

# Contribution of strategy use to performance on complex and simple span tasks

Heather Bailey · John Dunlosky · Michael J. Kane

Published online: 23 November 2010  
© Psychonomic Society, Inc. 2010

**Abstract** Simple and complex span tasks are widely thought to measure related but separable memory constructs. Recently, however, research has demonstrated that simple and complex span tasks may tap, in part, the same construct because both similarly predict performance on measures of fluid intelligence (Gf) when the number of items retrieved from secondary memory (SM) is equated (Unsworth & Engle, *Journal of Memory and Language* 54:68–80 2006). Two studies ( $n = 105$  and  $n = 152$ ) evaluated whether retrieval from SM is influenced by individual differences in the use of encoding strategies during span tasks. Results demonstrated that, after equating the number of items retrieved from SM, simple and complex span performance similarly predicted Gf performance, but rates of effective strategy use did not mediate the span-Gf relationships. Moreover, at the level of individual differences, effective strategy use was more highly related to complex span performance than to simple span performance. Thus, even though individual differences in effective strategy use influenced span performance on trials that required retrieval from SM, strategic behavior at

encoding cannot account for the similarities between simple and complex span tasks.

**Keywords** Working memory · Short-term memory · Secondary memory · Strategy use · Fluid intelligence

## Introduction

Span tests of short-term memory (STM), sometimes known as “simple” span tasks (Turner & Engle, 1989), present subjects with short lists of items to recall immediately in serial order, and they are designed to measure the rehearsal, coding, and storage functions of putative STM buffers. Simple span tasks have been included in psychometric test batteries for over 100 years (for review, see Ackerman, Beier & Boyle, 2005), and they currently figure prominently in tests of competing theoretical models of STM storage mechanisms (e.g., Brown, Neath, & Chater, 2007; Burgess & Hitch, 2006; Oberauer & Lewandowsky, 2008). So-called “complex” span tasks, in contrast, have been researched only for the last quarter century. For these span tasks, a secondary processing demand is inserted between each to-be-remembered item, as a means to mimic the function of the working memory (WM) system, to maintain accessibility of information in the service of ongoing cognitive activity. Widespread interest in complex span tasks has derived, in part, from their broad predictive power, i.e., individual differences in complex span performance are strong predictors of fluid aspects of cognition, from comprehension to learning to reasoning. They also often do so more strongly and indiscriminately than do individual differences in simple span (for reviews, see Conway, Kane, & Engle, 2003; Daneman & Merikle, 1996; Engle, Kane, & Tuholski, 1999). As we will discuss below,

---

H. Bailey (✉)  
Washington University,  
439 Psychology Building, One Brookings Drive,  
Campus Box 1125,  
Saint Louis, MO 63130, USA  
e-mail: hroth@artsci.wustl.edu

J. Dunlosky  
Kent State University,  
Kent, USA

M. J. Kane  
University of North Carolina at Greensboro,  
Greensboro, USA

however, recent studies have begun to question the extent to which simple and complex span tasks tap different underlying mechanisms, and the present study engages this question by examining the effects of encoding strategies on span performance and its correlates.

Because simple span tasks are designed to measure the operation of STM buffers, relatively free of the influence of executive mechanisms, they require only short-term retention (or “storage”) of short item lists, such as words, letters, or digits. The word span (WSPAN) task is a popular simple span task in which subjects study a list of words (e.g., fruit, eagle, work, etc.) that are presented one at a time. Lists vary in length—often between three and seven words. Subjects study each word as it is presented and then attempt to recall the words in serial order. In contrast, WM represents the storage and concurrent manipulation of information and is measured by complex span tasks, which involve a storage task along with an interleaving processing task. One popular complex span task is the reading span (RSPAN) task (Daneman and Carpenter, 1980). In one version of RSPAN (Turner & Engle, 1989), subjects view a sentence and an unrelated word (e.g., “The ice melted quickly under the warm sun. ? fruit”). They read the sentence aloud, decide whether it is coherent, read the word aloud (i.e., fruit), and then see the next sentence-word pair (i.e., “Lori left the engine on and parked the grape.? eagle”). Subjects typically complete between three and seven of these sentence-word pairs for a given trial. Following the final pair of each trial, they attempt to recall the words in serial order (e.g., fruit, eagle, etc.).

### Differences between simple and complex span performance

Initial evidence of differences between simple and complex span tasks came from research that compared the ability of span performance to predict individual differences in reading comprehension. In particular, Daneman and Carpenter (1980) argued that comprehension performance would be better predicted by tasks that measure the ability to simultaneously manipulate and store information (e.g., complex span tasks) than by tasks that measure the ability to store information alone (e.g., simple span tasks). To evaluate this prediction, subjects completed a complex span task (e.g., RSPAN), a simple span task (e.g., WSPAN), and measures of comprehension (e.g., pronoun reference task and verbal SAT scores). As expected, RSPAN, but not WSPAN, performance significantly predicted comprehension. Moreover, a meta-analysis conducted by Daneman and Merikle (1996) demonstrated that span tasks that involved solving math problems (e.g., operation span and computation span tasks) also significantly predicted

comprehension. Thus, as noted above, tasks with processing and storage components—regardless of whether the processing component is verbal or mathematical—often predict higher-order cognition better than do tasks with only a storage component. Similarly, Engle, Tuholski, Laughlin, and Conway (1999) demonstrated that simple and complex span tasks load onto two separate factors (i.e., STM & WM; see also Alloway, Gathercole, & Pickering, 2006; Cantor, Engle, & Hamilton, 1991), and that complex, but not simple, span performance significantly predicted unique variance in general fluid intelligence (Gf).

### Similarities between simple and complex span performance

In contrast to these studies, Colom, Rebollo, Abad, and Shih (2006) reported that simple span performance predicted Gf. They reanalyzed data from five previous studies using the model and analyses from Engle et al. (1999). These studies involved performance on a variety of complex span, simple span, and higher-order cognitive tasks. As in Engle et al. (1999), Colom et al. (2006) fit separate structural-equation models to data from each study and found that the STM and WM constructs were more strongly related than previous studies had originally reported. Most important, STM and WM tasks predicted Gf similarly in several data sets, and at times, STM was a stronger predictor (see also, Bayliss, Jarrold, Baddeley, & Gunn, 2005; La Pointe & Engle, 1990; Shah & Miyake, 1996). These outcomes suggest that performance on Gf tasks is predicted by a common mechanism shared between simple and complex span tasks, which Colom and colleagues argued to be “storage” (see also Tehan, Hendry, & Kocinski, 2001).

