

Variation in strategy use across measures of verbal working memory

Alexandra B. Morrison¹ · Gail M. Rosenbaum¹ · Damien Fair² · Jason M. Chein¹

Published online: 1 April 2016
© Psychonomic Society, Inc. 2016

Abstract The working memory (WM) literature contains a number of tasks that vary on dimensions such as when or how memory items are reported. In addition to the ways in which WM tasks are designed to differ, tasks may also diverge according to the strategies participants use during task performance. The present study included seven tasks from the WM literature, each requiring short-term retention of verbal items. Following completion of a small number of trials from each task, individuals completed a self-report questionnaire to identify their primary strategy. Results indicated substantial variation across individuals for a given task, and within the same individual across tasks. Moreover, while direct comparisons between tasks showed that some tasks evinced similar patterns of strategy use despite differing task demands, others showed markedly different patterns of self-reported strategy use. A community detection algorithm, aimed at identifying groups of individuals based on their profile of strategic choices, revealed unique communities of individuals who are dependent on specific strategies under varying demands. Together, the findings suggest that researchers using common WM paradigms should very carefully

consider the implications of variation in strategy use when interpreting their findings.

Keywords Working memory · Short term memory · Strategy · Mnemonics

A diverse range of topics in the cognitive and psychological sciences involve the study of short-term working memory (WM)—a system often conceptualized as a flexible mental workspace used for the storage and manipulation of information during the planning and execution of everyday cognitive behaviors (e.g., Baddeley, 2003; Cowan, 2001; Logie, 2003). Accordingly, there is widespread use of tasks designed to index the functioning of WM, and over the past few decades the number and variety of WM assessment tasks has proliferated. With such a widely utilized, yet diverse, collection of WM tasks, an obvious question arises: How similar are these tasks to one another with respect to the mental processes, and strategies, that they evoke? While traditional behavioral research (Cowan, 2001; Luck & Vogel, 2013), psychometric studies (Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004), and cognitive neuroimaging investigations (Nee et al., 2013; Wager & Smith, 2003) all provide some evidence that distinct WM tasks do draw upon shared processes, they also provide evidence that these tasks can operate quite differently from one another with respect to the behavioral phenomena they elicit (Morrison, Conway, & Chein, 2014; Oberauer, 2003; Ricker & Cowan, 2014), the variance they account for across individuals (Unsworth, Fukuda, Awh, & Vogel, 2014), and the specific brain circuitry that they activate (Chein, Moore, & Conway, 2011; Henson, Shallice, Gorno-Tempini, & Dolan, 2002; Zanto, Clapp, Rubens, Karlsson, & Gazzaley, 2016).

Alexandra B. Morrison and Gail M. Rosenbaum contributed equally to the present work.

Electronic supplementary material The online version of this article (doi:10.3758/s13421-016-0608-9) contains supplementary material, which is available to authorized users.

✉ Jason M. Chein
jchein@temple.edu

¹ Department of Psychology, Temple University, Weiss Hall, 1701 N. 13th St., Philadelphia, PA 19122-6085, USA

² Departments of Behavioral Neuroscience and Psychiatry, Oregon Health and Science University, Portland, OR, USA

One likely, but underinvestigated, source of these differences regards the cognitive strategies that individuals deploy when performing alternative WM tasks. Although there has been some limited exploration of strategy use in various WM paradigms (as discussed in greater detail below), many studies seem to be implemented with little thought to the issue of potential variation in strategy use. Others seem to assume, usually implicitly, that there is a *de facto* strategy associated with a given WM task, and that engagement in this strategy is more or less uniform across individuals. In fact, WM measures are sometimes explicitly chosen from among the alternatives because of their assumed strategic properties (i.e., the task is assumed to demand, or to preclude, some particular strategy), and interpretations frequently depend directly on the assumed, but usually not demonstrated or assessed, engagement (or nonengagement) of a particular strategy (e.g., phonological rehearsal of memory items).

In the present study, we set out to explore the distribution of self-reported strategy use across a varied set of WM assessment measures with the goal of providing a more extensive description of how strategy selection varies both within and across tasks, and among individuals. To this end, we sampled strategy patterns across a number of WM measures tested in the same population of individuals, in the same experimental setting, and using the same extended pool of verbal stimuli. Where possible, these tasks were designed to differ systematically with respect to specific properties, thus allowing for an evaluation of the particular features of a task that are most closely associated with alternative strategies.

Working Memory Tasks and Their Properties

WM research emerged historically from the study of short-term memory; with the emphasis shifting from simple, passive storage of information (short-term memory) to a focus on the ability to operate upon, and mentally “work” with, that information in the service of ongoing cognition (working memory; Baddeley & Hitch, 1974). Despite the shift in emphasis, many tasks that were originally conceived as assessments of short-term memory, including variants of the immediate serial recall and Sternberg item recognition tasks, remain among the most widely used measures in the WM literature (see D’Esposito & Postle, 2015; Hurlstone, Hitch, & Baddeley, 2014; Jonides et al., 2008; Logie, DellaSala, Laiacina, Chalmers, & Wynn, 1996; Oberauer & Lewandowsky, 2008). However, the shift in emphasis also brought with it the introduction of new tasks into the WM researcher’s arsenal (and increased the use of some previously less common tasks), including, for example, variants of the complex WM span task (Conway et al., 2005; Daneman & Carpenter, 1980; Oberauer, Lewandowsky, Farrell, Jarrod, & Greaves, 2012; Redick

et al., 2012; Unsworth, Heitz, Schrock, & Engle, 2005) and running memory span task (Broadway & Engle, 2010; Bunting, Cowan, & Saults, 2006; Cowan et al., 2005; Pollack, Johnson, & Knaff, 1959).

To pursue ever more nuanced aspects of the processes that underlie WM and its capacity limits, researchers have designed and deployed an increasingly extensive array of different WM measures. While there are far too many WM paradigms and variants to list exhaustively, a subset of these tasks is shown schematically in Fig. 1. All of these paradigms share the core requirement that some type of information must be maintained over the short term, and it is this essential feature that makes them “working memory” tasks. Moreover, a robust literature establishes that this shared aspect of WM tasks at least partially explains why they typically correlate positively with one another, and why they significantly predict individual differences in many complex cognitive abilities, such as reading comprehension, problem solving, and fluid intelligence (Kane et al., 2004; Unsworth et al., 2014).

Unfortunately, our understanding of the ways in which seemingly similar WM tasks diverge, and the implications of this divergence, is limited. We do know that task differences are important—for instance, even fundamental characteristics of WM, such as capacity estimates (see Cowan, 2001, for review), forgetting rates (Ricker & Cowan, 2014), and primacy and recency effects (Morrison et al., 2014; Oberauer, 2003) are known to vary across tasks—but we know surprisingly little about which specific qualities of each WM paradigm have this influence on the observed findings. One way to gain some traction in our understanding is to engage in task analysis (to evaluate the specific characteristics of a given task implementation) and to explore how task demands affect the strategies that a given task elicits or encourages.

Some dimensions of WM task variation can be fairly readily characterized (Fig. 2 provides a list of some of these dimensions). For example, an obvious difference between immediate serial recall and item recognition tasks lies in the way in which memory is reported (i.e., response demands); as their names suggest, immediate serial recall requires recall, while item recognition requires recognition. These two tasks additionally differ in terms of their emphasis on the serial order in which memoranda are maintained and reported; immediate serial recall requires ordered recollection whereas item recognition does not. To retain the requirement for item recall (as in immediate serial recall), but without the emphasis on serial order of presentation, some researchers have instead used a free recall paradigm (Beaman & Jones, 1998; Marsh, Sorqvist, Hodgetts, Beaman, & Jones, 2015; Neely & LeCompte, 1999; Unsworth, Spillers, & Brewer, 2010), in which participants are allowed to remember and report items in whatever order they desire. Another factor differentiating WM paradigms regards the duration of the retention interval, with some tasks

