The contributions of strategy use to working memory span: A comparison of strategy assessment methods

By: John Dunlosky, Michael J. Kane

*** Note: Figures may be missing from this format of the document

Abstract:
In two experiments, we tested whether individual differences in strategy production account for individual differences in performance on a working memory span task. We measured the strategies used during a standard experimenter-paced operation span (OSPAN) task by having participants make both set-by-set reports of strategy use for individual item sets and global reports of strategy use. In Experiment 1, although normatively effective strategies were self-reported on only a small proportion of OSPAN sets, individual differences in effective strategy use correlated with span performance. Experiment 2 replicated this outcome using a sample of 100 participants but, as important, it demonstrated that individual differences in effective strategy use did not mediate the relationship between OSPAN and measures of verbal ability. Discussion focuses on the interpretation of strategy–span relationships and the relative utility of general reports of strategy use versus the set-by-set reports introduced here for the OSPAN task.

Article:

What does performance on a working memory (WM) span task measure, and why do individual differences in span often correlate with measures of higher order cognition, such as reading comprehension and inductive reasoning? These questions have been central to two decades of theoretical and applied research about working memory. Answers arise from prominent theories of WM, most of which conceptualize individual differences in WM span as reflecting a general process or mechanism that influences the efficiency of the WM system. For instance, Hasher and colleagues argue that individual differences in span reflect the degree to which people inhibit irrelevant, or no-longer-relevant, information from WM when performing the task (e.g., Lustig, May, & Hasher, 2001; May, Hasher, & Kane, 1999). In contrast, Engle and his colleagues suggest that variability in span performance derives from individual differences in attention control capabilities that are critical to both active maintenance and inhibition (e.g., Engle & Kane, 2004; Engle, Kane, & Tuholski, 1999). Other general-process theories of WM are available (Miyake & Shah, 1999; see also Cowan, 2005; Oberauer, 2005), but importantly, they all emphasize that WM capacity differences—as measured by WM span—largely reflect the integrity of general processes or mechanisms that are central to the WM system (Conway, Kane, & Engle, 2003). Thus, the Zeitgeist of WM theory is echoed in Earl Hunt's (1999) observation that "the concept of memory span is clearly a system architecture concept" (p. 29).

Our aim is not to competitively evaluate general theories of WM architecture or mechanism. Instead, we critically examine an alternative class of hypotheses that has recently been forwarded. According to this alternative, differences in span performance are assumed to result from variability in the algorithms, or strategies, that individuals use while performing span tasks; moreover, the power of WM span tasks to predict variation in complex cognitive abilities may also derive from individual differences in strategy use (e.g., Baddeley, 2000; McNamara & Scott, 2001; Salthouse, 1996).

To illustrate how algorithmic hypotheses differ from architectural hypotheses, consider how they might explain individual differences in the operation span (OSPAN) task, a frequently used measure of WM capacity. In a typical administration of OSPAN, a mathematical operation and a word appear on a computer screen—for example, Is (9/3) – 2 = 2? *aunt*—and a participant must read the problem aloud—"Is (9/3) – 2 = 2?"—then

verify the answer as quickly and accurately as possible ("no"), and then immediately read the word aloud ("aunt"). At that time, the experimenter quickly advances the screen to the next operation–word pair. These operation– word pairs are individually presented in sets that usually range from 2 to 6 pairs per set. After the last operation–word pair of a set is presented, the participant is prompted to recall the words from that set in the order that they had appeared. Although participants need to process the operations quickly and accurately, they are instructed that their primary goal is to remember the words in serial order. Performance on this span task can be measured in a variety of ways, but it is almost always based on the number of words correctly recalled (Conway et al., 2005; but see Waters & Caplan, 1996).

The substantial individual differences that arise in WM span are readily explained in terms of variability in some relatively general mechanism but, in principle, they can be equally well accounted for by differences in strategic behaviour. Such an algorithmic, strategy mediation hypothesis explains individual differences in OSPAN in terms of the strategies that people use to learn and recall the word sets. Some strategies are likely to be relatively ineffective, such as passively reading each word, whereas others are likely to be more effective, such as using interactive imagery or sentences to associate the words within a set. By the strategy mediation hypothesis, individuals who perform best in span tasks are those who use more effective strategies (e.g., McNamara & Scott, 2001).

The present article describes the outcomes of two experiments involving the OSPAN task that empirically evaluate a core prediction of the strategy mediation hypothesis—namely, that individual differences in strategy use predict span performance. In doing so, we introduce a new measure of strategy use for the OSPAN task based on set-by-set reports and also compare it to a traditional measure of strategy use based on general retrospective reports.

Review of research relevant to the strategy mediation hypothesis
At least two lines of research bear on the strategy mediation hypothesis and motivated the approach that we adopted to test it. First, some studies have investigated whether instructions to use a particular strategy, such as imagery or rehearsal, can influence span (e.g., McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003). For instance, Turley-Ames and Whitfield compared performance on the OSPAN task immediately prior to, and following, strategy instruction. Strategy-instructed participants were told to use one of several mnemonics to remember the target words, such as repeatedly rehearsing the words aloud, creating a sentence or story from the words, or generating visual images of the words. Such strategy instructions, as well as more intensive training in the use of a particular strategy (McNamara & Scott, 2001), significantly boosted span scores.

A second line of work used self-paced viewing times to infer the contribution of strategy use to span performance. Engle, Cantor, and Carullo (1992) and Friedman and Miyake (2004) examined the time that participants allocated to various components of span tasks, such as viewing each to-be-remembered word, as an indication of the ways in which participants strategically approached the task. For instance, Engle et al. (1992, Exp. 1) allowed participants unlimited time to examine the different components of each operation–word pair (with these components presented via a self-paced, moving-window technique) in OSPAN. Viewing time on the to-be-remembered words correlated significantly with span performance (r = .29), suggesting that high-WM individuals read (or studied, or rehearsed) the target words for more time than did low-WM individuals. Thus, high spans appeared more strategic than did low spans when performing the OSPAN task. One limitation of this conclusion is that viewing times are an indirect measure of strategy production and may be influenced by a host of other factors that, in principle, could account for the aforementioned results (McNamara & Scott, 2001). Moreover, time spent on a task will only reflect strategic behaviour to the degree that different strategies result in different time signatures, which may not be the case for many mediational strategies, such as imagery, sentence generation, and rehearsal, which participants may use when attempting to maintain the to-be-remembered words. Thus, although research on viewing times has led to important insight into WM theory, this research does not directly indicate how often mediational strategies are used during standard WM span tasks.