In summary, some evidence indicates that only complex span performance predicts Gf, whereas other evidence indicates that both complex and simple span performance do. This discrepancy was explored by Unsworth and Engle (2006). They claimed that simple and complex span performance similarly predicts Gf only when both tasks require the same amount of information to be retrieved from inactive, or secondary, memory (SM). As opposed to SM, primary memory (PM) reflects currently activated representations. Importantly, PM is thought to be the activated portion of SM that is currently needed for processing and has, on average, a four-item capacity (cf. Cowan, 2001; Unsworth & Engle, 2006). Note here that although the majority of individuals can hold up to four items in PM, this capacity can vary. Some individuals can only maintain up to three items, whereas others can maintain up to five (Cowan, 2001). Thus, according to this view, when a list of items is presented during a simple span

task, on average, four are active in PM. When the fifth item is presented, PM is beyond its capacity so one item must be recovered from SM. During a complex span task, such as the RSPAN task, a sentence is read aloud and processed. Then the to-be-remembered word is maintained in PM until the next sentence is presented. Because the second sentence contains enough information to fill PM capacity, the first word must be recalled from SM. Thus, all to-be-remembered items on a complex span task are retrieved from SM with the exception of the final word, which is maintained in PM because no sentence follows (Unsworth & Engle, 2006).

According to this rationale, simple and complex span performance depends on maintenance of information in PM and a search through information in SM. Unsworth and Engle (2007) proposed that (1) this search is difficult because SM contains a large amount of potentially interfering information and (2) individuals use cues to narrow the search and increase the probability of successful retrieval. Importantly, complex span performance depends more upon this cue-dependent search than does simple span performance, because a higher proportion of items are retrieved from SM on complex versus simple span tasks. Further, performance on measures of Gf also may rely on this cue-dependent search of SM (Unsworth & Engle, 2007). For example, in the Raven Advanced Progressive matrices (RAPM; Raven, Raven, & Court, 1998)—a measure of Gf—subjects must decide which pattern completes a matrix design. To do so, subjects must infer, test, and remember multiple candidate rules as they process the individual items in the matrix and the set of potential solutions that are presented. Thus, performance on RAPM may involve active maintenance of the patterns in PM and retrieval of the rules from SM. This shared reliance on SM retrieval processes may be responsible for the correlations observed between complex span and Gf performance. By contrast, simple span performance depends less on SM, which may explain the lower correlations between simple span and Gf performance. Thus, equating the number of items retrieved from SM during complex and simple span tasks should result in a simple span-Gf relationship as well.

Unsworth and Engle (2006) evaluated this possibility by manipulating list length so the number of items retrieved from SM ranged from one to three on both simple and complex span tasks. More specifically, because four items are retrieved from PM on a simple span task, a list length of five yields four items from PM and one item from SM; a list length of six yields four items from PM and two from SM, and so on. Because all but the final item are retrieved from SM on a complex span task, for a list length of three, the first two items are drawn from SM and the final item from PM. After equating the number of items in SM, Unsworth and Engle (2006) found that performance on Gf

tasks was predicted equally well by simple and complex span tasks. Most important, Gf performance largely was predicted by performance on trials that required retrieval from SM.

This evidence sparked a debate regarding the degree to which individual differences in SM retrieval drive observed span-Gf relationships (see Unsworth & Engle, 2007). Mogle, Lovett, Stawski, and Sliwinski (2008) measured performance on several tasks that tap SM, such as word recognition and paired-associate recall. Performance on these SM tasks completely accounted for the variance shared between complex span and Gf performance. Because span performance did not account for any unique variance in Gf performance beyond that of SM, Mogle et al. (2008) concluded that SM is entirely responsible for span-Gf relationships. Results from another study, however, using similar methods and tasks, demonstrated that both retrieval from SM *and* active maintenance of information in PM account for the relationship between span and Gf (Unsworth, Brewer, & Spillers, 2009). Regardless of whether SM solely or jointly accounts for span-Gf relationships, evidence converges on the conclusion that retrieval from SM is important.

### The role of strategy use

A fundamental question now arises: Why does retrieval from SM predict span-Gf relationships? As mentioned above, Unsworth and Engle (2007) claimed that both span and Gf tasks require a cue-dependent search of information in SM. One way to increase the probability of successful retrieval from SM (i.e., the efficiency of the search through SM) is through the use of encoding strategies. Strategies may aid retrieval because they create a more distinctive trace for a particular set of to-be-remembered items in SM (Marschark, Richman, Yuille, & Hunt, 1987). Consider some of the tasks used by Mogle et al. (2008) and Unsworth et al. (2009) to measure SM: word recognition, free recall, and paired-associate recall. All of these tasks afford the use of verbal encoding strategies. For instance, chunking to-be-remembered items during encoding improves free recall performance (Ericsson, Chase, & Faloon, 1980). On a paired-associate recall task, Richardson (1998) demonstrated that people use a variety of strategies (e.g., imagery, verbal mediators, and repetition) while encoding word pairs. As expected, images and verbal mediators were associated with higher recall than was repetition, which we refer to as a less effective strategy because it was normatively associated with lower levels of performance. The same effect also has been observed in complex span performance (Bailey, Dunlosky, & Hertzog, 2009; Bailey, Dunlosky, & Kane, 2008; Dunlosky & Kane, 2007;

Kaakinen & Hyona, 2007; McNamara & Scott, 2001; St. Clair-Thompson, Stevens, Hunt, & Bolder, 2010). Specifically, span performance is significantly higher on trials in which individuals report using normatively effective strategies (e.g., imagery and sentence generation) as compared to trials in which they report using less effective ones (e.g., passive reading and rote repetition). From this point on, we will refer to these normatively effective strategies as *effective* and the normatively less effective strategies as *less effective* strategies.

Similar to their contribution to complex span performance, encoding strategies should be critical to simple span performance after the influence of PM is reduced, because when information is retrieved from PM, individual differences in strategy use do not predict recall performance (Baddeley & Ecob, 1970). That is, because performance is nearly perfect on short simple span trials of four or fewer items, differential strategy use does not strongly influence simple span performance. However, by reducing the influence of PM, and thereby increasing the influence of SM as in complex span and Gf tasks, strategy use may partly explain simple span-Gf relationships.