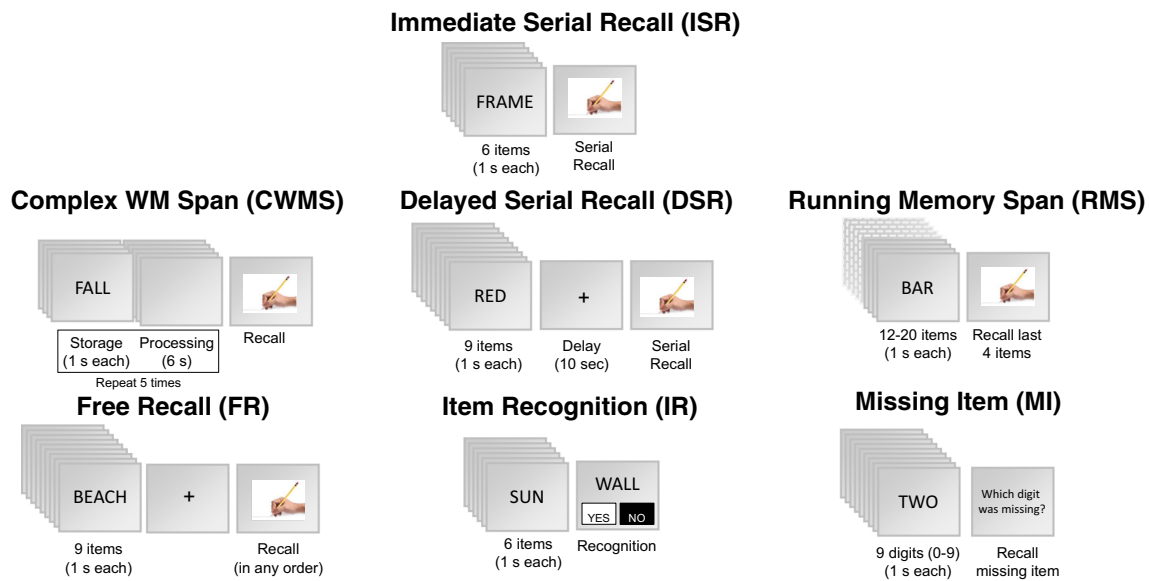


Fig. 1 Schematic of the seven working memory tasks included in the present study.

involving the assessment of memory (essentially) immediately after presentation (as in immediate serial recall), and others imposing a longer, though still “short-term,” delay interval between initial presentation and retrieval (as in delayed serial recall; e.g., Chein & Fiez, 2001; Farrell, 2006). WM tasks are also distinguished by the presence or absence of intervening distractors (i.e., extraneous stimuli or processing requirements) and the relative need for memory updating (replacing the previous contents of WM with newly relevant information). For instance, complex WM span tasks (e.g., Conway et al., 2005)

require that information be maintained and updated across periods of responding to a secondary processing task, while running memory span (e.g., Bunting et al., 2006) tasks require that the contents of WM be continuously updated such that earlier (now extraneous) information is replaced by more recently presented information.

There are, of course, many other dimensions that can be varied through the specific parameterization and implementation of a given WM paradigm; stimulus qualities, speed of item presentation, modality of presentation, and list length,

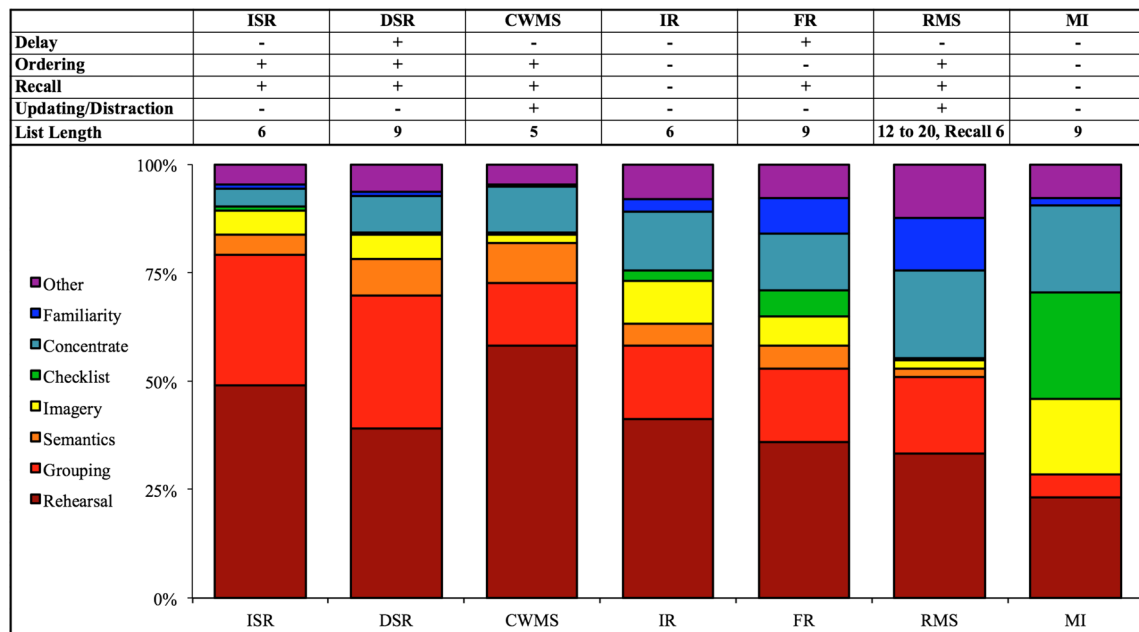


Fig. 2 Top: Demands present or absent in each task. Bottom: Graph of percentage of participants employing each of the more common strategy responses. “Other” includes participants who reported a strategy other than one listed as well as those who reported less common strategies,

such as Look and Sound. *Note.* ISR = immediate serial recall; DSR = delayed serial recall; CWMS = complex working memory span; IR = item recognition; FR = free recall; RMS = running memory span; MI = missing item.

to name a few. These factors also undoubtedly influence the specific processes and strategies that are evoked in support of task performance. Since there is an almost infinite number of WM tasks that can be created by systematically varying each factor, one cannot hope to exhaustively categorize all possible variants of WM tasks. However, by beginning with a finite and relatively common set of WM tasks and parameterizations, we can begin to understand whether (and how) strategy use varies as a function of the specific elements of a WM task.

Assessment of Working Memory Strategies

There is a small but informative extant literature on strategy use in WM tasks, with various different approaches used to collect strategy information. These approaches include (1) investigating the behavioral impacts of instructing participants to engage in a predetermined strategy (Carretti, Borella, & De Beni, 2007; McNamara & Scott, 2001; St Clair-Thompson, Stevens, Hunt, & Bolder, 2010; Turley-Ames & Whitfield, 2003); (2) allowing participants to control the pace of item presentation and measuring looking time as a correlate of the amount of strategic processing (Engle, Cantor, & Carullo, 1992); (3) designing stimuli in a manner that is more amenable to certain strategies, like grouping items or organizing them based on prior knowledge (Bor, Cumming, Scott, & Owen, 2004; Bor & Owen, 2007); (4) assessing overt verbal behaviors during the task's intertrial intervals (Lehmann & Hasselhorn, 2007); and (5) asking participants to self-report their strategy following each individual trial or group of trials, through either open-ended or structured questionnaires (e.g., Dunlosky & Kane, 2007; Richardson, 1998).

Studies using the last approach are the most common and have already been conducted with immediate serial recall (Logie et al., 1996), free recall (Hertzog, McGuire, & Lineweaver, 1998) and complex WM span measures, such as operation span and reading span (Bailey, Dunlosky, & Kane, 2011; Dunlosky & Kane, 2007; Friedman & Miyake, 2004). These prior studies are informative in several ways. First, they provide an initial characterization of the range of specific strategies that might be evoked when performing WM tasks. Prior investigations also consistently show that strategy use is not homogenous even within a given task, and that performance can be contingent on strategy choice—strategies such as visual imagery, forming sentences with memory items, and grouping items have been found to be normatively effective strategies, while simply reading or repeating (i.e., phonologically rehearsing) memory items have been found to be normatively less effective (Bailey et al., 2011; Dunlosky & Kane, 2007). Strategy use is also found to be a contributing factor to the relationship between performance on WM tasks and other types of cognitive tasks (e.g., reading comprehension

and episodic memory), in particular when the same strategy can be used to support performance in both the WM and other assessed cognitive tasks (Bailey, Dunlosky, & Kane, 2008). Furthermore, flexible use of strategies may be a characteristic of higher performing individuals (Dunlosky & Kane, 2007).

Although this body of prior work provides an important foundation for the present investigation, the specific methods used in past studies have important limitations with respect to the present study objectives. Here we aim to inform researchers asking questions such as, “If I choose a certain WM task, what types of strategies should I expect from participants?” and “To what extent can I assume that alternative WM tasks are interchangeable and will produce comparable distributions of strategy use?” Extensions of this research may ultimately help the field answer questions such as “Which WM task is best suited for my research question?” Results from prior studies are helpful but limited because of the restricted range of tasks considered, and because of variation in the experimental context, stimuli, and strategy assessment methods used across studies. Accordingly, the present study was designed to assess a range of self-reported strategies for a larger set of verbal WM tasks, each assessed under the same basic experimental conditions. Furthermore, while prior work on WM strategy use has focused on the relationship between performance and strategy choice, our immediate focus is not on the way in which strategy use can optimize performance but on the ways in which experimental design factors can impact strategy use.