Our goal was to evaluate directly whether strategy use influences WM span performance when the span task is conducted using a standard procedure that involves no strategy training and minimizes strategy use by allowing participants only limited time to read each to-be-remembered word (in contrast to the self-paced, moving-window technique). An assessment of strategy use in previous studies typically has involved using working memory tasks that increase the likelihood of strategy use, in contrast to the standard tasks often used in the literature to provide a more process-pure measure of WM ability. Accordingly, previous research aimed at evaluating the strategy mediation hypothesis may not generalize to the standard preparations used by the field, a limitation that we sought to circumvent in the present research. In particular, rather than allowing participants to pace themselves through the task, in our studies (as in most) the experimenter controlled the stimulus presentation. Because the experimenter-driven procedure presumably minimizes strategic behaviour (e.g., see Friedman & Miyake, 2004, Appendix), it arguably yields the most conservative test of the strategy mediation hypothesis, and the one most relevant to research using the standard, experimenter-controlled procedure.[1]

Of course, to estimate the degree to which strategies affect span scores in experimenter-paced procedures, we must obtain a measure of strategy use other than reading time. One obvious candidate is to simply ask participants, retrospectively, about the strategies that they had used to accomplish the task. This is precisely how prior studies on WM span have assessed strategy use (e.g., Friedman & Miyake, 2004; Harris & Qualls, 2000; McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003). After all trials of the span task were completed, participants explained in general how they tried to complete the task. These general retrospective reports have yielded some support for the strategy mediation hypothesis. For instance, Turley (1997, cited in Turley-Ames & Whitfield, 2003) found that many participants reported rehearsing the to-be-remember words in the OSPAN task, with more high-span than low-span participants reporting such strategy use (69% vs. 24%, respectively). Friedman and Miyake (2004) found, in their experimenter-controlled reading span task, that 59% of their participants reported using some phonological strategy at some point during a reading span task, whereas 45% and 18% reported using a semantic-based and imagery-based strategy, respectively. Moreover, participants who used the imagery-based strategy earned higher span scores than did those who did not.

One interpretation of these outcomes is that strategy use accounts for some of the individual differences in WM span, which supports the strategy mediation hypothesis. Even so, conclusions based on retrospective general reports are open to alternative interpretations. First, given that the reports are general, participants may develop the report in ways that do not accurately reflect how the task was actually performed (for detailed discussion, see Ericsson & Simon, 1984). For example, participants may draw on beliefs about how they ought to have performed the task, or they may remember how they performed a few trials of the task and generalize these particular instances to the remaining trials. Second, given that the reports are retrospective, participants may simply forget how they performed many of the individual trials. Of utmost importance here, individual differences may occur in any of these processes or biases, and they may or may not be associated with WM capacity. For instance, if low-span individuals are particularly susceptible to forgetting prior to making the general retrospective reports (see Kane & Engle, 2000), then any span/strategy correlation may not reflect true differences in strategy production per se. Retrospective general reports therefore do not provide definitive evidence about the strategy mediation hypothesis, namely that strategy use accounts for variation in WM span.

Current approach to evaluating the strategy mediation hypothesis
Our approach involved having participants make set-by-set reports of strategy use for the OSPAN task. In particular, sometime after a participant attempted to recall the to-be-remembered words of a given set, he or she reported the particular strategy—if any—used to remember the words for that set. This report was based on a specific probe that assessed the strategies that pilot work and prior research (e.g., Turley-Ames & Whitfield, 2003) suggested are sometimes used during the OSPAN task.

Prompt used to obtain set-by-set strategy reports. For concurrent reports (Experiment 1), the prompt for the report was presented immediately after a participant attempted recall for the words within a set. For retrospective reports (Experiments 1 and 2), all sets were presented for the OSPAN task first. Afterwards, each set was individually presented in its entirety above this prompt for the strategy report.

How did you originally try to remember the words from the series above?

State the number corresponding to the most appropriate response:

1 = read each word as it appeared
2 = repeated the words as much as possible
3 = used a sentence to link the words together
4 = developed mental images of the words
5 = grouped the words in a meaningful way
6 = did something else

These set-by-set reports demonstrate acceptable validity and have minimal reactive effects on task processing when they are used to measure strategy use in related domains, such as in associative learning and free recall (Dunlosky & Hertzog, 2001; Hertzog, McGuire, & Lineweaver, 1998). Even so, given that the present research is the first to adapt these reports to WM span, we report evidence relevant to their key properties in Experiment 1.[2] As important, set-by-set reports of strategy use are superior to general reports in numerous ways (Ericsson & Simon, 1980), but here we consider only one. Imagine that in response to a general retrospective prompt, two participants report that they used rehearsal and interactive imagery to complete the OSPAN task. However, one participant actually used imagery on 80% of the sets, whereas the other did on only 20%. This critical difference would not be reflected in the general reports, but it would be in set-by-set reports.[3] In this case, set-by-set reports would provide a more accurate description of strategy use at the level of individual sets, which itself is necessary for conducting fine-grained analyses conditionalized on set. Note, however, that our aim is not to dismiss entirely the potential benefit of general reports, which we collected in both experiments. Doing so not only afforded tests of the main prediction based on converging operations, but also allowed us to directly compare the two kinds of strategy report. To foreshadow, our results supported conclusions from earlier research that had exclusively used general reports of strategy use, but they also demonstrated some of the superior characteristics of set-by-set reports.

In summary, we evaluated a prediction of the strategy mediation hypothesis by examining the relations among individual differences in the measures of strategy use and in OSPAN performance. If strategy use accounts for individual differences in span performance, then the correlation between the use of effective strategies and span performance will be positive.

EXPERIMENT 1

Given that set-by-set strategy reports are new to investigations of span tasks, we used both concur-rent reports and retrospective reports. The former were made immediately after a participant attempted to recall the target words of each set, whereas the latter were made after a participant had completed all the sets of the OSPAN task. Thus, both general reports and retrospective set-by-set reports (hereafter, referred to as retrospective reports) occurred after the OSPAN task had been completed, but the former asked for a general report of strategy use across all sets whereas the latter involved reporting the strategy used for each set individually. To prompt the retrospective reports, each original set was re-presented individually, with all the equations and words from the set presented together. By using both kinds of set-by-set report, we could deter-mine (a) whether they demonstrated acceptable validity, (b) whether concurrent reports had reactive effects on task performance, and (c) whether retrospective reports accurately captured the strategies that participants reported using at study (i.e., whether retrospective reports suffered from forgetting).

To address these issues, we used methods from Dunlosky and Hertzog (2001), who demonstrated the validity of concurrent and retrospective strategy reports during associative learning. In Experiment 1, two groups of participants performed the standard OSPAN task, and both made retrospective strategy reports. The *retrospective-only* group made only retrospective reports, whereas the *concurrent + retro* group made both kinds of strategy report.

If the reports have construct validity, then OSPAN performance will be better for trials in which participants reported using normatively more effective strategies (e.g., imagery) than normatively less effective strategies (e.g., reading). If concurrent reports have reactive effects by influencing the use of strategies, then the profile of retrospective reports will differ for the concurrent + retro group and retrospective-only group. Finally, if the validity of retrospective reports is reduced due to forgetting, then the profile of reported strategies will differ markedly when the reports are made concurrently versus retrospectively. Importantly, we found that the reports demonstrated acceptable validity and were quite consistent for concurrent and retrospective reports. Thus, to achieve greater statistical power for analyses relevant to the strategy mediation hypothesis, we collapsed across the two groups and analysed retrospective reports where appropriate.