Given that using effective strategies at encoding may aid the search through SM at retrieval, they may be responsible for the similarity between complex span-Gf and simple span-Gf relationships after equating items retrieved from SM (Unsworth & Engle, 2006). Based on this assumption, we predict effective strategy use will be related to task performance on set sizes in which items are retrieved from SM. Most important, if strategy use accounts for individual differences in retrieval from SM, then we also predict effective strategies will mediate the relationship between span and Gf performance. That is, both span and Gf tasks require individuals to solve novel problems and generate ad hoc strategies, and previous research has identified effective and less effective strategies used on Gf tasks (Carpenter, Just, & Shell, 1990; Mitchum & Kelley, *in press*; Vigneau, Caissie, & Bors, 2005) and on similar tasks that require adaptive problem solving (Bethell-Fox, Lohman & Snow, 1984). Specifically, an effective strategy is to determine the emerging pattern between items and generate the correct response, whereas a less effective strategy is to scan the response options and individually eliminate the incorrect ones. Because span and Gf tasks both afford effective strategies, individuals who are strategic on span tasks could be those who are also strategic on Gf tasks.

To assess effective strategy use, subjects completed set-by-set strategy reports for all span tasks: After completing the final trial of a given span task, subjects were re-presented with each item set separately and asked to report whether they had used reading, repetition, sentence generation, imagery, grouping, or something else to remember the items from that set. We asked about these particular encoding strategies because

previous research has found that they are frequently used in verbal span tasks (Bailey et al., 2008, 2009; Dunlosky & Kane, 2007; Turley-Ames & Whitfield, 2003). The validity of these strategy reports for the OSPAN and RSPAN tasks has been established repeatedly by demonstrating significantly higher performance on trials in which subjects reported using effective strategies as compared to trials in which they reported using less effective strategies (e.g., Bailey et al., 2008; Dunlosky & Kane, 2007; Unsworth & Spillers, 2010). Further, making these strategy reports appears to have minimal reactive effects (Dunlosky & Kane, 2007), which indicates that reporting strategies on one span task trial has little influence on strategy selection or span performance, either on future trials or on future span tasks.

## Experiment 1

A main goal of the first experiment was to evaluate whether effective strategies have the same influence on simple and complex span performance after equating the number of items retrieved from SM. Another goal was to evaluate whether effective strategy use mediates span-Gf associations.

## Method

### Subjects

One hundred five undergraduates (75 females) from introductory psychology courses at Kent State University participated to complete a course requirement. Their mean age was 19.7 ( $SD = 3.9$ ) years and mean years of education was 12.1 ( $SD = 0.66$ ).

### Materials

**OSPAN task** We used the version of the OSPAN task described in Kane et al. (2004). Subjects read a mathematical operation aloud (e.g., “Is  $(3 \times 2) + 5 = 10$ ?”), reported whether it was correct, and then read a target word aloud (e.g., “phone”). Immediately thereafter, the experimenter pressed a key to present the next operation-word pair on-screen. Following the final pair of the trial, subjects recalled the target words in serial order. The OSPAN task consisted of 15 experimenter-paced trials that ranged from three to seven operation-word pairs. The words and the order of set sizes were initially randomized and that order was used for all subjects.

Performance on all span tasks was computed using partial-credit unit scoring (for details, see Conway et al., 2005). That is, performance on each trial was scored as the

proportion of correctly recalled items (e.g., trial 1:  $3/4 = 0.75$ , trial 2:  $3/3 = 1$ , and trial 3:  $4/6 = 0.67$ ), and overall performance was expressed as the mean proportion of correctly recalled items, [e.g.,  $(0.75 + 1 + 0.67)/3 = 0.81$ ].

**RSPAN task** We used a modified version of the RSPAN task from Kane et al. (2004). Subjects read a sentence aloud (e.g., "Mr. Owens left the lawnmower in the lemon."), reported whether it made sense, and then read an unrelated word aloud (e.g., "eagle"). Once the word was read aloud, the next sentence-word pair appeared on-screen. After the final pair of each trial, subjects wrote the target words in serial order. The RSPAN task consisted of 15 experimenter-paced trials that ranged from three to seven sentence-word pairs presented in random order.

**Word span task** On each trial, words were presented one at a time for one second apiece. The WSPAN task consisted of 15 trials: three trials of each list-length ranging from three to seven words. Subjects read each word aloud when it appeared and, following the final word on a trial, they typed them in serial order.

**Letter span task** The letter span (LSPAN) task was identical to WSPAN except it presented individual letters rather than words. All 26 letters of the alphabet were used randomly throughout the lists.

**Set-by-set strategy reports** Subjects completed strategy reports after the final trial on each span task. We used the set-by-set strategy reports created by Dunlosky and Kane (2007), in which the stimuli from each trial were re-presented on-screen along with the six strategy options (reading, repetition, sentence generation, imagery, grouping, or a different strategy). For example, after the 15th trial on the OSPAN task, the equations and words from the first trial were presented and subjects indicated which strategy they used to remember those words. After reporting their strategy for the first trial, the equations and words from the second trial were presented.

**Letter sets test** One Gf measure was the letter sets test from the Kit of Reference Tests for Cognitive Factors (Ekstrom, French, Harman, & Dermen, 1976), in which each item consisted of five sets of letters with four letters in each set. Four of the sets followed the same rule for ordering the letters, but the fifth set did not. Subjects were instructed to find the rule that relates the four sets of letters to each other; once they identified the rule, they indicated the fifth set, which did not fit the rule. Subjects had seven minutes to complete as many as possible and were instructed not to guess, but only to answer those sets in which they knew the rule. We scored performance as the proportion of correct answers.

**Locations test** Another measure of Gf was the locations test from the Kit of Reference Tests for Cognitive Factors (Ekstrom et al., 1976). Subjects were shown several sets, each of which consisted of five rows of dashes that were separated into groups. In each of the first four rows, one dash was replaced by an "X"; and in the fifth row, five of the dashes were replaced by numbers (i.e., labeled 1–5). Subjects attempted to find the rule for the location of the "X" in the first four rows of dashes and then to choose the correct location, which was denoted by a number, for an "X" in the fifth row. Subjects had 6 minutes to complete as many sets as possible and were instructed only to indicate an answer for those sets in which they knew the rule. Again, we scored performance as proportion correct.

## Procedure

Subjects completed all tasks in a 1.5-hour session. They first filled out a consent form, followed by the OSPAN task, a demographics questionnaire, the WSPAN task, letter sets test, the RSPAN task, the locations test, and finally, the LSPAN task.