Present Study

The present work explores differences in strategy use in seven verbal WM tasks by examining differences among individuals in the same task and within individuals across multiple tasks. We examine the data from several points of view. With respect to tasks, we evaluate the distribution of strategies employed for a given task, the relative uniformity of intratask strategy choice, intertask variation in strategy selection, and the relationships between strategy choice and specific task dimensions/characteristics. With respect to individuals, we evaluate consistency of strategy choice across tasks and the existence of communities of individuals who behave similarly to one another across variations in task demands.

Method

Participants

Two-hundred and twenty Temple University undergraduate students participated in this study (137 female, 76 male, 7

did not specify their sex) with a mean age just under 21 years ($M = 20.80$, $SD = 4.67$, range 17 to 56). All participants were awarded course credit.

Working Memory Tasks

Selection of Tasks Each participant completed seven verbal WM tasks (see Fig. 1). This subset of tasks was selected following a review of the WM literature, which produced an extensive archive of tasks, varying on many dimensions. Inclusion of a particular WM measure in the present study was based on three constraints: (1) The frequency of use in the literature, with the aim of including more common tasks; (2) the ability to form relatively parallel/analogous versions of each paradigm in a way that would support task comparisons and isolate specific relevant task properties; and (3) the amount of time needed to instruct and execute each measure. It was with these constraints in mind that the following six tasks were selected: immediate serial recall, delayed serial recall, free recall, item recognition, running memory span, and complex WM span. These six tasks are alike in many respects but vary on specific dimensions, as clarified in Fig. 2.

A seventh task, the missing item task (Beaman & Jones, 1997; Buschke, 1963; Jones & Macken, 1993), was selected to be more disparate from the other six tasks and, hence, to evoke what we expected would be a demonstrably different pattern of self-reported strategy use, thus serving as a control task for the sensitivity of the self-report strategy measure to variation in strategy use. The missing item task differs from the other tasks we administered in several ways. In the present study, it was the only task that required a different set of stimuli (digits spelled out zero through nine). Rather than requiring the retention of a novel set of words, this task relies on memory for a set of items known to the participants prior to the onset of a trial, and with an inherent ordering structure that differs from the order of presentation during the task. Since the complete stimulus set is known a priori, and all but one item from the set is presented in each trial, participants are thought to use a unique strategy to support task performance, wherein each presented item is mentally noted and “checked off” the mental list of potential items (Beaman & Jones, 1997).

Task Parameterization The specific parameters for each of the seven tasks were selected with two goals in mind: consistency with methods from the prior literature and maximal matching of parameters in order to support comparison of tasks differing only according to key variables of interest. To support comparison of strategies across tasks, a common set of procedures and stimuli were used throughout—all tasks required memory for verbal memoranda, were presented visually, and used sequential, rather than simultaneous (a single array of items) presentation. Items were also always presented

for 1 second each, with no gap between successive stimulus presentations (no interstimulus interval) in any of the tasks, with the exception of the complex WM span task, wherein an intervening processing interval imposed a necessary interstimulus delay (but there was still no interval between the memory items and the processing task).

The goal of maximizing the match between tasks sometimes required deviation from the most typical implementation parameters in the extant literature, and the most notable of these deviations are detailed below. Likewise, in order to maintain consistency with the past literature while also supporting certain pairwise task comparisons, it was necessary at times to set parameters (e.g., list length, timing of processing decisions in the complex WM span task, the inclusion a delay prior to retrieval) in such a way that actually confounded pairwise comparisons between more disparate tasks. Again, these potential confounds are considered in the discussion.

Stimuli and Procedure

All seven tasks involved memory for verbal items. Word items used for all but the missing item task were selected from the MRC Psycholinguistic Database (Wilson, 1988) with the following constraints: stimuli were all one syllable, contained a maximum of six letters, had an imagability score of at least 500, and had a written frequency of at least 50. Within these constraints, 70 words were selected (see Supplemental Materials A). For each task, one of the lists of 10 phonologically distinct words was selected, except for running memory span (which required 20 words per trial; see task explanation below) and the missing item task (which used number words as stimuli). For a given task the same word list was used repeatedly across a practice trial and six experimental trials. On a given task trial, items were sampled randomly without replacement from among the items on the selected list. The missing item task necessarily involves memory for a known set of items, and accordingly, the digits 1 through 9 spelled out as words (e.g., one, two, three) were always used for this task.

Individuals participated in a single 1-hour session. Sessions were run in groups of between 2 and 10 participants. Each session consisted of all seven WM tasks. Task stimuli were projected onto a 5-ft. × 7-ft. screen. Participants viewed the screen from 7 to 10 feet away. Words were presented in white on a black background in Arial font, size 26, subtending approximately 0.44×1.42 – 2.48 degrees of visual angle (depending on the length of the word). For all tasks, responses were provided in written form, and given on a task-appropriate response sheet provided immediately prior to testing of each task. For recall tasks, the response sheets included blank spaces where each of the presented words could be written for each trial. For the one recognition task (item recognition), a checkbox for “yes” and “no” responses could be checked off

for each trial. For the complex WM span task, additional slots were included for responding to the intervening math processing items.

For each task, participants were instructed verbally and additionally read written instructions. Every task included a practice trial, followed by a brief question and answer period to ensure understanding. Visibility of stimuli was confirmed by all participants following the practice trial for each task, and this was followed by six experimental trials for each task. Following each task, participants completed a strategy questionnaire for that task only, and after the completion of the strategy questionnaire, participants were introduced to the next task. This cycle of tasks and questionnaires continued for the remainder of the session.

The order of task presentation was counterbalanced across experimental group sessions. Based on the counterbalancing scheme, each individual task occurred sometimes as the first task in the session (i.e., it was the first task attempted by participants and was therefore the first task for which strategy assessment was conducted), and sometimes as a later task in the session. The counterbalancing also assured a varied sequence of task ordering for each session, such that task history would not have a systematic impact on strategy selection.

Tasks

Immediate Serial Recall Participants were shown six words, one at a time, for 1 second each, without an interstimulus interval (ISI). Immediately following presentation of the last item, participants were prompted, by an icon showing a hand in the act of writing, to report all six items in the order in which they had been presented.

Item Recognition This task differed from the immediate serial recall task in only the response requirements. Six items were presented one at a time, for 1 second each, without an ISI. At the conclusion of each trial, participants were then shown a single probe word and asked to respond “yes” if the word was on the list and “no” if it was not. The word was present on the list 50 % of the time and absent 50 % of the time (lures were chosen from the four unsampled stimulus items on the 10-item list). A set size of six was chosen to maintain a consistent set size between the immediate serial recall task and the item recognition task. While item recognition tasks are often implemented with simultaneous presentation of the memoranda, for consistency with other tasks we used sequential presentation.

Delayed Serial Recall Nine words were displayed one at a time, for 1 second each, and with no ISI. Following item presentation there was a delay of 10 seconds, during which a fixation cross was presented on the center of the screen. After

the delay, participants were prompted (by icon) to recall the nine words in serial order.

Free Recall The free recall task was identical to the delayed serial recall task, with the exception of the response requirements. Nine words were presented sequentially, for 1 second each, and with no ISI. After a 10-second delay/fixation period, participants were prompted to report the words in any order they chose. While the free recall task has been previously implemented with and without a delay (e.g., Beaman & Jones, 1998), a delay before recall was included here to allow for a close comparison between the free recall task and the delayed serial recall task. Although atypical for this task in the literature, in order to be consistent with the other tasks, stimuli were also sampled from the same 10-item list of words across all six trials (and thus, although presentation order varied, only one item differed from one trial to the next). Since participants were free to recall the items in any order, it could be expected that free recall of the items would become relatively stereotyped, and more accurate, across trials.

Complex WM Span Participants were shown five words and, between the presentation of each word, were asked to judge the veracity of a solved arithmetic equation (e.g., $(5 + 3) / 2 = 4$). Memoranda were displayed for 1 second each. In prototypical versions of this type of complex WM span task (referred to as the operation span task), either an experimenter ensures that each successive item is presented immediately after responding to the equation is completed (Kane & Engle, 2003) or automated timing is tailored to an individual’s speed of solving a set of practiced mathematical operations (Unsworth et al., 2005). In order to present the task to a group of participants, we allotted a fixed time for participants to solve the math problems (6 seconds), based on the average response duration obtained from a large sample ($N > 100$) of prior participants who had completed the automated version of the task in our research lab (Chein, 2008).