Method

Participants and materials

A total of 70 undergraduates from introductory psychology courses at the University of North Carolina at Greensboro (UNCG) participated to complete a course requirement. A total of 35 participated in each strategy report group. Macintosh computers were used to present trials during the OSPAN task and to record strategy reports. For recall during each trial of the span task, participants wrote the target words on a sheet of paper.

Procedure

Participants performed the span task individually. An experimenter sat in the room to pace the task and to record the participants' answers to the equations on the span task. Participants were first given detailed instructions about the span task.

We used the version of OSPAN described by Kane et al. (2004). Each trial presented an equation and word on-screen, and the participant read the equation aloud, verified whether or not the provided answer was correct, and then immediately read the word aloud. Participants were instructed that they could take as much time as they needed to accurately verify each equation, but they were to begin reading each equation aloud as soon as it appeared. They were also required to read the target word aloud immediately upon verifying the equation aloud. As soon as the participant named the word aloud, the experimenter pressed a key to advance to either the next set or the recall cue (which was a display of three question marks centred on-screen). In each set, the participant saw 2–5 equation–word pairs, and the sets were presented in the same initially randomized sequence to all participants (with three trials at each set size, and three 2-item sets for practice). Thus, the participant was unaware of the current set size until the recall cue was presented. At the recall cue, the participant attempted to write down all of the words from the current set in correct serial order. Every third mathematical error that was committed by the participant (if any) was followed by the experimenter reminding the participant, after recall, to be more accurate in responding to the equations. The OSPAN task was scored by calculating the percentage of items in each set that were recalled in correct serial position and then averaging these percentages across sets. This scoring procedure yields more normal distributions and better reliability than do many other commonly used scoring methods (see Conway et al., 2005).

Participants in the concurrent + retro group were also given instructions about the set-by-set strategy reports. The instructions stressed that "we want you to do your best to tell us how you tried to remember the words". The strategies from the report prompt (see above) were each described in one sentence, and no information about the differential effectiveness of the strategies was provided. These particular response options were chosen based on previous research about strategy use and free recall (Hertzog et al., 1998), strategy use and span performance (McNamara & Scott, 2001), and global reports from a small number of pilot participants. During the three practice trials (as well as during the entire OSPAN task), participants first attempted to recall all the words from a given trial and then immediately saw the prompt to make a strategy report for that trial. Although these reports are literally retrospective, we use the label concurrent because the interval between the task and report was minimal (and hence the strategies would still be accessible from immediate memory) and to ensure consistent terminology with previous research.

The retrospective-only group did not receive any strategy instructions prior to the practice trials or the OSPAN task. After participants from both groups finished the OSPAN task, they made the global report of strategy use, prompted by the question, "In as much detail as possible, please tell us how you tried to remember the words during this task. For instance, what strategies—if any—did you use?" At this point, participants had as much time as they needed to type a report of how they performed the OSPAN task. Next, the instructions for the set-by-set strategy reports were presented. Both groups then proceeded to make the retrospective strategy reports. For these reports, the equations and words in each set were presented together on the screen (in the same order that they had originally been presented) along with the prompt for the set-by-set report (see above). Each set was presented one at a time and in the same order in which it had appeared during the OSPAN task. Each participant proceeded through the reports at his or her own pace.

Results
Set-by-set strategy reports
Proportions of each strategy reported. For each participant, we computed the proportion of sets that he or she reported using each of the strategies. Means across individual participant's values are reported in Table 1. Given the number of possible analyses that could be conducted for this data set, we pivot our analyses and discussion around issues that are most relevant to the strategy mediation hypothesis and to describing the properties of the strategy reports.

First, consider whether participants reported using any strategies during the experimenter-driven OSPAN task, which was conducted with the usual intent to reduce strategic behaviour. Consistent with the task demands, participants most often reported just reading the to-be-recalled words as they performed the task. Even so, it is clear that strategy use was not entirely suppressed. On over 20% of the trials, participants reported using normatively effective strategies (sentence, imagery, or grouping), and on about a third of the trials they reported using a normatively less effective strategy (repeating or rehearsing).

Not all participants used each of the strategies. Based on the retrospective reports alone, 30% of the participants reported using only reading or rehearsing, whereas only 63% of the participants reported using at least one of the normatively effective strategies. Despite the fact that more than half the participants reported using a normatively effective strategy during the task, 39% of the sample endorsed using an effective strategy on only 30%, or fewer, of the trials. Thus, although individuals were strategic in some cases, widespread use of normatively effective strategies was not evident, with close to 70% of the participants using them on fewer than a third of the trials.

Second, we examined whether retrospective reports were consistent with concurrent reports. A 2 (kinds of report: concurrent vs. retrospective) x 6 (strategy reported) analysis of variance (ANOVA) was conducted on the proportion of reported strategy use for the concurrent + retro group. Participants reported differential use of the individual strategies—there was a main effect for strategy reported, $F(5, 170) = 14.7$, MSE = 0.12. More important, as evident from inspection of Table 1, the profile of strategies reported did not significantly differ for concurrent and retrospective reports, $F(1, 34) < 1.0$. Such consistency in the two kinds of report demonstrates that participants' retrospective reports largely represent the kinds of strategy purportedly used while performing the OSPAN task; there appears to be little forgetting of strategy when items are re-presented for the retrospective report.

We also examined whether concurrent reports had any reactive effects on strategy use by comparing the profile of retrospective reports for those who made concurrent reports versus those who did not. A 2 (group: retrospective-only vs. concurrent + retro) x 6 (strategy reported) ANOVA on retrospective reports revealed a main effect for the kinds of strategy reported, $F(5, 340) = 20.8$, MSE = 0.062. Of most importance, however, the main effect of group, $F(1, 68) < 1.0$, and the interaction of group with reported strategy, $F(5, 340) < 1.0$, did not approach statistical significance. Nevertheless, inspection of Table 1 suggests that making concurrent reports resulted in somewhat more retrospective reports of merely reading the words.

Table 1. *Proportion of sets that participants reported studying by a given strategy*

Report type/Group	Read	Repeat	Sentence	Imagery	Grouping	Other
			Strategy reported			
			Experiment 1			
Concurrent reports						
Concurrent + retro	.42 (.06)	.32 (.05)	.06 (.02)	.07 (.03)	.10 (.04)	.03 (.01)
Retrospective reports						
Concurrent + retro	39 (.06)	.34 (.06)	.07 (.02)	.07 (.03)	.10 (.04)	.04 (.02)
Retrospective-only	.28 (.05)	.35 (.06)	.09 (.03)	.11 (.03)	.09 (.03)	.07 (.02)
			Experiment 2			
Retrospective reports						
Retrospective-only	.33 (.03)	.35 (.03)	.06 (.01)	.09 (.01)	.10 (.01)	.07 (.01)

Note: See text for detailed description of the various kinds of strategies reported.