## Results

First, scores on the simple and complex span tasks were computed along with their correlations with the Gf tasks. Next, we compared reported strategy use on simple span and complex span tasks to evaluate whether one type of span task lended itself more to the use of effective strategies. Finally, we examined the contribution of effective strategy use to performance on both types of span task and to the span-Gf relationships. To help interpret these analyses, the descriptive statistics from each task are reported in Table 1. Also included in Table 1 are the proportions of effective strategies used on each span task.<sup>1</sup> These variables were derived by first computing the mean proportion of span-task trials that each subject reported, imagery, sentence generation, and grouping, and then by averaging across all subjects. Note here that the strategy reports appeared to have minimal reactive effects across tasks, as effective strategies were reported equally often on the first (OSPAN:  $M = 0.22$ ) and second complex span task (RSPAN:  $M = 0.25$ ),  $t(132) = 1.15$ ,  $p = 0.25$ , and on the

<sup>1</sup> Given the different nature of their stimuli (i.e., letters versus words), LSPAN and the other span tasks may differentially afford effective strategies. Thus, we conducted strategy-related analyses comparing LSPAN to each of the other span tasks. For both experiments, the proportion of effective strategies reported on LSPAN did not significantly differ from the proportion of effective strategies reported on the WSPAN, OSPAN, or RSPAN tasks.

**Table 1** Descriptive statistics for span tasks, Gf tasks, and effective strategies (Experiment 1)

Task	Mean	Median	SD	Skewness	Kurtosis
OSPAN	0.53	0.51	0.15	-0.14	1.3
RSPAN	0.51	0.49	0.16	0.50	0.52
LSPAN	0.88	0.91	0.09	-1.9	5.4
WSPAN	0.81	0.83	0.09	-0.43	-0.60
Locations	0.59	0.61	0.24	-0.24	-0.99
Letters	0.80	0.88	0.21	-1.4	1.6
OStrategy	0.30	0.20	0.30	0.85	-0.30
RStrategy	0.36	0.27	0.33	0.64	-0.83
LStrategy	0.28	0.13	0.35	1.0	-0.31
WStrategy	0.22	0.07	0.27	1.2	0.35

SD standard deviation, *Locations* locations test, *Letters* letters sets, *OStrategy* proportion of trials effective strategies were reported on the OSPAN task, *RStrategy* proportion of trials effective strategies were reported on the RSPAN task, *LStrategy* proportion of trials effective strategies were reported on the LSPAN task, *WStrategy* proportion of trials effective strategies were reported on the WSPAN task

first (WSPAN:  $M = 0.23$ ) and second simple span task (LSPAN:  $M = 0.25$ ),  $t(127) = 0.64$ ,  $p = 0.52$ . Finally, the reliabilities for all of the span tasks and strategy reports are reported along the diagonal in Table 2. For the span tasks, we computed reliability using Cronbach's alpha. For the strategy reports, we computed the proportion of effective strategies reported on odd and even trials for each participant, and then correlated these values across participants. To compute split-half reliability, the correlations were adjusted using the Spearman-Brown formula.

**Table 2** Zero-order correlations for span tasks, Gf tasks, and effective strategy use (Experiment 1)

Task	1	2	3	4	5	6	7	8	9	10
1 OSPAN	(0.86)									
2 RSPAN	<b>0.69</b>	(0.90)								
3 LSPAN	<b>0.30</b>	<b>0.45</b>	(0.85)							
4 WSPAN	<b>0.36</b>	<b>0.46</b>	<b>0.46</b>	(0.80)						
5 Locations	<b>0.21</b>	<b>0.23</b>	<b>0.34</b>	<b>0.20</b>	(0.63)					
6 Letters	<b>0.43</b>	<b>0.46</b>	<b>0.45</b>	<b>0.29</b>	<b>0.41</b>	(0.78)				
7 OStrategy	<b>0.35</b>	0.12	0.02	0.03	0.01	-0.03	(0.94)			
8 RStrategy	<b>0.23</b>	<b>0.31</b>	0.11	0.16	0.04	0.12	<b>0.31</b>	(0.96)		
9 LStrategy	0.15	0.16	0.15	-0.03	0.06	0.11	<b>0.22</b>	<b>0.40</b>	(0.97)	
10 WStrategy	<b>0.22</b>	0.16	0.10	0.13	0.01	0.00	<b>0.39</b>	<b>0.56</b>	<b>0.31</b>	(0.94)

*Locations* locations test, *Letters* letters sets, *OStrategy* the proportion of trials effective strategies were reported on the OSPAN task, *RStrategy* the proportion of trials effective strategies were reported on the RSPAN task, *LStrategy* the proportion of trials effective strategies were reported on the LSPAN task, *WStrategy* the proportion of trials effective strategies were reported on the WSPAN task

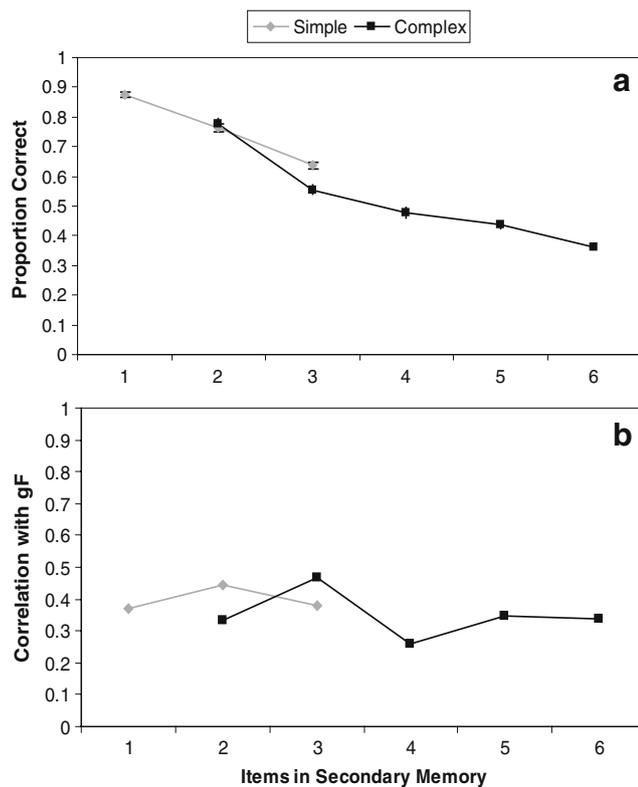
Correlations that are bolded are significant at  $p < 0.05$ . The reliability for each task is given in parentheses

## Performance on simple and complex span tasks

Collapsed across all set sizes, simple span performance ( $M = 0.84$ ,  $SE = 0.01$ ) was significantly higher than complex span performance ( $M = 0.52$ ,  $SE = 0.01$ ),  $t(103) = 27.10$ ,  $p < 0.001$ ,  $d = 2.90$  (Cohen, 1988). More important, performance also was compared after equating items retrieved from SM (Fig. 1a). This comparison was made with two items in SM (complex set size 3 & simple set size 6) and with three items (complex set size 4 & simple set size 7). With two items drawn from SM, performance on complex and simple span tasks was not significantly different,  $t < 1.0$ ; however, with three items from SM, simple span performance ( $M = 0.64$ ,  $SE = 0.01$ ) significantly exceeded that from complex span ( $M = 0.55$ ,  $SE = 0.02$ ),  $t = 4.72$ ,  $p < .001$ ,  $d = 0.52$ .

