (2005). A meta-analysis of working memory impairments in children with attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 44(4), 377–384.
- May, A. (2011). Experience-dependent structural plasticity in the adult human brain. *Trends in Cognitive Sciences*, 15(10), 475–482.
- McLean, J. F., & Hitch, G. J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology*, 74(3), 240–260. <https://doi.org/10.1006/jecp.1999.2516>.
- McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. Retrieved from *Memory & Cognition*, 29(1), 10–17. <<http://www.ncbi.nlm.nih.gov/pubmed/11277453>>.
- Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, 49(2), 270.
- Miller, G. A. (1956). The magic number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 91–97.
- Minear, M., Brasher, F., Guerrero, C. B., Brasher, M., Moore, A., & Sukeena, J. (2016). A simultaneous examination of two forms of working memory training: Evidence for near transfer only. *Memory & Cognition*, 44(7), 1014–1037.
- Montgomery, J. W. (2000). Verbal working memory and sentence comprehension in children with specific language impairment. Retrieved from *Journal of Speech, Language, and Hearing Research: JSLHR*, 43(2), 293–308. <<http://www.ncbi.nlm.nih.gov/pubmed/10757685>>.
- Montgomery, J. W., & Evans, J. L. (2009). Complex sentence comprehension and working memory in children with specific language impairment. *Journal of Speech, Language, and Hearing Research: JSLHR*, 52(2), 269–288. [https://doi.org/10.1044/1092-4388\(2008/07-0116\)](https://doi.org/10.1044/1092-4388(2008/07-0116)).
- Morey, C. C. (2018). The case against specialized visual-spatial short-term memory. *Psychological Bulletin*, 144, 849–883.
- Morey, C. C., & Miron, M. D. (2016). Spatial sequences, but not verbal sequences, are vulnerable to general interference during retention in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(12), 1907.
- Newbury, D. F., Bishop, D. V., & Monaco, A. P. (2005). Genetic influences on language impairment and phonological short-term memory. *Trends in Cognitive Sciences*, 9(11), 528–534.
- Newell, K. M. (1991). Motor skill acquisition. *Annual Review of Psychology*, 42(1), 213–237.
- Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. *Cognitive skills and their acquisition*, 1(1981), 1–55.
- Noack, H., Lövdén, M., Schmiedek, F., & Lindenberger, U. (2009). Cognitive plasticity in adulthood and old age: Gauging the generality of cognitive intervention effects. *Restorative Neurology and Neuroscience*, 27(5), 435–453.
- Norris, D., Hall, J., & Gathercole, S. E. (in preparation). How do we perform backward serial recall?
- Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, 19(5), 779–819. <https://doi.org/10.3758/s13423-012-0272-4>.
- 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>.
- Page, M. P., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. Retrieved from *Psychological Review*, 105(4), 761–781. <<http://www.ncbi.nlm.nih.gov/pubmed/9830378>>.
- Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.
- Parmentier, F. B., Elford, G., & Maybery, M. (2005). Transitional information in spatial serial memory: Path characteristics affect recall performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 412.
- Pearson, D. G., Ball, K., & Smith, D. T. (2014). Oculomotor preparation as a rehearsal mechanism in spatial working memory. *Cognition*, 132(3), 416–428.
- Pearson, D. G., & Sahraie, A. (2003). Oculomotor control and the maintenance of spatially and temporally distributed events in visuo-spatial working memory. *The Quarterly Journal of Experimental Psychology Section A*, 56(7), 1089–1111.
- Phillips, N. L., Parry, L., Mandalis, A., & Lah, S. (2017). Working memory outcomes following traumatic brain injury in children: A systematic review with meta-analysis. *Child NeuroPsychiatry*, 23(1), 26–66.
- Pickering, S. J., Gathercole, S. E., Hall, M., & Lloyd, S. A. (2001). Development of memory for pattern and path: Further evidence for the fractionation of visuo-spatial memory. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 54(2), 397–420. <https://doi.org/10.1080/713755973>.
- Pimperton, H., & Nation, K. (2012). Poor comprehenders in the classroom: Teacher ratings of behavior in children with poor reading comprehension and its relationship with individual differences in working memory. *Journal of Learning Disabilities*. <https://doi.org/10.1177/0022219412454172>.