Finally, to further evaluate the consistency of reports, for each participant we computed the percentage of items that were matched with the same strategy report for the concurrent and retrospective reports. The mean across participants was 67% (SEM = 3.9). This value indicates moderate intraindividual consistency, and it is similar in magnitude to corresponding values of consistency for the two kinds of strategy report for paired-associate learning (Dunlosky & Hertzog, 2001).

In summary, participants reported using a variety of strategies while performing the OSPAN task. Retrospective reports did not differ substantively from concurrent reports across participants, and the intraindividual consistency for the two kinds of report was moderate, which suggests that forgetting had only a minor influence on the validity of set-specific retrospective reports. And finally, the concurrent reports did not have any major reactive effects, although they may have had a subtle (and statistically non-significant) influence on strategy production. Given that retrospective reports could not have reactive effects on the OSPAN task (because they occurred after the OSPAN task had been performed), we used them exclusively in Experiment 2.

Proportion of each strategy reported as a function of set size. Another advantage that set-by-set reports enjoy over global reports is that analyses of strategy use can be conducted as a function of within-participant factors, such as set size. Thus, we are able to address whether strategy use might change reactively as set sizes increased, such as whether strategic behaviour was suppressed with the longer, more difficult sets. Although these analyses were largely exploratory, we report strategy use as a function of set size mainly for archival reasons.

Because the number of observations was quite small for this conditional analysis, we collapsed across smaller set sizes (set size = 2 and 3) and across larger set sizes (set size = 4 and 5). We then computed the proportion of normatively effective strategies (imagery, sentence generation, and grouping) and the proportion of normatively less effective strategies (reading and rehearsing) as a function of set size. To increase statistical power of these analyses, we also collapsed across groups and based analyses on retrospective reports.

Means across individual's proportions of reported strategies are presented in Table 2. As evident by inspection of these numbers, the proportion of effective versus less effective strategies reported did not differ as a function of set size. The main effect of strategy report (effective vs. less effective) was significant, $F(1, 69) = 31.7$, MSE = 0.38, whereas the main effect for set size and the interaction were not, $Fs < 1.6$, MSEs < 0.08.

Table 2. *Proportion of sets that participants reported studying by a given strategy as a function of set size*

	Strategy reported	
Set size	Less effective	Effective
Experiment 1		
Smaller	.68 (.04)	.28 (.04)
Larger	.68 (.04)	.26 (.03)
Experiment 2		
Smaller	.67 (.03)	.27 (.03)
Larger	.70 (.03)	.21 (.03)

Note: Less effective strategy = reports of reading and rehearsing. Effective strategy = reports of imagery, sentence generation, and grouping. Smaller refers to set sizes of 2 and 3; larger refers to set sizes of 4 and 5. Values are based on retrospective reports, and standard errors of the means are reported in parentheses. Values do not sum to 1.0 within rows because reports of "other strategy" were excluded from these analyses.

Span performance as a function of strategy reported and group. For each participant, we computed an OSPAN score conditionalized on the strategy reported for individual sets, so that each participant had a separate proportion-correct score for each type of strategy reported. This conditional analysis provides support for the construct validity of the reports if strategies that are normatively the most effective for memory—that is, imagery, sentence generation, and grouping—also yield superior levels of span performance. As evident from inspection of Table 3, the mean conditional values across participants were greater for the normatively effective strategies than for the less effective strategies.

Concerning inferential analysis, because few participants reported using all six strategies, an ANOVA excluded the majority of participants due to missing values. Because our goal was to establish that normatively effective strategies (imagery, sentence generation, and grouping) benefited performance above more passive strategies (reading and repeating), we compared a composite score derived from the effective strategies with one from the two more passive strategies. (The category of "other strategy" was not included because we had no a priori expectation as to whether participants' idiosyncratic strategies would tend to be more or less effective.) For retrospective reports (collapsed across the retrospective-only and concurrent þ retro groups), the proportion correct on the span task was significantly greater for normatively effective strategies *(M = .75, Mdn = .80, SEM = .03)* than for the normatively less effective strategies *(M = .60, Mdn = .58, SEM = .02), t(36) = 3.64, p < .05.* Conditionalized on the concur-rent reports, the proportion correct was also significantly greater for the effective strategies *(M = .83, Mdn = .87, SEM = .04)* than for the less effective strategies *(M = .66, Mdn = .67, SEM = .03), t(16) = 2.75, p < .05.* Thus, as expected, span performance was higher when participants reported using mediational strategies that involved associating the to-be-remembered words within a set than when they just read the words or used more passive rehearsal strategies. These outcomes establish the construct validity of the set-by-set strategy reports.

Table 3. *Performance on the operation span task as a function of the strategy reported for the set*

Group/Report	Strategy reported					
	Read	Repeat	Sentence	Imagery	Grouping	Other
	Experiment 1					
Concurrent report						
Concurrent + retro	.61 (.04)	.72 (.04)	.86 (.06)	.82 (.05)	.77 (.06)	.33 (.12)
Retrospective report						
Concurrent + retro	.65 (.04)	.63 (.05)	.77 (.06)	.70 (.09)	.71 (.11)	.29 (.11)
Retrospective-only	.61 (.05)	.50 (.04)	.81 (.06)	.69 (.07)	.85 (.06)	.54 (.08)
	Experiment 2					
Retrospective report						
Retrospective-only	.63 (.03)	.61 (.05)	.73 (.05)	.71 (.05)	.74 (.04)	.61 (.05)

Note: Performance is defined as proportion of target words recalled in serial position from a set. See text for detailed description of the various kinds of strategies reported.

Given that reports of effective strategies demonstrated construct validity, we also evaluated whether reports of effective strategies showed acceptable levels of reliability. To do so, we computed the reliability of the retrospective reports (because all participants made these reports, and hence this estimate would capitalize on the full data set) by computing a split-half correlation between the proportion of effective strategies reported across the odd and even sets. The correlation between this odd/even split (corrected by the Spearman–Brown formula) demonstrated an acceptable level of reliability, $r = .91, p < .05$.

General strategy reports

Each participant provided a general strategy report upon completion of the OSPAN task, immediately before making the retrospective set-by-set reports. The general reports were independently scored by two raters, both blind to the OSPAN score, in a way that was comparable to the set-by-set reports. In particular, general strategy reports were categorized as follows: less effective strategies alone (i.e., participant reported the use of reading or rehearsing or both),[5] less effective/effective strategies (i.e., participant reported using either reading or rehearsing and one of the effective strategies—imagery, sentence generation, or grouping), effective strategies alone (i.e., participant reported using imagery, sentence generation, or grouping), and other (i.e., some idiosyncratic strategy). Agreement between the raters was high (91%), and lack of agreement was settled by a third rating from the senior author.

Table 4. *Proportion of participants using standard strategies based on general strategy reports*

Group	General strategy reported			
	Less effective	Less effective/ Effective	Effective	Other
	Experiment 1			
Concurrent + retro	.57	.23	.20	0
Retrospective-only	.63	.06	.29	.03
	Experiment 2			
Retrospective-only	.60	.16	.10	.14

Note: Less effective = reading and repeating; less effective/ effective = the former strategy plus reporting the use of an effective strategy; effective strategy = imagery, sentence generation, or grouping; other = idiosyncratic strategy not represented by the previous categories.