## Predicting Gf

The main analysis relevant to replicating Unsworth and Engle (2006) involved comparing the strength of the complex span-Gf correlation to that of the simple span-Gf correlation. Although the individual span tasks had similar relationships with the Gf tasks (Table 2), we further evaluated the relationships by using composite variables, which are better estimates of constructs because they capture common variance among the measures. Because performance on the OSPAN and RSPAN tasks was significantly related ( $r = 0.69$ ), we averaged the standardized scores (i.e., z-scores) of each complex task; we did the same for the two simple span tasks ( $r = 0.46$ ) and the two Gf tasks ( $r = 0.41$ ). The correlations between span and Gf composite measures, after equating the number of items in SM, are shown in Fig. 1b. The complex span-Gf correlation



**Fig. 1** Mean proportion correct on simple and complex span tasks (panel a). Average correlation between performance on the Gf tasks and simple or complex span tasks (panel b). Error bars represent the standard error of the means (Experiment 1)

( $r = 0.34$ ) was not significantly different from the simple span-Gf correlation ( $r = 0.44$ ) when two items were in SM,  $t(102) = 1.0$ ,  $p = 0.21$ , or when three items were in SM (complex-Gf = 0.47; simple-Gf = 0.38),  $t(102) < 1.0$ . Contrary to our predictions, simple span performance did not predict Gf performance better when items were retrieved from SM ( $r = 0.45$ ) versus PM only ( $r = 0.34$ ), but this effect approached significance,  $t(102) = 1.43$ ,  $p = 0.10$ . Interestingly, though, simple span performance accounted for a significant amount of unique variance in Gf performance (8.5%) after controlling for complex span performance. More specifically, simple span trials uniquely predicted significant Gf variance (3.4%) when items were retrieved from SM, but not when they were retrieved from PM (0.7%).

#### Span performance as a function of reported strategy use

We tested the contribution of strategy use at the group-level by comparing mean span performance on trials in which normatively effective strategies—imagery, sentence generation, and grouping—were reported to the mean span performance on trials in which normatively less effective strategies—reading and repetition—were reported (for the proportions of effective strategies reported on each task, see Appendix). On complex span tasks, the mean proportion of

correctly recalled items was significantly higher on trials in which subjects reported using effective strategies ( $M = 0.58$ ) than on trials in which they reported using less effective strategies (0.52),  $t(84) = 3.02$ ,  $p < 0.01$ ,  $d = 0.31$ . Moreover, on simple span trials that required retrieval from SM, performance was significantly higher when effective strategies were used (0.82) than when less effective strategies were used (0.73),  $t(78) = 5.41$ ,  $p < 0.001$ ,  $d = 0.66$ . However, on simple span trials that required retrieval from PM only, strategy use did not influence span performance ( $t < 1.0$ ), because performance was at ceiling (effective = 0.97 versus less effective = 0.96), which was expected given that retrieval from PM should be highly successful regardless of the strategy used (Baddeley & Ecob, 1970).

Another way to evaluate the contribution of strategies is at the individual level. To do so, we computed an across-subject correlation between the proportion of trials on which effective strategies were reported on each span task and the corresponding span performance (e.g., effective strategy use on the RSPAN task and RSPAN performance). These correlations (Table 2) yielded two important findings. First, across all four span tasks, the proportions of effective strategy use were significantly related ( $r_s = 0.22$ – $0.56$ ), indicating that subjects who were strategic on one span task were also strategic on the other span tasks. Second, the correlations between effective strategy use and span performance were significant for only the OSPAN ( $r = 0.35$ ) and RSPAN ( $r = 0.31$ ) tasks; effective strategy use did not predict LSPAN or WSPAN performance, even when only larger set sizes were examined (5–7;  $r = 0.14$ ,  $p = 0.10$ ).

#### Effective strategy use as a mediator

To assess whether effective strategy use is responsible for the span-Gf correlations reported above, we conducted hierarchical linear regressions separately for the complex and simple span tasks. These analyses compared the amount of variance that span performance accounted for in Gf performance before and after controlling for strategy use. More specifically, a given span composite was entered as a predictor of Gf on the first step. Then on the second step, the proportion of effective strategy use from that span composite was entered first as a predictor followed by the corresponding span composite (see Table 3).

Complex span performance accounted for 17% of the variance in Gf performance, both before and after controlling for effective strategy use. The same held for simple span, which accounted for 21% of the Gf variance before and after controlling for strategy use. (This latter effect is not surprising, of course, because simple span scores did not correlate significantly with simple span strategy use.) Finally, regression analyses were conducted with effective strategy use as a mediator of the simple span-Gf relationships for those trials in

**Table 3** Summary of hierarchical regression analyses for variables predicting criterion task performance (using composite variables)

Variable	$R^2$	$\beta$	$F$	Variable	$R^2$	$\beta$	$F$
Experiment 1							
Gf composite				Gf composite			
Step 1				Step 1			
Complex span	0.17	0.41*	20.42	Simple span	0.20	0.44*	24.48
Step 2	0.18 <sup>a</sup>			Step 2	0.20		
Strategy use <sup>b</sup>	0.01	0.04	0.17	Strategy use <sup>b</sup>	0.00	0.04	0.17
Complex span	0.17	0.44*	10.62	Simple span	0.20	0.44*	12.13
Simple span							
Step 1							
Complex span	0.27	0.52*	37.76				
Step 2	0.29						
Strategy use <sup>c</sup>	0.05	0.23*	5.51				
Complex span	0.24	0.50*	20.70				
Experiment 2							
Gf composite				Gf composite			
Step 1				Step 1			
Complex span	0.09	0.29*	13.94	Simple span	0.14	0.38*	24.38
Step 2	0.13			Step 2	0.16		
Strategy use <sup>b</sup>	0.02	0.13	2.31	Strategy use <sup>b</sup>	0.00	0.03	0.14
Complex span	0.11	0.37*	9.59	Simple span	0.16	0.40*	10.82
Simple span							
Step 1							
Complex span	0.28	0.53*	58.67				
Step 2	0.45						
Strategy use <sup>c</sup>	0.13	0.36*	7.86				
Complex span	0.33	0.58*	29.67				