Running Memory Span This task tests memory for the final items of a list of unpredictable length. Successive stimuli were presented for 1 second each, one at a time, and with no ISI. The length of the list varied unpredictably across trials (12–20 items); thus, we used a pool of 20 words for this task (rather than 10, as in other tasks). When the list ended, participants were prompted (by icon) to report the last six words that had been shown in the presented order.

Missing Item For the missing item task, the set of digits zero through nine were spelled out as words. Nine of the 10 digits were presented for 1 second each, sequentially, and without an ISI. After item presentation was completed, participants were prompted to report the single item that was not presented in the trial.

Strategy Assessment and Distribution Analyses

A strategy questionnaire was administered immediately following each task (see Supplemental Materials A). The 10 strategies included on the questionnaire were constructed based on a comprehensive search through the strategy literature, open-ended strategy reports provided by participants at the conclusion of a 1-month WM training study, where training included complex WM span tasks (Chein & Morrison, 2010),¹ and open-ended strategy reporting provided during a brief pilot of the present experiment. The 10 strategies identified from these sources were assessed via endorsement of the following statements, with the words in boldface lettering intended to highlight the characteristic quality of each strategy (words shown in parenthesis after each statement indicate our nomenclature for these strategies but were not shown on the questionnaire):

I silently **repeated** the items (Rehearsal), I remembered the words in **groups** (Grouping), I used the **meaning** of the words to remember or connect them (Semantic), I thought about **other things** that could relate to the words (Association), I pictured the way the words **looked** on the screen (Look), I created a visual **image** based on the meaning of the words (Imagery), I thought about the way the words **sounded** (Sound), I simply **concentrated** on the words (Concentrate), I answered based on what words seemed recent or **familiar** (Familiarity), I expected certain words to appear and mentally **checked them off** as they arrived (Checklist).

On the questionnaire, participants could also indicate that they had used another unidentified strategy, no particular strategy, or didn't understand the task instructions/demands.

The order in which the list of strategies was presented was counterbalanced across sessions so that selection would not be biased according to location in the list (i.e., because a given strategy appeared more often near the top or bottom of the page). However, for a given participant, the strategy list always appeared in the same order throughout the session. During analysis it was determined that some statements were endorsed very infrequently, and the associated strategies were ultimately collapsed into a single Other category along with any additional previously unidentified strategies that participants described. Additionally, the None category was grouped

with Concentrate (see Results for further explanation of collapsed categories).

The strategy questionnaire included two pages that were intended to assess primary and secondary strategy use, respectively. On page one of the questionnaire, participants were asked to endorse a single strategy from the list as *the* strategy that best described their approach to the task. This selection was considered their “primary strategy.” On page two, which was identical to the first page with the exception of the instructions at the top of the page, participants were asked to indicate other strategies that were used to a lesser extent (i.e., secondary strategies) during task performance. The current paper focuses on primary strategies only. Secondary strategies were very erratically and variably reported (e.g., many participants never reported them, while others reported using nearly every strategy at some point within a single task). We were accordingly concerned about the validity and reliability of this component of the instrument and opted to focus exclusively on primary strategy use.

Main analyses utilized chi-square testing to determine which tasks evoked consistent and varied patterns of strategy choice. Additional descriptive analyses explored within-individual variation in strategy choice across specific task pairings. Finally, as a further way to identify patterns within the strategy-choice data, we also deployed community detection algorithms based on graph theory (using the Connectivity Toolbox in MATLAB; Rubinov & Sporns, 2010). These data-driven algorithms allowed us to explore whether participants naturally grouped into “communities” based on their across-task strategy distribution profiles. A detailed explanation of these analyses can be found in Supplemental Materials E.

Results

Strategy Variation Within and Across Tasks

An initial aim of the present study was to provide a characterization of the range and distribution of strategies that individuals engage when performing alternative WM tasks. Across all tasks, the strategy assessment data indicated that certain strategies were often reported as a primary strategy, while others were rarely reported. In order to support direct comparisons of the strategy distribution patterns across tasks, infrequently reported strategies (i.e., those reported less than 2.5 % of the time, representing less than about 35 total reports of use) were collapsed together with cases where participants reported using a strategy not listed (those reporting Other) to create a single Other category (Supplemental Materials B provides a chart with the complete, uncollapsed data). Collapsed strategies included Look, Sound, and Association.

Additionally, some participants reported using no strategy (selected None) on certain tasks (2.23 %). Because it was

¹ In Chein and Morrison (2010), participants engaged in complex WM span tasks for a total of 10 hours. Strategy use was not discussed with participants during training, but following training participants were asked to detail the strategies they engaged. Participants' time on task was substantially more than a typical single session study of WM, and they reported a rich and varied set of strategies that were very informative in the design of the present study.

reported infrequently, we considered grouping these responses in the Other category. Alternatively, it occurred to us in retrospect that the None response might be used in a fashion equivalent to the Concentrate option. Like Concentrate, selection of None could indicate a “default” approach to the task at hand with no other supplementary strategy. The latter grouping of None was corroborated by the fact that the frequency of use of these two strategy responses (None, Concentrate) was strongly correlated across tasks ($r = .64, p = .06$, one-tailed; the correlation does not reach significance because there were only seven contributing cases). Accordingly, in final analyses, we grouped these two response types (None and Concentrate) into a single class.² Last, data were treated as missing for the rare occasions when participants reported that they did not understand the task (1.33 % of the time for each task, on average).

At an initial, descriptive level there were several similarities in strategy distribution across all of the tasks (see Fig. 2 for a graphical representation of the strategy distributions, and Supplemental Materials B for an uncollapsed version, as well as a table with exact percentages). Unsurprisingly, the most commonly reported strategy was Rehearsal, making up 39.85 % of all strategy reports. Grouping was the next most popular strategy, comprising 18.87 % of primary strategy reports. In fact, for all tasks except the missing item task, Rehearsal and Grouping made up at least half of the primary strategies reported.

Despite commonalities across tasks, there were also notable differences. Chi-square tests were used to determine whether strategy use differed significantly across tasks (see Table 1). Indeed, even though the majority of the tasks shared the two most popular strategies, a chi-square test including all task strategy frequencies indicated significant differences, $\chi^2(42) = 455.45, p < .001$. As anticipated, missing item showed the most disparate strategy distribution, and pairwise chi-square tests confirmed that the frequency of strategy choices in the missing item task differed significantly from every other task strategy distribution (all $ps < .001$). Of note, participants reported using the Checklist strategy more often than in any other task (24.63 % in missing item vs. 1.82 % across all other tasks), though this strategy was by no means universal in this task.

Because the missing item strategy distribution was so different from the other tasks, we repeated chi-square analyses while excluding the missing item task, which still produced a highly significant result, $\chi^2(35) = 189.45, p < .001$. To further explain this omnibus effect, additional chi-square tests were conducted comparing the strategy differences between pairs of tasks. Including the pairwise comparisons between missing item and every other task, we ran a total of 21 pairwise chi-squares. To control for family-wise error rate, we used

² We ran the analyses both grouping None with Other and excluding None, and the results did not qualitatively change.

Table 1 Chi-square results

	χ^2	<i>df</i>	<i>p</i>
All tasks	455.45	42	<.001
All tasks (Without MI)	189.45	35	<.001
ISR vs. IR	25.83	7	<.001
ISR vs. CWMS	4.69	7	<.001
ISR vs. DSR	8.40	7	.30
ISR vs. FR	40.57	7	<.001
ISR vs. RMS	65.09	7	<.001
ISR vs. MI	139.84	7	<.001
IR vs. CWMS	27.05	7	<.001
IR vs. DSR	19.94	7	.006
IR vs. FR	9.72	7	.21
IR vs. RMS	33.94	7	<.001
IR vs. MI	76.51	7	<.001
CWMS vs. DSR	23.39	7	<.002
CWMS vs. FR	42.80	7	<.001
CWMS vs. RMS	57.79	7	<.001
CWMS vs. MI	135.73	7	<.001
DSR vs. FR	33.02	7	<.001
DSR vs. RMS	52.34	7	<.001
DSR vs. MI	130.80	7	<.001
FR vs. RMS	24.77	7	<.001
FR vs. MI	74.50	7	<.001
RMS vs. MI	112.11	7	<.001

Note. MI = missing item; ISR = immediate serial recall; IR = item recognition; CWMS = complex working memory span; DSR = delayed serial recall; FR = free recall; RMS = running memory span.

a Bonferroni-corrected cutoff *p* value of .0024 (.05/21 = .0024). Of interest, only three of the 21 tests did *not* reach significance even after correction: immediate serial recall and delayed serial recall, $\chi^2(7) = 8.40, p = .30$; item recognition and free recall, $\chi^2(7) = 9.72, p = .21$; and item recognition and delayed serial recall, $\chi^2(7) = 19.94, p = .006$. Chi-square tests were significant for all other pairs of tasks, suggesting that the strategy distributions across most tasks demonstrated significant variability.