The mean proportion of participants reporting the use of the various strategies is presented in Table 4. Consistent with the set-by-set reports, the majority of participants reported reading or repeating the to-be-remember words ("less effective" strategy in Table 4). Nevertheless, over 30% of the participants reported using at least one of the normatively effective strategies, regardless of whether they had made concurrent reports while performing the span task.

In Table 5, we report mean proportion of correct span performance as a function of the four categories of general reports. In contrast to the comparable values for the set-by-set reports (Table 3), span performance was not consistently greater across groups when participants reported that they had used a more effective strategy than a less effective one. Excluding idiosyncratic reports ("Other" in Table 5) from the analyses, a 2 (group) x 3 (kind of general strategy report) ANOVA revealed a main effect of group, $F(1, 63) = 6.58$, MSE = 0.017, whereas kind of report, $F(1, 63) = 1.47$, MSE = 0.017, p = .24, and the interaction, $F(1, 63) = 2.30$, MSE = 0.017, p = .11, did not reach statistical significance. Thus, if all we had available were these general reports, we would conclude that participants' strategies in OSPAN had no relation to their performance.

Table 5. *Performance on the operation span task as a function of general strategy reports*

| | General strategy reported | | | |
Group	Less effective	Less effective/Effective	Effective	Other
	Experiment 1			
Concurrent + retro	.63 (.03)	.71 (.06)	.70 (.04)	—
Retrospective-only	.62 (.03)	.45 (.10)	.65 (.04)	.69 (.)[a]
	Experiment 2			
Retrospective-only	.64 (.02)	.62 (.04)	.68 (.04)	.62 (.03)

Note: Less effective = reading and repeating; less effective/effective = the former strategy plus reporting the use of an effective strategy; effective strategy = imagery, sentence generation, or grouping; other = idiosyncratic strategy not represented by the previous categories.
[a]Only one participant contributed to this entry.

Discussion

The method used in Experiment 1 provided a rich set of data for evaluating the strategy mediation hypothesis, and, as important, it demonstrated the validity of the set-by-set strategy reports. Participants reported using associative and elaborative rehearsal on a minority of trials in the OSPAN task, and these reports were consistent across concurrent and retrospective probes. Moreover, strategy reports significantly predicted OSPAN scores, such that trials on which participants reported using a normatively effective strategy yielded higher recall than did trials on which participants reported using a normatively ineffective strategy. General retrospective strategy reports, in contrast, were not associated with task performance, thus suggesting some superiority (in reliability, validity, or both) of set-by-set reports.

A major aim of Experiment 2 was to replicate key findings from Experiment 1 in a manner to provide a more complete evaluation of the strategy mediation hypothesis. In particular, measures of span performance demonstrate criterion validity in that they often correlate with higher order measures of cognition, such as reading comprehension. Such span–ability relationships are explained differently by the strategy mediation hypothesis and by general-process theories. For the latter, the relationship manifests because the particular process that varies across individuals—such as inhibitory or attentional control—influences both span and ability test performance because both rely critically upon that process. By contrast, according to the strategy mediation hypothesis, individuals who more often use effective strategies for span tasks are also more likely to use effective strategies to perform other cognitive tasks. Consistent with the strategy mediation hypothesis, results from Experiment 1 established that effective strategies are used with some frequency, but we do not know whether the individual differences in strategy use mediate any of the relationships between WM span and cognitive ability.

EXPERIMENT 2

To further evaluate the strategy mediation hypothesis, the procedure of Experiment 2 was nearly identical to the one used in Experiment 1, except for the following changes. First, as mentioned earlier, concurrent set-by-set reports were dropped in favour of retrospective reports, which showed reasonable validity and cannot have reactive effects on how participants complete the span task because these reports occur after the OSPAN task is complete. Second, to investigate whether strategy use can account for any span– ability relationships, each participant also completed three standard measures of cognitive (verbal) ability, which included tests of analogical reasoning and reading comprehension. If strategy use in part mediates the relationship between span and these verbal-ability measures, then the span–ability correlation will be reduced when it is partialled on individual differences in the use of effective strategies. Finally, 100 students participated in the full protocol to obtain sufficient statistical power for significantly revealing small-to-moderate correlations.

Method
Participants and materials
A total of 100 undergraduates from UNCG introductory psychology courses participated to fulfil a course requirement. Macintosh computers presented trials during the OSPAN task and recorded strategy reports. For recall during each trial of the span task, participants wrote the target words on a sheet of paper.

Verbal ability tasks
We used three standardized tests to measure verbal ability (Verbal Analogies, Reading Comprehension, and Inference); the first two were modified from the originals by presenting only a subset of items (for detailed descriptions of these modified tasks, see Kane et al., 2004). Kane et al. found these tasks to demonstrate adequate reliability (coefficient alphas of approximately .70–.80) and to correlate significantly with OSPAN (rs of approximately .40–.45).

The Verbal Analogies test was adapted from the Air Force Officer Qualifying Test (AFOQT; Berger, Gupta, Berger, & Skinner, 1990). Each of the 18 items presented an incomplete analogy, which participants had to complete by choosing among five available options. All the items and response choices consisted of fairly high-frequency words, and so the test measured verbal inductive reasoning more than it measured vocabulary knowledge. Participants had 5 min to complete the test. The Reading Comprehension test was also adapted from the AFOQT. Each of the 14 items presented a brief passage on a different topic, and participants had to choose the best completion to the passage's final sentence from among five available choices. Participants had 9 min to complete the test.

The Inference task ("Inferences") was taken from the Educational Testing Service Kit of Factor-Referenced Tests (ETS; Ekstrom, French, Harman, & Dermen, 1976). Each item required participants to read a short paragraph about a topic and to choose the one conclusion (of five) that could be drawn from the passage without assuming any information beyond that provided by the passage. The passages dealt with such topics as fatal car accidents after dark, the endangerment of sea otters, and different cultures' uses of the Stonehenge site. Participants had 6 min to complete the 10 test items from "Part 1" of the ETS test.

Procedure
Participants performed the span task individually as described in Experiment 1. The major methodological change was that all set-by-set reports were made retrospectively. After participants completed the span task, the pencil-and-paper tasks were administered in the following order: analogies, reading comprehension, and then inferences.

Results
Set-by-set strategy reports
Proportion of each strategy reported. For each participant, we computed the proportion of sets that he or she reported using each of the strategies. As evident from inspection of Table 1, participants most often reported

either just reading words or repeating them as they performed the task. Even so, as in Experiment 1, participants reported using the normatively most effective strategies (sentence, imagery, or grouping) on over 20% of trials. Also as in Experiment 1, not all participants used each of the strategies; 24% of the participants reported using only reading or rehearsing, and whereas 71% of the participants reported using at least one of the effective strategies, 31% of the participants endorsed using an effective strategy on 30% or fewer of the trials. Thus, although individuals were strategic in some cases, widespread use of effective strategies was not evident, with over 50% of the participants using them on fewer than a third of the trials.