\* $p < 0.001$ <sup>a</sup> Total amount of variance in criterion task performance accounted for by effective strategy use and span performance<sup>b</sup> Proportion of effective strategies reported on the particular span task<sup>c</sup> Combined proportion of effective strategies reported on complex and simple span tasks

which items were retrieved from only SM. Performance on large simple span set sizes (i.e., set sizes 5–7) and Gf tasks shared 20% of their variance and, again, effective strategy use did not account for any shared variance (see Table 3 for regression statistics). Thus, effective strategy use cannot explain the relationship between performance on span and Gf tasks.

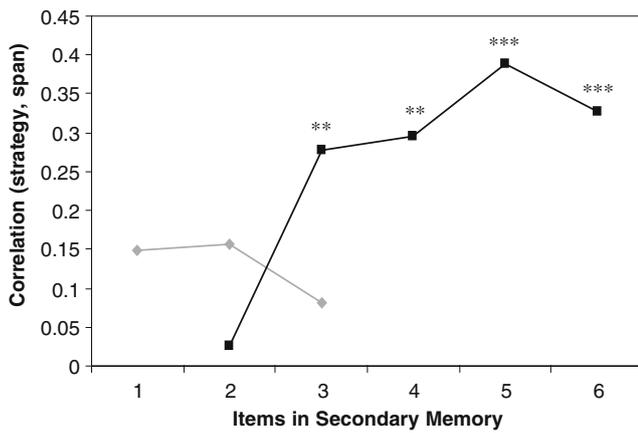
## Discussion

One goal of Experiment 1 was to evaluate the performance and predictive power of simple and complex span performance after equating items in SM. As predicted, simple and complex span performance was relatively similar when items in SM were equated, and both similarly predicted Gf performance. However, unlike Unsworth and Engle's

findings, simple span performance significantly predicted Gf performance before equating retrieval from SM.

Our main goal was to estimate the contribution of effective strategy use to performance on simple and complex span tasks. We predicted that effective strategy use would correlate with span performance when items were retrieved from SM. Contrary to this prediction, effective strategy use significantly predicted performance only on complex span tasks (Fig. 2). If effective strategies are more important to retrieval from SM than from PM (Baddeley & Ecob, 1970), then the non-significant strategy-simple span relationship may be due to the fact that simple span tasks only required that between one and three items be retrieved from SM, whereas complex span tasks required up to six items.

Results from the first experiment were limited to comparing performance on simple and complex span tasks at only two data points: when two and three items were retrieved from



**Fig. 2** Average correlation between effective strategy use and span performance as a function of the number of items in secondary memory. (Experiment 1)

SM. Thus, to better assess the contribution of effective strategies to simple and complex span performance, more points of comparison were included in the second experiment.

### Experiment 2

The purpose of Experiment 2 was to further estimate individual differences in strategy use on span tasks. To do so, we used methods similar to those in Experiment 1, except more set sizes were added to the simple span tasks, which ranged from two to nine, and set size 7 was excluded from complex span tasks. The new set sizes made it possible to compare span performance at five points: when one, two, three, four and five items were ostensibly retrieved from SM (i.e., simple set sizes 5–9; complex set sizes 2–6). Also, we added the RAPM as a third Gf measure.

### Methods

#### Subjects

One hundred fifty-two undergraduates from introductory psychology courses at Kent State University participated to complete a course requirement.

#### Materials

**Complex span tasks** OSPAN and RSPAN were the same as those in Experiment 1. Each consisted of 15 experimenter-paced trials ranging from two to six to-be-remembered words. After the final trial on each task, subjects completed set-by-set strategy reports.

**Simple span tasks** WSPAN and LSPAN were conducted in the same manner as in Experiment 1 except set sizes ranged from two to nine. As with the complex span tasks, subjects completed set-by-set strategy reports following the final trial on each task.

**Fluid intelligence (Gf) tasks** The letter sets and locations test were conducted in a similar manner as in Experiment 1 except subjects had eight minutes to complete as many trials as possible. Additionally, subjects completed the same 18 trials from the RAPM used by Stanovich and Cunningham (1993). In this task, a 3 × 3 matrix display was presented on the computer screen, with eight geometric figures and the ninth figure (i.e., the bottom, right-hand figure) missing. Subjects selected one of eight choices of figures to complete the pattern. Subjects had 15 minutes to complete as many displays as possible. Performance was scored as proportion correct.

#### Procedure

Data collection for this experiment took place as a part of a larger study consisting of 8 one-hour sessions with one-week delays between each session. The order in which tasks were completed was the same across subjects: demographics questionnaire (session 1), LSPAN (session 1), locations test (session 1), OSPAN (session 1), RSPAN (session 2), WSPAN (session 3), letter sets test (session 5), and finally, RAPM (session 7).

**Table 4** Descriptive statistics for span tasks, Gf tasks, and effective strategies (Experiment 2)

Task	Mean	Median	SD	Skewness	Kurtosis
OSpan	0.65	0.66	0.12	0.05	-0.31
RSpan	0.63	0.63	0.13	-0.40	1.3
LSpan	0.85	0.85	0.07	-0.66	1.1
WSpan	0.75	0.76	0.10	-0.44	4.6
RAPM	0.37	0.36	0.18	0.69	-0.02
Locations	0.58	0.69	0.35	-1.4	1.5
Letters	0.69	0.72	0.15	-0.94	0.65
OStrategy	0.22	0.13	0.26	1.0	0.37
RStrategy	0.25	0.13	0.28	1.1	0.34
LStrategy	0.24	0.19	0.26	0.95	0.19
WStrategy	0.24	0.08	0.30	1.1	0.21