To examine which task dimensions influenced strategy selection and which did not, we conducted an additional series of chi-square analyses based on the following specific task dimensions (see Fig. 2 and Table 2): delay length (immediate vs. delayed), list length (short vs. long), order requirement (ordered vs. free), response demands (recall vs. recognition), and updating/distraction demands (passive storage vs. storage with updating/distraction). Tasks falling within each dimensional class were combined and chi-square values computed. Strategy distributions did not differ based on delay or list length ($ps > .10$). However, strategy selection did vary based on serial order requirements (ordering), response demands (recall), and the updating/distraction dimension ($ps < .001$).

Table 2 Chi-squares collapsed across task dimensions

Dimension	Tasks +	Tasks -	χ^2	<i>df</i>	<i>p</i>
Delay	DSR, FR	ISR, IR, CWMS, RMS	11.35	7	.12
Ordering	ISR, CWMS, DSR, RMS	IR, FR	41.08	7	<.001
Recall	ISR, CWMS, DSR, FR	IR, RMS	49.79	7	<.001
Updating/ Distraction	CWMS, RMS	ISR, IR, DSR, FR	40.77	7	<.001
List Length	6 items—ISR, IR	9 items—DSR, FR	11.94	7	.10

Note. DSR = delayed serial recall; FR = free recall; ISR = immediate serial recall; IR = item recognition; CWMS = complex working memory span; RMS = running memory span.

Exploration of Cross-Task and Strategy “Contamination”

Our experimental design necessitated that participants complete several WM tasks successively, providing a strategy report after each task. This design made it possible that earlier exposure to the strategy-use questionnaire could influence participants’ strategy choice in subsequent tasks (e.g., introduce novel strategies that the participant might want to try later in the session). Participants’ strategy reports may also have been influenced by exposure to prior tasks in the session. To statistically address these possible “contamination” effects, we ran additional chi-square tests comparing the distribution of reported strategies when a task occurred as the first task in the session (ensuring no prior exposure to the strategy report sheet or other tasks) to the distribution when the same task was presented later in the experiment. Only one of the tasks (item recognition) significantly differed in strategy distribution when it was presented first compared to later, $\chi^2 = 18.51$, $p = .01$. However, this test did not survive correction for multiple comparisons (.05/7 tasks = $p < .007$). Therefore, there is very limited evidence that a different variety of strategies were reported after exposure to the strategy sheet and/or other tasks. Stacked graphs displaying strategy reports for participants’ completing each task as the first in the session, compared to those who completed the same task later in the session, are presented in Supplemental Materials C.

Task Performance as a Function of Strategy

Table 3 summarizes performance in each of the WM tasks. While the main focus of the present study was to investigate strategy use and variation across WM tasks, some prior work has found a link between participants’ strategy use and their task performance (Bailey et al., 2011; Dunlosky & Kane, 2007). Accordingly, we ran a series of tests to investigate whether task performance varied systematically with strategy choice. Specifically, for each task independently, we conducted a one-way ANOVA on task accuracy, treating strategy choice as a between-subjects factor. These analyses indicated no significant variation in performance as a function of strategy choice (all Bonferroni-corrected and uncorrected $ps > .1$).

Additional post hoc analysis of the performance data is reported in Supplemental Materials D.

Strategy Variation across Individuals

We were interested not only in how certain tasks encourage the use of distinct strategies but also in how individuals choose different strategies across tasks. It is possible that some participants are more likely to keep strategy use constant across tasks, while others are more likely to switch strategies depending on the task at hand. Figure 3 shows frequency counts detailing the correspondence (data falling on the main diagonal) or lack of correspondence (data falling off of the diagonal) in individuals’ strategy choices across several task pairings. Figure 3a displays the correspondence between immediate serial recall and delayed serial recall, whose overall strategy distributions did not significantly differ from one another. Despite the similar distributions, about half of the participants indicated use of different primary strategies (i.e., switched strategies) to perform the two tasks. In item recognition compared to free recall (see Fig. 3b), whose strategy distributions also did not significantly differ, only about a third of participants maintained their strategy across the tasks. So, it seems that even when the overall distribution of strategies is similar across tasks, individuals often do not maintain consistent

Table 3 Mean (standard deviation) of response accuracy for each task

Task	Mean (SD)
Immediate serial recall (avg. % recalled in position, out of 6 items)	66 % (21)
Delayed serial recall (avg. % recalled in position, out of 9 items)	33 % (15)
Complex working memory span (avg. % recalled in position, out of 5 items)	75 % (3)
Item recognition (avg. % old/new recognition accuracy)	88 % (19)
Free recall (avg. % recalled out of 9 items)	66 % (15)
Running memory span (avg. % recalled in position, out of 6 items)	35 % (16)
Missing item (avg. % items accurately checked off as missing)	59 % (26)

a Individual Strategy Variation from Immediate to Delayed Serial Recall
 -53% maintained strategy
 -64% Rehearsed or Grouped in both

		Immediate Serial Recall							
		R	G	S	I	C	Rec	Ch	O
Delayed Serial Recall	R	52	15	3	1	2	0	1	2
	G	18	35	1	1	1	1	0	2
	S	7	1	2	2	0	1	0	0
	I	2	2	1	5	0	0	0	0
	C	9	2	0	0	2	0	1	0
	Rec	1	0	1	0	0	0	0	0
	Ch	0	0	1	0	0	0	0	0
	O	5	1	0	2	1	0	0	4

b Individual Strategy Variation from Item Recognition to Free Recall
 -34% maintained strategy
 -66% Rehearsed or Grouped in both

		Item Recognition							
		R	G	S	I	C	Rec	Ch	O
Free Recall	R	44	8	3	2	2	3	0	5
	G	9	8	2	3	7	0	3	2
	S	3	2	1	1	2	0	0	2
	I	3	5	1	3	0	1	0	0
	C	3	2	1	5	7	1	0	4
	Rec	6	3	0	3	4	0	1	0
	Ch	4	4	0	0	2	0	1	1
	O	6	2	1	1	2	1	0	2

c Individual Strategy Variation from Immediate Serial Recall to Complex Working Memory Span
 -48% maintained strategy
 -65% Rehearsed or Grouped in both

		Immediate Serial Recall							
		R	G	S	I	C	Rec	Ch	O
Complex WM Span	R	68	29	3	3	2	1	0	4
	G	7	14	2	3	1	0	1	0
	S	6	3	2	3	0	0	0	0
	I	1	2	0	0	0	0	0	0
	C	7	1	1	0	2	1	1	3
	Rec	1	0	0	0	0	0	0	0
	Ch	0	1	0	0	0	0	0	0
	O	2	3	0	1	1	0	0	2

Fig. 3 Number of participants using each strategy in (a) immediate serial recall compared to delayed serial recall; (b) item recognition compared to free recall; and (c) immediate serial recall compared to complex WM span. R = Recall; G = Grouping; S = Semantics; I = Image; C = Concentrate; Rec = Recent; Ch = checklist; O = Other.

strategies across tasks. Indeed, the degree of individual variation is comparable in those cases to that shown in Fig. 3c, which displays the correspondence between immediate serial recall and complex WM span, which were significantly different in overall strategy distribution.