Proportion of each strategy reported as a function of set size. As in Experiment 1, we computed the proportion of reports of effective and less effective strategies as a function of smaller (set size = 2 and 3) and larger (4 and 5) set sizes. Means across individual's proportions of reported strategies are presented in Table 2. The proportion of effective versus less effective strategies reported did not differ as a function of set size. The main effect of strategy report (effective vs. less effective) was significant, $F(1, 99) = 87.4$, $MSE = 0.23$, whereas the main effect for set size and the interaction were not, $Fs < 2.30$, $MSEs < 0.08$.

Span performance as a function of strategy reported. For each participant, we computed an OSPAN score (proportion of items recalled in correct serial order) conditionalized on the strategy reported for individual sets. As evident from inspection of Table 3, the mean across participants' conditional values was greater for the normatively more effective strategies than the less effective strategies. As in Experiment 1, few participants reported using all six strategies, and so an ANOVA excluded the majority of participants due to missing values. Thus, we again compared a composite score summed across the three effective, more active strategies and the two less effective, more passive strategies. The proportion correct on the span task was significantly greater for normatively effective strategies (M = .70, Mdn = .73, SEM = .03) than for the normatively less effective strategies (M = .6 1, Mdn = .62, SEM = .02), $t(67) = 2.90$, p , .05.

As in Experiment 1, given that the reports of effective strategy use showed construct validity, we also examined the reliability of the reports. We computed the reliability of reporting effective strategies by computing a split-half correlation between the proportion of effective strategies reported for the odd and even sets. The correlation between this odd/even split (corrected by the Spearman–Brown formula) demonstrated an acceptable level of reliability, $r = .77$, p , .05.

General strategy reports
Each participant provided one general strategy report. For correlational analysis reported below, the general strategy reports were scored as follows. If a participant reported using only the normatively less effective strategies (read or repeating), the report received a score of 0 (see Footnote 5). If a participant reported using both less effective and effective strategies (e.g., reading and imagery), the report received a score of 1. If only normatively effective strategies (imagery, sentence generation, or grouping) were reported, the score was 2. The mean proportion of participants reporting the use of the various strategies is presented in Table 4. Consistent with the set-by-set reports, the majority of participants reported reading or repeating the to-be-remember words ("less effective" strategy in Table 4). Nevertheless, about 25% of the participants reported using at least one of the normatively effective strategies. Also, note that this estimate of effective strategy use differs from the estimate using set-by-set reports in which 71% of the participants reported using an effective strategy on one or more sets. In both cases, participants report using effective strategies, but the set-by-set reports appear more sensitive to detecting strategic behaviour when it occurs relatively infrequently across sets.

In Table 5, we report mean span performance as a function of the four categories of general reports. In contrast to the comparable values for the set-by-set reports (Table 3), span performance was not consistently greater when participants reported that they had used a more effective strategy than a less effective one. Excluding idiosyncratic reports ("Other" in Table 5) from the analyses, span performance across the other three conditions did not differ significantly, $F(2, 83) < 1$.

Relations among span, strategy use, and measures of verbal ability

Zero-order correlations were computed among variables that are relevant to evaluating the strategy mediation hypothesis. To provide a measure of the proportion of effective strategy use for each participant, we summed the proportions of reported strategy use across set-by-set reports of imagery, sentence generation, and grouping. General strategy reports were scored as describe above, and given that other idiosyncratic strategies were rare, they were excluded from the correlational analysis. Scores on the three measures of verbal ability (analogies, reading comprehension, and inference) were converted to z-scores.

Table 6. *Pearson r correlations among focal variables in Experiment 2*

	Variable				
	1	2	3	4	5
1. Span performance					
2. Effective strategy use	.30*				
3. General report	.00	.44*			
4. Analogy	.25*	.02	.12		
5. Reading comp	.20*	.16	.24*	.55*	
6. Inference	.05	.10	.09	.44*	.45*

Note: Effective strategy use—proportion of set-by-set retrospective reports summed across proportion of reported imagery use, sentence generation, and grouping. General reports exclude idiosyncratic strategies.
*p < .05.

Correlations among the key measures are reported in Table 6. As expected from the strategy mediation hypothesis, and the ANOVAs presented above, the correlation between effective strategy use (as measured by set-by-set reports) and span performance was significant (r = .30, p < .05). The corresponding value for general reports, however, was not significantly different from 0 (r = .00). Moreover, by a Williams t test for dependent correlations, these two correlations differed significantly from each other, t(97) 2.96, p < .01.

Span performance correlated significantly with the measures of analogical reasoning and reading comprehension, although the correlations were relatively low. We have no explanation for these levels of span–cognition correlation other than bad luck. All of our tests were administered in exactly the same way, with the same materials, as in Kane et al. (2004), who found robust correlations with span. Nevertheless, although the present correlations are relatively low, they are well within the range of those reported elsewhere in the literature (for a review, see Daneman & Merikle, 1996). Most important, above-zero correlations between span performance and cognitive performance were essential for evaluating the degree to which individual differences in strategy production might mediate the relationship between working memory and cognitive ability. We recomputed the correlation between OSPAN and verbal-ability scores, partialled on the proportion of effective strategy use. Because the analogies and comprehension measures were moderately correlated, we combined the two in conducting this analysis (combined-measure correlation with span, r = .26, p < .05). This span–ability correlation was influenced minimally by partialling on the derived measure of effective strategy use or when partialling individually on each of the three kinds of normatively effective strategy (both rs = .24).

Discussion

The critical outcomes from Experiment 2 replicated those from Experiment 1: The set-by-set reports had construct validity, accounted for variability in span performance, and were more highly related to span performance than were general reports. As important, new results from Experiment 2 provided a further test of the strategy mediation hypothesis. Namely, another prediction from this hypothesis is that individual differences in strategy use would mediate the relationship between span performance and performance on measures of higher order cognition. To evaluate this prediction, we chose tasks that have often been used by researchers

interested in the criterion validity of span tasks. Span significantly correlated with measures of verbal ability, and, most important, this span–ability correlation was not influenced when individual differences in strategy use were statistically factored out. Given that the span–ability correlations reported here were relatively low, perhaps strategy use would mediate some of this relationship if the correlations had been nearer the higher end of the range of those reported in the literature (Daneman & Merikle, 1996). Even so, our results indicate that strategy use is unlikely to mediate all of the shared variance between span and cognitive ability. Of course, this conclusion may not hold for other sets of span and ability measures, but it is also important to note that other researchers have reported that strategy use (as measured by general reports or by viewing times) also failed to statistically mediate span– ability relationships on other tasks (Engle et al., 1992; Friedman & Miyake, 2004; Turley-Ames &Whitfield, 2003).