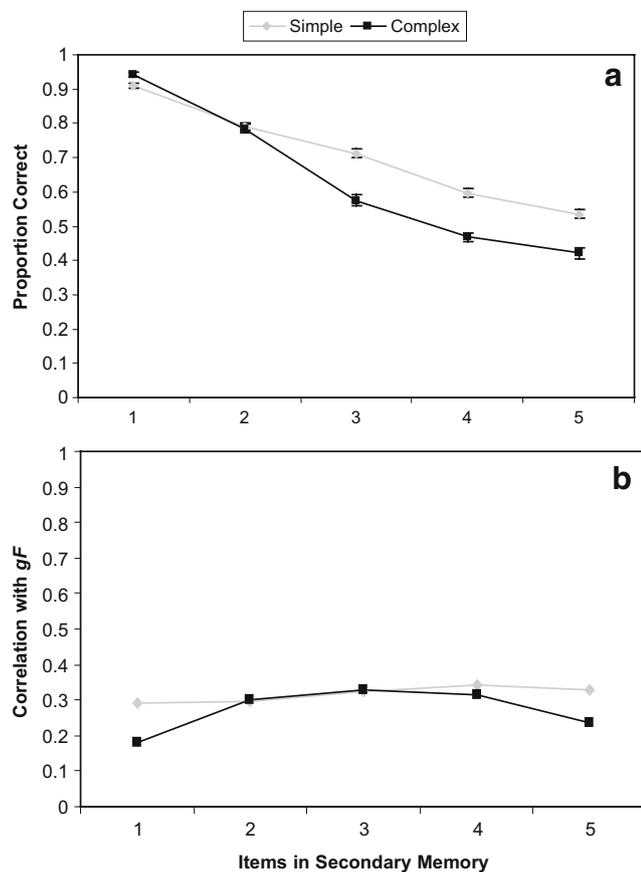
SD Standard deviation, RAPM Raven Advanced Progressive matrices, Locations locations test, Letters letters sets, OStrategy proportion of trials effective strategies were reported on the OSPAN task, RStrategy proportion of trials effective strategies were reported on the RSPAN task, LStrategy proportion of trials effective strategies were reported on the LSPAN task, WStrategy proportion of trials effective strategies were reported on the WSPAN task

## Results

Analyses of span performance and strategy use were conducted as in [Experiment 1](#). To help interpret them, [Table 4](#) shows descriptive statistics and the proportion of effective strategy use on each task. Note that, as compared to [Experiment 1](#), removing the largest complex span set size resulted in an increase in overall OSPAN and RSPAN scores, and including larger simple span sets resulted in a decrease in overall LSPAN and WSPAN scores.

### Performance on simple and complex span tasks

Collapsed across all set sizes, simple span performance ( $M = 0.81$ ,  $SE = 0.01$ ) was significantly higher than complex span performance ( $M = 0.64$ ,  $SE = 0.01$ ),  $t(149) = 21.76$ ,  $p < 0.001$ ,  $d = 1.76$ . This effect remained significant after equating the number of items retrieved from SM ([Fig. 3a](#)) with two exceptions: (1) when one item was retrieved from SM, performance was higher on complex versus simple span and (2) when two items were retrieved from SM (i.e.,



**Fig. 3** Mean proportion correct on simple and complex span tasks ([panel a](#)). Average correlation between performance on the Gf tasks and simple or complex span tasks ([panel b](#)). Error bars represent the standard error of the means ([Experiment 2](#))

complex set size 3 versus simple set size 6, span performance was statistically equivalent.

### Predicting Gf

[Table 5](#) presents correlations among scores on all span tasks. To evaluate whether complex and simple span similarly predicted Gf, we again created z-score composite variables for each construct (for OSPAN and RSPAN,  $r = 0.65$ ; for LSPAN and WSPAN,  $r = 0.53$ ; among the three Gf tasks,  $r_s = 0.43$ – $0.48$ ). As in [Experiment 1](#), we calculated the span-Gf correlations as a function of the number of items retrieved from SM ([Fig. 3b](#)). Replicating [Unsworth and Engle \(2006\)](#), complex and simple span similarly predicted Gf; the complex-Gf correlation was not significantly different from the simple-Gf correlation when one, two, three, four, or five items were retrieved from SM (all  $t_s < 1.0$ ,  $p_s > 0.05$ ). Although it approached significance, simple span performance did not predict Gf performance better when items were retrieved from SM ( $r = 0.38$ ) versus PM only ( $r = 0.24$ ),  $t(146) = 1.59$ ,  $p = 0.09$ . However, as in [Experiment 1](#), simple span performance predicted Gf performance beyond complex span performance. In fact, simple span trials uniquely predicted significant Gf variance (5.5%) when items were retrieved from SM, but not when they were retrieved from PM.

### Span performance as a function of reported strategy use

The contribution of effective strategy use at the group-level was assessed by comparing span performance when effective versus less effective strategies were reported (for the proportions of effective strategies reported on each task, see [Appendix](#)). On the complex span tasks, recall was significantly higher on trials in which subjects reported using effective strategies ( $M = 0.69$ ) than on those they reported using less effective strategies (0.64),  $t(106) = 2.40$ ,  $p < 0.01$ ,  $d = 0.36$ . Moreover, simple span scores were significantly higher when effective versus less effective strategies were used on trials that required retrieval from SM (effective = 0.76 versus less effective = 0.70,  $t(101) = 4.05$ ,  $p < 0.001$ ,  $d = 0.41$ ) as well as for trials that required retrieval from PM (effective = 0.98 versus less effective = 0.96,  $t(89) = 2.29$ ,  $p < 0.05$ ,  $d = 0.15$ ).

We also examined effective strategy use at the individual level by correlating the performance and proportion of effective strategies reported on a span task (e.g., OSPAN performance and proportion of effective strategy use on the OSPAN task, see [Table 5](#)). Again, the proportions of effective strategy use on all span tasks were significantly correlated ( $r_s = 0.21$ – $0.48$ ) indicating that subjects who were more strategic on one span task also were generally more strategic on others. Moreover, the correlations between the proportion

**Table 5** Zero-order correlations for span tasks, Gf tasks, and effective strategy use (Experiment 2)

Task	1	2	3	4	5	6	7	8	9	10	11
1 OSPAN	(0.83)										
2 RSPAN	<b>0.65</b>	(0.85)									
3 LSPAN	<b>0.43</b>	<b>0.57</b>	(0.82)								
4 WSPAN	<b>0.36</b>	<b>0.53</b>	<b>0.52</b>	(0.88)							
5 RAPM	0.10	<b>0.28</b>	<b>0.23</b>	<b>0.18</b>	(0.68)						
6 Locations	0.13	<b>0.30</b>	<b>0.30</b>	<b>0.30</b>	<b>0.44</b>	(0.61)					
7 Letters	<b>0.21</b>	<b>0.43</b>	<b>0.35</b>	<b>0.41</b>	<b>0.43</b>	<b>0.48</b>	(0.89)				
8 OStrategy	<b>0.25</b>	<b>0.26</b>	0.05	<b>0.17</b>	<b>0.25</b>	-0.03	0.09	(0.91)			
9 RStrategy	<b>0.25</b>	<b>0.38</b>	0.07	0.10	<b>0.33</b>	0.09	<b>0.16</b>	<b>0.48</b>	(0.91)		
10 LStrategy	<b>0.18</b>	0.11	<b>0.19</b>	-0.05	<b>0.25</b>	0.11	<b>0.06</b>	<b>0.33</b>	<b>0.38</b>	(0.90)	
11 WStrategy	0.12	0.06	0.12	0.00	<b>.19</b>	0.04	0.04	<b>0.21</b>	<b>0.47</b>	<b>0.36</b>	(0.95)