- Postle, B. R., Berger, J. S., Goldstein, J. H., Curtis, C. E., & D'Esposito, M. (2001). Behavioral and neurophysiological correlates of episodic coding, proactive interference, and list length effects in a running span verbal working memory task. *Cognitive, Affective, and Behavioral Neuroscience*, 1(1).
- Ramus, F., Marshall, C. R., Rosen, S., & van der Lely, H. K. (2013). Phonological deficits in specific language impairment and developmental dyslexia: Towards a multi-dimensional model. *Brain*, 136(Pt 2), 630–645. <https://doi.org/10.1093/brain/aws356>.
- Rapport, M. D., Orban, S. A., Kofler, M. J., & Friedman, L. M. (2013). Do programs designed to train working memory, other executive functions, and attention benefit children with ADHD? A meta-analytic review of cognitive, academic, and behavioral outcomes. *Clinical Psychology Review*, 33(8), 1237–1252. <https://doi.org/10.1016/j.cpr.2013.08.005>.
- Raven, J. (2003). Raven progressive matrices. In S. R. McCullum (Ed.), *Handbook of nonverbal assessment* (pp. 223–237). Boston, MA: Springer.
- Redick, T. S., & Engle, R. W. (2011). Integrating working memory capacity and context-processing views of cognitive control. *Quarterly Journal of Experimental Psychology*, 64(6), 1048–1055. <https://doi.org/10.1080/17470218.2011.577226>.
- Redick, T. S., Shipstead, Z., Wiemers, E. A., Melby-Lervåg, M., & Hulme, C. (2015). What's working in working memory training? An educational perspective. *Educational Psychology Review*, 27(4), 617–633.
- Salamé, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, 21(2), 150–164.
- Salmi, J., Nyberg, L., & Laine, M. (2018). Working memory training mostly engages general-purpose large-scale networks for learning. *Neuroscience & Biobehavioral Reviews*.
- Schuchardt, K., Bockmann, A.-K., Bornemann, G., & Maehler, C. (2013). Working memory functioning in children with learning disorders and specific language impairment. *Topics in Language Disorders*, 33(4), 298–312.
- Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, 50(2), 138–166.
- Shelton, J. T., Elliott, E. M., Matthews, R. A., Hill, B., & Gouvier, W. D. (2010). The relationships of working memory, secondary memory, and general fluid intelligence: Working memory is special. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(3), 813.
- Sheslow, D., & Adams, W. (1990). *Wide Range Assessment of Memory and Learning (WRAML)*. Wide Range Inc.
- Shipstead, Z., Redick, T. S., & Engle, R. W. (2010). Does working memory training generalize? *Psychologica Belgica*, 50(3–4).
- Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, 138(4), 628–654. <https://doi.org/10.1037/a0027473>.
- Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. (2016). Do “brain-training” programs work? *Psychological Science in the Public Interest*, 17(3), 103–186.
- Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Harvard University Press.
- Sonuga-Barke, E. J. S., Brandeis, D., Cortese, S., Daley, D., Ferrin, M., Holtmann, M., & Grp, E. A. G. (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>.
- Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin*



- & Review, 1–20.
- 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.
- Sprenger, A. M., Atkins, S. M., Bolger, D. J., Harbison, J. I., Novick, J. M., Chrabaszcz, J. S., & Dougherty, M. R. (2013). Training working memory: Limits of transfer. *Intelligence*, 41(5), 638–663. <https://doi.org/10.1016/j.intell.2013.07.013>.
- St Clair-Thompson, H. L., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology*, 30(2), 203–219. <https://doi.org/10.1080/01443410903509259>.
- Strobach, T., & Schubert, T. (2016). *Video game training and effects on executive functions cognitive training*. Springer 117–125.
- Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology*, 96(3), 471.
- Swanson, H. L., Xinhua, Z., & Jerman, O. (2009). Working memory, short-term memory, and reading disabilities: A selective meta-analysis of the literature. *Journal of Learning Disabilities*, 42(3), 260–287. <https://doi.org/10.1177/0022219409331958>.
- Szucs, D., Devine, A., Soltesz, F., Nobes, A., & Gabriel, F. (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex*, 49(10), 2674–2688. <https://doi.org/10.1016/j.cortex.2013.06.007>.
- Szucs, D., Nobes, A., Devine, A., Gabriel, F. C., & Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Frontiers in Psychology*, 4, 444. <https://doi.org/10.3389/fpsyg.2013.00444>.
- Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review*, 120(3), 439.