Although we identified strategies that people often spontaneously use to learn verbal materials used on the OSPAN task, perhaps participants were also using other kinds of strategy when performing this task that were not tapped by the set-by-set strategy reports, which may account for some of the span–ability relationships. The influence of other strategies could be investigated in future research by developing variants on the set-by-set reports introduced here. Nevertheless, at least for the current set of verbal tasks—as well as those investigated by Engle et al. (1992), Turley-Ames and Whitfield (2003), and Friedman and Miyake (2004)—the criterion validity of the span task is apparently due to factors other than strategies that are commonly used in verbal learning. If anything, these prior studies, which used nonstandard versions of OSPAN that encouraged strategy use, indicate that variability in strategy use is a nuisance variable that sup-presses the correlation between measures of WM span and verbal ability.

GENERAL DISCUSSION
In the present research, we had two major inter-related aims: introduce a measure of strategy use at the level of individual item sets that can be used with span tasks and, as important, implement this measure so as to evaluate the strategy mediation hypothesis.

A central prediction from this hypothesis is that individual differences in the production of effective strategies will account for individual differences in span scores. Certainly, other researchers to date have presented affirmative evidence by demonstrating (a) that individual differences in self-pacing during the span task are related to span performance (Engle et al., 1992; Friedman &Miyake, 2004) and (b) that instructing individuals to use strategies can boost span scores (Turley-Ames & Whitfield, 2003). However, it is important to note that this past research used versions of the span task that either allowed or encouraged strategy use. Thus, although evidence from these studies has been informative, it does not speak directly to the question of whether strategy production influences span performance as it is typically measured. We questioned the role that strategies played in span tasks that are implemented in a standard manner, which involves experimenter pacing that is meant to suppress strategic behaviour so as to obtain a more valid measure of WM capacity. Across two experiments, individual differences in effective strategy use were evident. As important, these differences significantly correlated with performance on the span task. These outcomes confirm a critical prediction of the strategy mediation hypothesis.

Given support for this prediction, an important theoretical question becomes, how should the significant strategy–span correlation be interpreted? Two possibilities are most salient in the present context. According to the strategy-as-cause account, some individuals may in general be more strategic than are others across many tasks. In this case, differences in strategy use cause differences in span. Alternatively, individual differences in WM capacity may afford the use of strategies, which in turn can contribute to span scores.

According to this strategy-as-effect account, all individuals would presumably be similarly strategic across tasks, but for difficult tasks in which little— if any—instruction is provided (such as standard span tasks), having a higher WM capacity affords the production of effective strategies while performing the task, particularly if producing those strategies is cognitively demanding. In this case, differences in the integrity of

the system architecture cause differences in the use of algorithms, whereas in the former case, the opposite is true.

These two accounts are not mutually exclusive, and hence both may provide a partial account of strategy–span correlations. Nevertheless, several diverse sources of evidence demonstrate the empirical validity of the strategy-as-effect account. For instance, Dunlosky and Thiede (2004) instructed high-span and low-span participants to strategically allocate study time across a set of paired-associate items. Participants were instructed in how to specifically allocate time. Namely, their goal was to learn only 6 items from a list of 30 paired-associate items, and hence all participants were explicitly instructed to select the easiest items (and to select few of them) for restudy. For selection of items for restudy, items were presented individually, which makes the task particularly difficult. Whereas high-span participants selected the easier items over more difficult items, the low-span participants tended to do just the opposite and on average selected over 16 items of the list, which contrasted with the explicit instructions. In this case, higher WM capacity afforded the successful implementation of a more effective (and instructed) strategy.

A second source of support for strategy-as-effect comes from studies that have instructed or trained people to use strategies during span tasks. For example, Turley-Ames and Whitfield (2003) found that high-WM participants benefited little when instructed to use rehearsal, imagery, or semantic-chaining strategies during OSPAN, because they appeared to have been performing optimally in the pretest, prior to any instruction. In contrast, low-span participants showed marked gains in span score following instruction, but only for the maintenance rehearsal mnemonic. Low spans did not appear to be able to use the more demanding imagery or chaining strategies with much success, and few reported using such strategies spontaneously.

As a final example, in children, instructions to use an imagery strategy to remember lists of verbal materials seem to benefit only those who are older than 6 or 7 years; 4- to 6-year-olds typically fail to improve their recall following imagery instructions (e.g., Guttmann, Levin, & Pressley, 1977; Pressley & Levin, 1977). These findings suggest that some minimal cognitive capacity is required to successfully engage in visual imagery as a mnemonic aid. Indeed, Pressley, Cariglia-Bull, Deane, and Schneider (1987) tested directly whether short-term memory (STM) capacity would predict children's ability to use an imagery strategy. First- through sixth-graders were instructed to use imagery, or not, in order to learn lists of concrete sentences, and all participants completed a battery of three verbal STM span tasks. A composite STM measure predicted sentence learning significantly more strongly in the imagery condition (r = .71) than in the control condition (r = .42), and after partialling age from the correlations, STM span was a significant predictor of learning for only the imagery participants (r = .51). These findings suggest that imagery production is capacity demanding and that greater STM capacity allows more effective use of this strategy. Indeed, when external sup-ports, such as pictures, are provided for an imagery strategy, young children (Guttman et al., 1977) and those with low STM capacity (Ryan, Ledger, &Weed, 1987) can use an imagery strategy successfully to learn new verbal materials. In sum, we find these converging sources of evidence for the strategy-as-effect to be compelling, but we acknowledge that they cannot rule out the possibility that a higher order latent construct of strategy use also may contribute to the strategy-span correlations.

Importantly, the methods developed in the present research could be adapted to further evaluate the two accounts of strategy–span correlations described above as well as to identify when (and why) strategy use mediates span– ability relationships. To this end, the present research also supported some key conclusions about measures of strategy use. First, consistent with previous research (e.g., Friedman & Miyake, 2004; McNamara & Scott, 2001), when making general reports of strategy use, some participants report the use of effective strategies. Second, the general reports were significantly correlated with the set-by-set reports. This relationship provides evidence for the convergent validity of the measures of strategy use. Nevertheless, the general and item-specific reports shared only 27% of their variance in Experiment 1 and only 19% in Experiment 2. Third, and most important, the set-by-set reports introduced for the OSPAN task in the present research demonstrated greater utility in exploring strategy effects. Notably, in contrast to the general reports, individual differences in set-by-set reports of effective strategy use correlated significantly with individual

differences in span performance. Perhaps if participants were forced to estimate how often effective strategies were used during the OSPAN task when making general reports, this frequency of effective strategy use would correlate significantly with span performance. But even if general reports could be improved, using set-by-set reports would still afford a more fine-grained analysis aimed at describing possible variability in strategy use for a given individual or task. For instance, set-by-set reports revealed that the same profile of effective strategies was used across both small and larger set sizes. This outcome—which could not be revealed by general reports—suggests that strategic behaviour was consistent across the OSPAN task. Thus, although global reports can certainly be valuable for investigating strategy production, based on the present outcomes we also can strongly recommend the use of set-by-set retrospective reports for exploring strategy-based hypotheses of working memory.