Strategy Use Profiles

The lack of within-individual consistency, especially for tasks with similar overall strategy distributions, was intriguing. To

further probe individual strategy variation, we next turned to a community detection algorithm based on graph theory. This data-driven approach explored the possibility that there are naturally occurring groups of individuals who vary their strategies the same way under changing task demands. The algorithm we applied required a complete strategy profile for each individual (e.g., would not accommodate participants with missing strategy data on one or more tasks) and fails for cases where the identical strategy is reported for every task. Accordingly, participants with any missing data ($N = 55$) and those who made consistent use of only one strategy ($N = 12$) were excluded from these analyses. Also, we excluded missing item strategy data from these analyses in order to focus on variation in strategy use within more proximate tasks. As in the analyses above, we maintained the strategy grouping of None and Concentrate, and excluded participants who did not understand the task. For this analysis, we also added the Checklist strategy to the Other category, as it was very sparsely reported in tasks other than the missing item task. After these exclusions, a total of 153 participants remained. Strategy data from these participants was submitted to network analysis, which revealed a highly modular and robust network with three distinct subgroups ($Q = .58$; threshold = .55).

To better understand the results of this grouping algorithm, and to explore what factors may have driven the robust groupings, we examined how participants in each group employed different strategies across tasks. This examination revealed striking differences in strategy use between groups, as shown in Fig. 4 (the Semantics and Imagery strategies had almost no variation by group, and for simplicity, these strategies are not shown in the figure). These groupings were driven primarily by the tasks for which Rehearsal (the most-reported strategy across all groups and tasks) was *less often* reported, as well as by the predominating alternate strategies for those who did not employ Rehearsal. For instance, no participants in Group 1 employed Rehearsal in the running memory span task; instead, they generally reported Concentrate or Other for that task. Meanwhile, fewer members of Group 2 reported using Rehearsal during item recognition; both relative to their use of this strategy on other tasks and compared to the use of rehearsal by the other two groups during item recognition. As Group 1 did for running memory span, Group 2 tended to report Concentrate or Other as the primary alternative strategy during item recognition. Group 3 were less frequent “rehearsers” during free recall, during which they tended to make greater use of the Familiarity, Other, and Concentrate strategies.

The correlation heat map presented in Table 4 provides another useful visualization for understanding the subtyping analysis. It is clear from this table that Group 1 differs most greatly from Groups 2 and 3 on Running Memory Span, while Group 2 differs on Item Recognition and Group 3 differs on Free Recall.

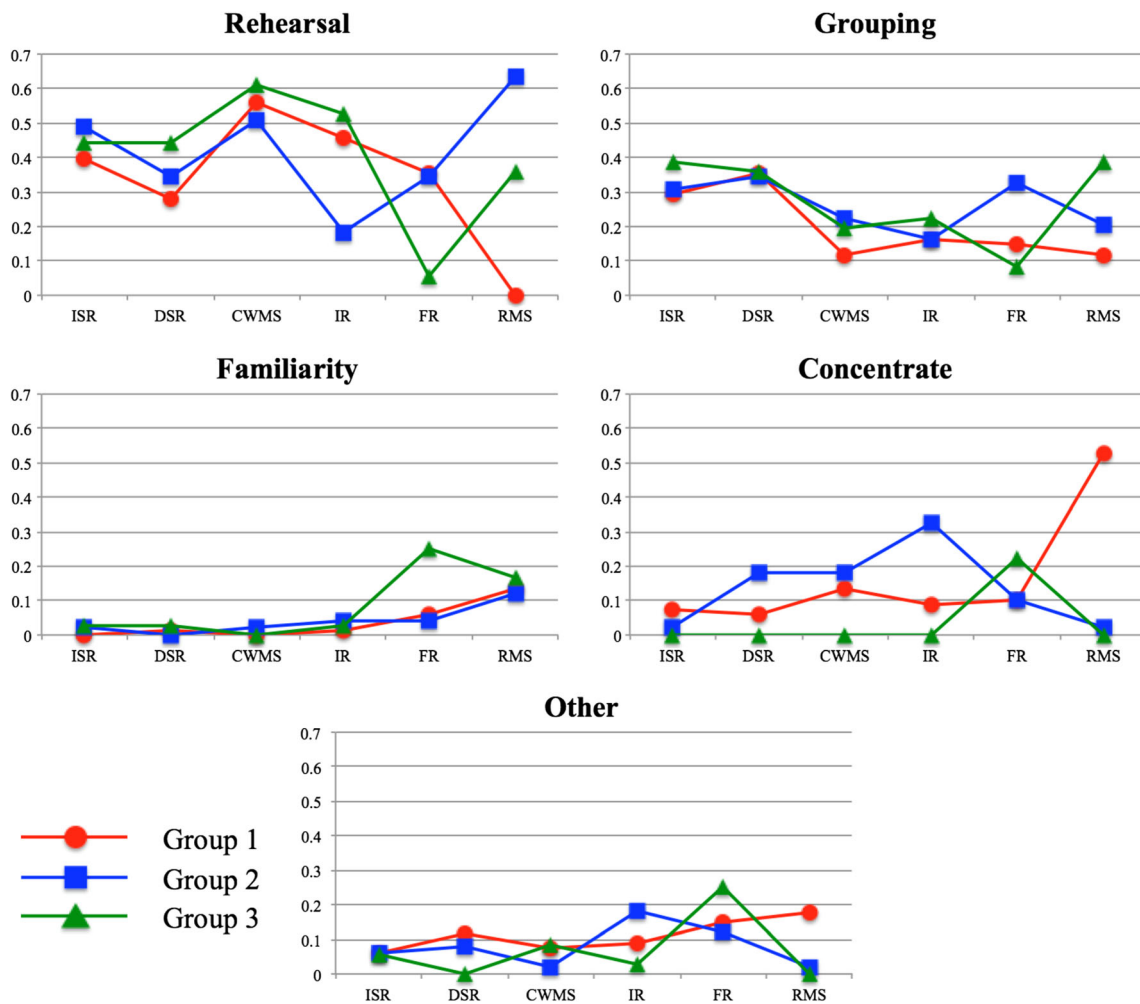


Fig. 4 Charts displaying the relative proportion of reported strategy use across tasks and divided by subtype (Group). Group 1: *N* = 60; Group 2: *N* = 32; Group 3: *N* = 53. Note. ISR = immediate serial recall; DSR =

delayed serial recall; CWMS = complex working memory span; IR = item recognition; FR = free recall; RMS = running memory span.

Discussion

Various factors drive investigators to make use of a given WM task—the prevalence of use in the prior literature associated with the phenomenon under investigation, emphasis on specific aspects of WM (e.g., executive vs. domain-specific storage), the degree to which the task is amenable to experimental constraints or suited to the population of interest, the amount of time it takes to obtain reliable data from the task, and so forth. These factors all relate to the appropriateness of the paradigm to the research question under investigation. However, given the very wide use

of WM assessment tasks, surprisingly little is known about how specific task demands actually affect the strategies that participants choose. Without knowing about how task properties affect strategy use, it is difficult (if not impossible) to actually determine the appropriateness of a given paradigm.

The broad goal of the present study was to initiate a more extensive investigation of the relationship between WM tasks and strategy use. One of our specific aims was to simply characterize the range of strategies employed during tasks common to the WM literature. We investigated the hypothesis that even modest changes in task parameters (e.g., recall vs.

Table 4 Correlations in strategy use between subtyping groups for each of the tasks.

	ISR	DSR	CWMS	IR	FR	RMS
1v2	.98	.90	.95	.36	.83	-.26
1v3	.97	.90	.96	.97	-.07	-.27
2v3	.98	.85	.91	.17	-.12	.81

Note. Darker cells represent negative or weak positive correlations while lighter cells represent strong positive correlations. ISR = immediate serial recall; DSR = delayed serial recall; CWMS = complex working memory span; IR = item recognition; FR = free recall; RMS = running memory span.

recognition) could lead to different strategy allocation by participants. Our analyses revealed that strategy distributions differed significantly across most, but not all, tasks. Further analyses showed that some parameters (an ordering demand, updating requirement, or distraction) appear to affect the distribution of strategy use, while others (presence/absence of a delay, length of the memory list) do not, at least within the boundary conditions that we investigated.

The finding that strategy use is heterogeneous, even within a given WM task, is consistent with the extant literature on strategy use during WM tasks (Logie et al., 1996; Hertzog et al., 1998). However, while others have found performance to be contingent on the choice of strategy (Bailey et al., 2011; Dunlosky & Kane, 2007; Friedman & Miyake, 2004), we did not find that performance within any of the tasks varied as a function of strategy use. The lack of a robust influence of strategy on performance may be related to the small number of trials that were included per task in the present study. While we prioritized inclusion of several tasks over inclusion of a larger number of trials per task, in future studies inclusion of a greater number of trials per task may increase the power to replicate the performance differences observed in earlier studies.