- Takeuchi, H., & Kawashima, R. (2012). Effects of processing speed training on cognitive functions and neural systems. *Reviews in the Neurosciences*, 23(3), 289–301. <https://doi.org/10.1515/revneuro-2012-0035>.
- Takeuchi, H., Sekiguchi, A., Taki, Y., Yokoyama, S., Yomogida, Y., Komuro, N., & Kawashima, R. (2010). Training of working memory impacts structural connectivity. *The Journal of Neuroscience*, 30(9), 3297–3303. <https://doi.org/10.1523/JNEUROSCI.4611-09.2010>.
- Takeuchi, H., Taki, Y., & Kawashima, R. (2010). Effects of working memory training on cognitive functions and neural systems. Retrieved from *Reviews in the Neurosciences*, 21(6), 427–449. <<http://www.ncbi.nlm.nih.gov/pubmed/21438192>>.
- Tenison, C., & Anderson, J. R. (2016). Modeling the distinct phases of skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(5), 749.
- Thomas, J. G., Milner, H. R., & Haberlandt, K. F. (2003). Forward and backward recall: Different response time patterns, same retrieval order. *Psychological Science*, 14(2), 169–174.
- Thompson, J. M., Hamilton, C. J., Gray, J. M., Quinn, J. G., Mackin, P., Young, A. H., & Ferrier, I. N. (2006). Executive and visuospatial sketchpad resources in euthymic bipolar disorder: Implications for visuospatial working memory architecture. *Memory*, 14(4), 437–451. <https://doi.org/10.1080/09658210500464293>.
- Towse, J. N., Hitch, G. J., & Horton, N. (2007). Working memory as the interface between processing and retention: A developmental perspective. Retrieved from *Advances in child development and behavior*, 35, 219–251. <<http://www.ncbi.nlm.nih.gov/pubmed/17682327>>.
- Towse, J. N., Hitch, G. J., & Hutton, U. (1998). A reevaluation of working memory capacity in children. *Journal of Memory and Language*, 39(2), 195–217.
- Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language*, 49, 446–468.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory & Language*, 28, 127–154.
- Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114(1), 104–132. <https://doi.org/10.1037/0033-295X.114.1.104>.
- von Bastian, C. C., Langer, N., Jancke, L., & Oberauer, K. (2013). Effects of working memory training in young and old adults. *Memory & Cognition*, 41(4), 611–624. <https://doi.org/10.3758/s13421-012-0280-7>.
- von Bastian, C. C., & Oberauer, K. (2013b). Effects and mechanisms of working memory training: A review. *Psychological Research*. <https://doi.org/10.1007/s00426-013-0524-6>.
- von Bastian, C. C., & Oberauer, K. (2013a). Distinct transfer effects of training different facets of working memory capacity. *Journal of Memory and Language*, 69(1), 36–58.
- Wass, S., Scerif, G., & Johnson, M. H. (2012). Training attentional control and working memory—Is younger, better? *Developmental Review*, 32(4), 360–387.
- Waters, G. S., Rochon, E., & Caplan, D. (1992). The role of high-level speech planning in rehearsal: Evidence from patients with apraxia of speech. *Journal of Memory and Language*, 31(1), 54–73.
- Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence*. Psychological Corporation.
- Wechsler, D. (2008). *Wechsler adult intelligence scale—fourth edition (WAIS-IV)*. San Antonio, TX: NCS Pearson498.
- Weicker, J., Villringer, A., & Thöne-Otto, A. (2016). Can impaired working memory functioning be improved by training? A meta-analysis with a special focus on brain injured patients. *Neuropsychologia*, 30(2), 190–212. <https://doi.org/10.1037/neu0000227>.
- Weisz, J., & Hawley, K. (2001). *Procedural and coding manual for identification of evidence-based treatments*. Los Angeles: University of California.
- Westerberg, H., & Klingberg, T. (2007). Changes in cortical activity after training of working memory—a single-subject analysis. *Physiology & Behavior*, 92(1), 186–192.
- Wiley, J., Jarosz, A. F., Cushen, P. J., & Colflesh, G. J. (2011). New rule use drives the relation between working memory capacity and Raven's Advanced Progressive Matrices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 256.
- Xerri, C., Merzenich, M. M., Peterson, B. E., & Jenkins, W. (1998). Plasticity of primary somatosensory cortex paralleling sensorimotor skill recovery from stroke in adult monkeys. *Journal of Neurophysiology*, 79(4), 2119–2148.