In summary, new evidence from set-by-set strategy reports was brought to bear on the strategy mediation hypothesis, which is that individual differences in strategy use contribute to span performance (Experiments 1 and 2) and mediate span–ability relationships (Experiment 2). Evidence using a standard experimenter-paced OSPAN task demonstrated that strategy use accounted for individual differences in span performance, although a strategy-based account was insufficient for explaining the present relationship between span and verbal ability.

Notes:

1 We chose the term experimenter-driven task because although each participant paces his or her own performance on the task (such as in solving word problems or reading to-be-remembered words aloud), the experimenter paces each participant individually through the procedure. This prevents participants from taking undue time to rehearse the memoranda. Accordingly, experimenter-driven tasks are used to minimize strategic behaviours that would potentially undermine the validity of the span task (a concern that Friedman & Miyake, 2004, demonstrated to be valid).

2 In a study of children, Swanson (1992) developed a large battery of WM tasks, some of which included a concurrent strategy report as a secondary processing task. For example, in an auditory digit-sequencing task, participants heard sequences of numbers, such as addresses or phone numbers, embedded in sentences. After hearing each sentence, children answered a comprehension question about it and then pointed to a visual depiction of one of four possible strategies they would use to remember the digit sequence, such as rote rehearsal, grouping, and rhyming. After strategy selection, participants recalled the digits. Although these strategy reports would be useful in evaluating the strategy mediation hypothesis, Swanson did not report any data about the participants' strategy selection or its relation to performance.

3 Of course, while making general reports, participants could elect to volunteer estimates of how often each strategy was used. In our studies, however, this never occurred. Moreover, even if it had occurred (or if we had asked participants to report this information), the specific trials in which each strategy had been used could not be ascertained from the general reports, which would rule out conditional analyses (e.g., span performance as a function of strategy use on individual trials) that have bearing on the effectiveness of individual strategies.

4 In the present article, we used the term effective strategy to refer to imagery, sentence generation, and grouping, and the term less effective strategy to refer to repetition and reading. The terms are not meant to reflect the absolute influence of these strategies on performance, but merely their relative and normative effects as demonstrated by decades of research on human learning and memory. To foreshadow, subsequent analysis of span performance as a function of strategy use supports the use of the present distinction, because in fact performance was higher for the normatively effective strategies than for the less effective strategies.

5 Turley-Ames and Whitfield (2003) did report that performance was somewhat better when participants reported repeating items than merely reading them. Although both strategies are normatively inferior to the other mediational strategies, we reanalysed span performance based on general reports for these two relatively passive strategies. In Experiment 1, span performance was significantly greater for those who reported that they

repeated (M span = 0.67, n = 17) than for those who reported reading (M span = 0.57, n = 18), t(33) = 2.29. Note, however, that this effect did not replicate in Experiment 2, which involved a substantially larger sample size: M span = 0.63 for repeaters (n = 23), and M span = 0.61 for readers (n = 44), t(65) = 0.80. Thus, in the present context, differences between these two general reports were not robust. Furthermore, because using this breakdown in other analyses involving general reports does not yield different conclusions from those discussed in the text and because this particular distinction is less relevant to our current aims we do not discuss it further.

REFERENCES

Baddeley, A. D. (2000). Short-term and working memory. In E. Tulving & F. I. M. Craik (Eds.), The Oxford handbook of memory (pp. 77–92). New York: Oxford University Press.

Berger, F. R., Gupta, W. B., Berger, R. M., & Skinner, J. (1990). Air Force Officer Qualifying Test (AFOQT) form P: Test manual (AFHRL-TR-89 – 56). Brooks Air Force Base, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Conway, A. R. A., 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, 12,769–786.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. Trends in Cognitive Sciences, 7, 547–552.

Cowan, N. (2005). Working memory capacity. New York: Psychology Press.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. Psychonomic Bulletin &Review, 3,422–433.

Dunlosky, J., & Hertzog, C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. Memory & Cognition, 29,247–253.

Dunlosky, J., & Thiede, K. W. (2004). Causes and constraints of the shift-to-easier-materials effect in the control of study. Memory & Cognition, 32, 779–788.

Ekstrom, R. B., French, J. W., Harman, M. H., & Dermen, D. (1976). Manual for kit of factor-referenced cognitive tests. Princeton, NJ: Educational Testing Service.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,972–992.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), The psychology of learning and motivation (pp. 145–199). New York: Academic Press.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the pre-frontal cortex. In A. Miyake & P. Shah (Eds.), Models of working memory: Mechanisms of active maintenance and executive control (pp. 102–134). New York: Cambridge University Press.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. Psychological Review, 87, 215–251.

Ericsson, K. A., & Simon, H. A. (1984). Protocol analysis: Verbal reports as data. Cambridge, MA: MIT Press.

Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. Journal of Memory and Language, 51, 136–158.

Guttmann, J., Levin, J. R., & Pressley, M. (1977). Pictures, partial pictures, and young children's oral prose learning. Journal of Educational Psychology, 69, 473–480.

Harris, J. L., & Qualls, C. D. (2000). The association of elaborative or maintenance rehearsal with age, reading comprehension, and verbal working memory performance. Aphasiology, 14, 515 –526.

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, 15, 85 –106.

Hunt, E. (1999). What is a theory of thought? In R. J. Sternberg (Ed.), The nature of cognition (pp. 3–49). Cambridge, MA: MIT Press.

Kane, M. J., & Engle, R. W. (2000). Working memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26, 333–358.

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, 189 –217.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. Journal of Experimental Psychology: General, 130, 199–207.

May, C. P., Hasher, L., & Kane, M. J. (1999). The role of interference in memory span. Memory and Cognition, 27, 759–767.

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. Memory and Cognition, 29,10–17.

Miyake, A., & Shah, P. (1999). Models of working memory: Mechanisms of active maintenance and executive control. New York: Cambridge University Press.

Oberauer, K. (2005). Binding and inhibition in working memory—individual and age differences in short-term recognition. Journal of Experimental Psychology: General, 134, 368 –387.

Pressley, M., Cariglia-Bull, T., Deane, S., & Schneider, W. (1987). Short-term memory, verbal competence, and age as predictors of imagery instructional effectiveness. Journal of Experimental Child Psychology, 43, 194–211.

Pressley, M., & Levin, J. R. (1977). Task parameters affecting the efficacy of a visual imagery learning strategy in younger and older children. Journal of Experimental Child Psychology, 24, 53 –59.

Ryan, E. B., Ledger, G. W., &Weed, K. A. (1987). Acquisition and transfer of an integrative imagery strategy by young children. Child Development, 58, 443–452.

Salthouse, T. A. (1996). The processing speed theory of adult age differences in cognition. Psychological Review, 103, 403–428.

Swanson, H. L. (1992). Generality and modifiability of working memory skilled and less skilled readers. Journal of Educational Psychology, 84, 473–488.

Turley-Ames, K. J., &Whitfield, M. M. (2003). Strategy training and working memory task performance. Journal of Memory and Language, 49, 446–468.

Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). Psychological Review, 103, 761–772.