The heterogeneity in strategy choice that we observed has important implications when trying to draw inferences across studies using disparate WM tasks. Sometimes assuming common underlying strategies may be justified, but other times these assumptions may not be valid. For instance, the present results endorse the conclusion that participants employ a similar range of strategies in an immediate serial recall task and a delayed serial recall task, and, accordingly, make a stronger case for directly comparing performance obtained from these two tasks. In contrast, it seems that ordering demands are a strong determinant of strategy choice, and hence, comparisons between immediate serial recall and item recognition, for example, may not be well justified.

Equally profound are the implications of these findings as they pertain to studies using neuro-investigative (e.g., fMRI, EEG) methods. The implications are twofold. First, if different WM tasks are associated with unique strategy distributions, and by inference, unique neural substrates, then researchers need to use caution in comparing activation patterns evoked from a given task to those evoked from a different task, despite both being putative measures of WM. Second, if participants in a study sample use heterogeneous strategies in a single task, then the neural processes they engage while completing that task should also be expected to be heterogeneous, thus adding substantial noise to already noisy methods. As a safeguard of sorts, researchers may want to probe their participants to understand which strategies they are using and examine group differences in brain activation based on strategy use, provided that the sample is large enough.

Our subtyping analysis also affords novel insights. This analysis suggests that within any study sample it is likely that

robust groups of participants will approach the same tasks differently. These findings are particularly interesting because they expose particular contexts in which participants will stray from the predominating rehearsal strategy. For example, based on the present results, it can be expected that a subgroup of participants will approach free recall differently from the majority of other participants, responding based on what is familiar or by concentrating on the words rather than engaging in effortful rehearsal. Once again, this is an important finding, not only with respect to behavioral research but also for neuro-investigative studies. For instance, we might expect that activation profiles will be similar for members of the same strategy group but disparate for members of opposing groups.

To take the potential implications one step further, the great degree of heterogeneity we observe in strategy selection—across task demands, within individuals (across tasks), and between groups of individuals—raises questions about the very existence of a common underlying WM “system” or resource. Indeed, the results lend some support to an emerging perspective that abandons the notion that WM functions as a singular cognitive system that is recruited whole cloth in support of performance and assumes instead that the specific processes marshaled in a given task environment are entirely a function of the particular materials that need to be processed, the exact storage and output requirements of the task, and the particular skills/knowledge possessed by the participant (e.g., Macken, Taylor, & Jones, 2015).

While the present experiment provides an important step toward understanding the strategies that participants use in various WM tasks, there are several limitations that need to be acknowledged. Foremost, in designing our tasks we needed to control task parameters so that (1) tasks would be easy to administer to large groups of participants and (2) we could make comparisons across tasks that differed according to specific parameters. In some cases, this led the tasks to diverge slightly from similar tasks used in past literature. This was true for the free recall task, which in atypical fashion used repeated presentations of the same pool of items, and for the delayed serial recall task, which used a larger stimulus set (9 items) than is the mode. A particularly relevant deviation, we think, occurred with the complex WM span task. Complex WM span tasks are usually subject-paced, with new memory items given directly after the solution to the intervening math problem is given. However, our operation span task could not be subject-paced because we administered the task to groups of participants who may have varied in their speed at solving math problems. We therefore opted to allow participants a fixed amount of time to solve the math problem. From the perspective of understanding strategy selection, this could have been problematic because those who solved the math problems more quickly (i.e., early in the allotted time period) would have had extra time to engage in active strategy

selection, and conversely, participants who did not have enough time to solve the math problem may have been forced to adopt passive strategies. Future studies testing participants individually may benefit from using subject-specific timing for operation span and other tasks.

Yet another issue of note is related to the stimulus sets we deployed. We used partially “open” pools of stimuli (i.e., we varied the specific subset of words presented from one trial to the next within a task). This practice can be contrasted with studies using “closed” sets of items (i.e., using the same limited set of words on each trial of a task) to isolate short-term processes from semantic and long-term memory phenomena (see Baddeley, 2012). While the use of open versus closed sets might impact strategy choice in some contexts, our chi-square analysis (see Table 2) showed that strategy distributions for tasks where the stimulus pool was more open (tasks where only 6 items out of 10 possible items were sampled across trials; ISR, IR) did not significantly differ from those using a more closed set (tasks where 9 of 10 possible items were sampled across trials; DSR, FR). Hence, where evidence is available in the present study the patterns don’t indicate that variation in strategy choice was a function of the degree to which stimulus sets were open or closed.

Another limitation of the present study stems from reliance on a self-reported strategy questionnaire. Despite the detailed description of strategies that we provided, it is likely that there was variability in participants’ interpretations of the subjective and introspective experiences associated with a given strategy, and in their interpretations of the specific language of the questionnaire in relation to those subjective experiences. Since we only have their final self-reports, we cannot readily assess the degree to which subjects reliably understood the described strategies and whether they uniformly associated these descriptions with the strategic approaches we intended to index. Nevertheless, the higher rate of use of the Checklist strategy in association with the missing item task does provide some evidence that participants were indeed trying to make appropriate use of the questionnaire. A somewhat related concern is that our strategy labels (and descriptions) may have been too coarse to capture nuanced, but important, differences within strategic categories. For example, Lehmann and Hasselhorn (2007) have made the argument that within the broad category of “rehearsal” there may be a distinction between “labeling” (merely saying the word), “single-word rehearsal” (rehearsing each word one at a time), and “cumulative rehearsal” (rehearsing at least two words in succession). As rehearsal was our most commonly reported strategy, it is possible that participants were actually engaging in different types of rehearsal. These distinctions may have been helpful in understanding the strategies employed by individuals, especially in the subtyping analysis.

Additionally, it is important to point out that participants may have been biased in their reporting after seeing the

strategy list for the first time. Indeed, having been exposed to the questionnaire after completion of the first task may have encouraged participants to select strategies that they would not have considered had they not seen the option on the questionnaire. Yet another possible source of bias in reporting may have emerged because participants completed several WM tasks within a single session, and experience with the demands of one task may have influenced subsequent tasks. These sources of bias were partially attenuated by counterbalancing task order. To further evaluate concerns about how the chronology of the tasks within a session might influence strategy reports, for each task separately, we compared the distribution of reported strategies in the subsample of participants that started the experiment with a given task to the distribution in those who completed the same task later in the session. We did not find robust differences within any of the tasks according to the task’s relative position in the session, making it somewhat less likely that factors associated with the order of tasks within a session influenced the strategy reports.

Some additional caveats we should acknowledge are tied to the use of chi-square tests and the assumptions associated with such tests. Chi-square tests assume that frequencies across cells are independent of one another, and this assumption was not met in the analyses that considered more than one task per participant (because a given participant’s responses were included in multiple cells). Luckily, the fact that strategy distributions were similar for those participants who completed a given task at the start of the session (with no violation of the independence assumption) to the entire session data suggests that the violation of this assumption did not strongly impact the outcomes. However, analysis of only the data from the first task encountered dramatically reduces statistical power and reduces the expected count of several cells below five occurrences, which violates another common rule in application of the chi-square statistic. Therefore, in the present study, it is difficult to accommodate the assumption of independence while also meeting guidelines for expected cell counts. While the chi-square results should be interpreted with these limitations in mind, concerns regarding these assumptions are, to some extent, lessened by the addition of the subtyping analyses, which also supported the conclusion that strategy differs between individuals in the same task and within an individual between tasks.

While there are some clear caveats to be considered, this is the first study to explicitly probe strategy use across a variety of WM tasks and to describe some differences in strategy use across tasks and across individuals. We hope that these data will serve as a useful reference to those wishing to understand and predict strategy variation across these tasks, and that the findings will also encourage more in depth contemplation of the implications of differences in strategy use across WM tasks. Ultimately, the findings suggest a need to be more circumspect in the selection of particular WM tasks for use in

research, more thoughtful in the interpretation of results emerging from studies using these tasks, and more careful in drawing comparisons across studies using different tasks.

Acknowledgements This project was supported in part by funding from the National Institute on Drug Abuse, NIH (DA031436) and the Office of Naval Research (ONR N00014-11-1-0399). We would like to thank the many research assistants and students who assisted with execution of the study, especially Naomi Streeter, Danielle Adinolfi, and Suwathna Reel.

References

- Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4(10), 829–839. doi:10.1038/Nrn1201
- Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63, 1–29. doi:10.1146/annurev-psych-120710-100422
- Baddeley, A., & Hitch, G. J. (1974). Working memory. *The Psychology of Learning and Motivation*, 8, 47–89.
- Bailey, H., Dunlosky, J., & Kane, M. J. (2008). Why does working memory span predict complex cognition? Testing the strategy affordance hypothesis. *Memory & Cognition*, 36(8), 1383–1390. doi:10.3758/Mc.36.8.1383
- Bailey, H., Dunlosky, J., & Kane, M. J. (2011). Contribution of strategy use to performance on complex and simple span tasks. *Memory & Cognition*, 39(3), 447–461. doi:10.3758/s13421-010-0034-3
- Beaman, C. P., & Jones, D. M. (1997). Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology Learning Memory and Cognition*, 23(2), 459–471. doi:10.1037/0278-7393.23.2.459
- Beaman, C. P., & Jones, D. M. (1998). Irrelevant sound disrupts order information in free recall as in serial recall. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 51(3), 615–636.
- Bor, D., Cumming, N., Scott, C. E., & Owen, A. M. (2004). Prefrontal cortical involvement in verbal encoding strategies. *European Journal of Neuroscience*, 19(12), 3365–3370. doi:10.1111/j.1460-9568.2004.03438.x
- Bor, D., & Owen, A. M. (2007). A common prefrontal-parietal network for mnemonic and mathematical recoding strategies within working memory. *Cerebral Cortex*, 17(4), 778–786. doi:10.1093/cercor/bhk035
- Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods*, 42(2), 563–570. doi:10.3758/Brm.42.2.563
- Bunting, M., Cowan, N., & Saults, J. S. (2006). How does running memory span work? *Quarterly Journal of Experimental Psychology*, 59(10), 1691–1700. doi:10.1080/17470210600848402
- Buschke, H. (1963). Relative retention in immediate memory determined by the missing scan method. *Nature*, 200, 1129–1130.
- Carretti, B., Borella, E., & De Beni, R. (2007). Does strategic memory training improve the working memory performance of younger and older adults? *Experimental Psychology*, 54(4), 311–320. doi:10.1027/1618-3169.54.4.311
- Chein, J. M. (2008). Data from the automated operation span task that was collected in ongoing projects. Unpublished raw data.
- Chein, J. M., & Fiez, J. A. (2001). Dissociation of verbal working memory system components using a delayed serial recall task. *Cerebral Cortex*, 11(11), 1003–1014. doi:10.1093/cercor/11.11.1003
- Chein, J. M., Moore, A. B., & Conway, A. R. A. (2011). Domain-general mechanisms of complex working memory span. *NeuroImage*, 54(1), 550–559. doi:10.1016/j.neuroimage.2010.07.067
- Chein, 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. doi:10.3758/Pbr.17.2.193
- Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769–786.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavior Brain Science*, 24(1), 87–114. discussion 114–185.
- Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, 51(1), 42–100. doi:10.1016/j.cogpsych.2004.12.001
- Daneman, M., & Carpenter, P. A. (1980). Individual-differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450–466. doi:10.1016/S0022-5371(80)90312-6
- D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, 66, 115–142. doi:10.1146/annurev-psych-010814-015031
- Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *Quarterly Journal of Experimental Psychology*, 60(9), 1227–1245. doi:10.1080/17470210600926075
- Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual-differences in working memory and comprehension: A test of 4 hypotheses. *Journal of Experimental Psychology Learning Memory and Cognition*, 18(5), 972–992. doi:10.1037/0278-7393.18.5.972
- Engle, R. W., Tuoholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128(3), 309–331.
- Farrell, S. (2006). Mixed-list phonological similarity effects in delayed serial recall. *Journal of Memory and Language*, 55(4), 587–600. doi:10.1016/j.jml.2006.06.002
- Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, 51(1), 136–158. doi:10.1016/j.jml.2004.03.008
- Henson, R. N. A., Shallice, T., Gorno-Tempini, M. L., & Dolan, R. J. (2002). Face repetition effects in implicit and explicit memory tests as measured by fMRI. *Cerebral Cortex*, 12(2), 178–186. doi:10.1093/cercor/12.2.178
- Hertzog, C., McGuire, C. L., & Lineweaver, T. T. (1998). Aging, attributions, perceived control, and strategy use in a free recall task. *Aging Neuropsychology and Cognition*, 5(2), 85–106. doi:10.1076/ane.5.2.85.601
- 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. doi:10.1037/a0034221
- Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology Learning Memory and Cognition*, 19(2), 369–381. doi:10.1037/0278-7393.19.2.369
- Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, 59, 193–224. doi:10.1146/annurev.psych.59.103006.093615

- Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*(1), 47–70. doi:10.1037/0096-3445.132.1.47
- 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. doi:10.1037/0096-3445.133.2.189
- Lehmann, M., & Hasselhorn, M. (2007). Variable memory strategy use in children's adaptive intratask learning behavior: Developmental changes and working memory influences in free recall. *Child Development*, *78*(4), 1068–1082. doi:10.1111/j.1467-8624.2007.01053.x
- Logie, R. H. (2003). Spatial and visual working memory: A mental workspace. *Psychology of Learning and Motivation: Advances in Research and Theory: Cognitive Vision*, *42*, 37–78.
- Logie, R. H., DellaSala, S., Laiacona, M., Chalmers, P., & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory & Cognition*, *24*(3), 305–321. doi:10.3758/Bf03213295
- Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. doi:10.1016/j.tics.2013.06.006
- Macken, B., Taylor, J., & Jones, D. (2015). Limitless capacity: A dynamic object-oriented approach to short-term memory. *Frontiers in Psychology*, *6*, 293. doi:10.3389/fpsyg.2015.00293
- Marsh, J. E., Sorqvist, P., Hodgetts, H. M., Beaman, C. P., & Jones, D. M. (2015). Distraction control processes in free recall: Benefits and costs to performance. *Journal of Experimental Psychology Learning Memory and Cognition*, *41*(1), 118–133. doi:10.1037/A0037779
- McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition*, *29*(1), 10–17.
- Morrison, A. B., Conway, A. R. A., & Chein, J. M. (2014). Primacy and recency effects as indices of the focus of attention. *Frontiers in Human Neuroscience*, *8*. doi:10.3389/Fnhum.2014.00006
- Nee, D. E., Brown, J. W., Askren, M. K., Berman, M. G., Demiralp, E., Krawitz, A., & Jonides, J. (2013). A meta-analysis of executive components of working memory. *Cerebral Cortex*, *23*(2), 264–282. doi:10.1093/cercor/bhs007
- Neely, C. B., & LeCompte, D. C. (1999). The importance of semantic similarity to the irrelevant speech effect. *Memory & Cognition*, *27*(1), 37–44. doi:10.3758/Bf03201211
- Oberauer, K. (2003). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language*, *49*(4), 469–483. doi:10.1016/S0749-596x(03)00080-9
- Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review*, *115*(3), 544–576. doi:10.1037/0033-295X.115.3.544
- 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. doi:10.3758/s13423-012-0272-4
- Pollack, I., Johnson, L. B., & Knaff, P. R. (1959). Running memory span. *Journal of Experimental Psychology*, *57*(3), 137–146. doi:10.1037/H0046137
- Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, *28*(3), 164–171. doi:10.1027/15-5759/A000123
- Richardson, J. T. E. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review*, *5*(4), 597–614. doi:10.3758/Bf03208837
- Ricker, T. J., & Cowan, N. (2014). Differences between presentation methods in working memory procedures: A matter of working memory consolidation. *Journal of Experimental Psychology Learning Memory and Cognition*, *40*(2), 417–428. doi:10.1037/A0034301
- Rubinow, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, *52*(3), 1059–1069. doi:10.1016/j.neuroimage.2009.10.003
- St Clair-Thompson, H., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology*, *30*(2), 203–219. doi:10.1080/01443410903509259
- Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language*, *49*(4), 446–468. doi:10.1016/S0749-596x(03)00095-0
- Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26. doi:10.1016/j.cogpsych.2014.01.003
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505.
- Unsworth, N., Spillers, G. J., & Brewer, G. A. (2010). The contributions of primary and secondary memory to working memory capacity: An individual differences analysis of immediate free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *36*(1), 240–247. doi:10.1037/a0017739
- Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: A meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(4), 255–274. doi:10.3758/Cabn.3.4.255
- Wilson, M. D. (1988). The MRC Psycholinguistic Database: Machine readable dictionary. *Behavioural Research Methods, Instruments and Computers*, *20*, 6–11.
- Zanto, T. P., Clapp, W. C., Rubens, M. T., Karlsson, J., & Gazzaley, A. (2016). Expectations of task demands dissociate working memory and long-term memory systems. *Cerebral Cortex*, *26*(3), 1176–1186.