RAPM Raven Advanced Progressive matrices, Locations locations test, Letters letters sets, OStrategy proportion of trials effective strategies were reported on the OSPAN task, RStrategy proportion of trials effective strategies were reported on the RSPAN task, LStrategy proportion of trials effective strategies were reported on the LSPAN task, WStrategy proportion of trials effective strategies were reported on the WSPAN task  
Correlations that are bolded are significant at  $p < 0.05$ . The reliability for each task is given in parentheses

of effective strategy use and span performance were significant for OSPAN ( $r = 0.25$ ), RSPAN ( $r = 0.38$ ), and LSPAN ( $r = 0.19$ ), but not for WSPAN ( $r = 0.00$ ).

Effective strategy use as a mediator

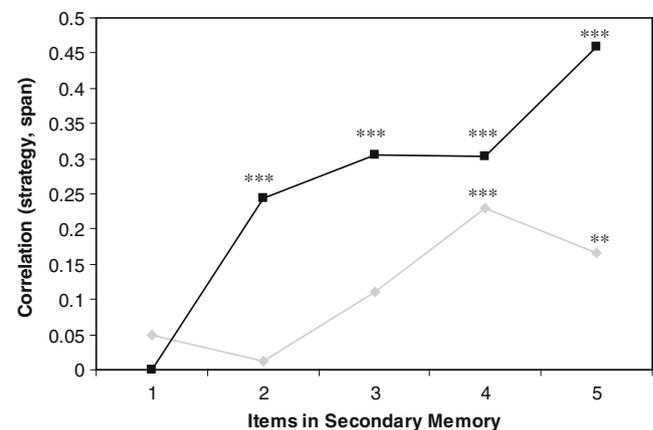
Finally, we conducted hierarchical linear regression analyses to assess whether strategy use mediated span-Gf relationships. Before controlling for strategy use, complex span and Gf tasks shared 9% of their variance, whereas after controlling for strategy use, they shared 11% (see Table 3). The same analyses were conducted for the simple span tasks but only for the trials that required retrieval from SM (i.e., set sizes 5–9). In step 1, simple span and Gf shared 14% of the variance, whereas after controlling for effective strategy use on simple span tasks, they shared 16%. Thus, effective strategy use suppressed both the complex-Gf and the simple-Gf relationships. Although these suppressor effects were not significant, other researchers have reported similar, and statistically significant, effects (Turley-Ames & Whitfield, 2003).

Discussion

Previous research has proposed that span performance predicts Gf performance when the tasks require retrieval of information from SM. Evidence supporting this claim was reported by Unsworth and Engle (2006), who found that when simple span performance requires retrieval from SM, it predicted Gf performance significantly and as strongly as did complex span performance. In the current

experiment, we also compared complex and simple span performance and their predictive power. Replicating the results from Unsworth and Engle (2006) and Experiment 1, complex and simple span performance similarly predicted Gf performance after equating retrieval from SM. However, overall simple span performance predicted Gf performance before equating retrieval from SM.

Experiment 2 further evaluated whether effective strategy use influences retrieval from SM and, in so doing, contributes to the relationship between span and Gf. Overall, span performance was higher when effective strategies were reported, indicating that individual differences in reported strategy use predicted span performance. However, as in Experiment 1, effective strategy use (i.e.,



**Fig. 4** Average correlation between effective strategy use and span performance as a function of the number of items in secondary memory. (Experiment 2)

imagery, sentence generation, and grouping) had more of an impact on complex than simple span (Table 4 and Fig. 4). Finally, although strategy use predicted span performance, it was not responsible for complex span-Gf or simple span-Gf relationships. In fact, after controlling for rates of effective strategy use, the relationships increased numerically, but not significantly.

## General discussion

The main goals of the current experiments were to evaluate whether effective strategy use at encoding (1) is similarly related to complex and simple span performance and (2) mediates the associations between span and Gf. Pertaining to the first goal, we evaluated whether effective strategies (e.g., imagery) were more important for retrieval from SM than from PM. Results supported previous research (Baddeley & Ecob, 1970) in that effective encoding strategies were more essential to the retrieval of information from SM than PM. At the group-level, both simple and complex span performance was higher when effective (versus less effective) strategies were used. However, at the individual level, effective strategies were more influential to complex versus simple span performance even after equating retrieval from SM (Figs. 2 and 4), which we discuss later. Note that we focused on encoding strategies in the current experiments, but other strategies can be used on these span tasks; for instance, covert retrieval practice has been shown to differentiate performance on complex and simple span tasks (McCabe, 2008; *in press*), and thus, may account for the span-Gf relationships.

As for the second goal, recall that Unsworth and Engle (2007) argued that performance on span and Gf tasks are related, in part, because both require a cue-delimited search of SM. We designed the current experiments to examine whether effective encoding strategies represented one method of improving the SM search. However, the results demonstrated that, whereas effective strategies did improve span performance, they did not mediate the span-Gf relationships.

We acknowledge that the span-Gf correlations that are typically reported in this literature are stronger than those we observed in Experiment 2 (see Kane, Hambrick, & Conway, 2005), thus providing less covariance for strategy use to explain. However, the magnitude of these correlations is not likely due to our choice in tasks because the lowest observed correlation was between OSPAN and RAPM, which are standard measures of span and Gf, respectively (Conway et al., 2002; Engle et al., 1999; Kane et al., 2004; Mogle et al., 2008). Although these two measures were not significantly related to one another, they both demonstrated convergent validity (i.e., OSPAN and RSPAN were correlated; RAPM, locations test and letters

sets were correlated). Further, performance on the other two Gf tasks—locations and letters—was significantly related to both complex and simple span performance. In both experiments, composite variables were used to reduce task-specific variance and to capture the common variance between the span tasks and between the Gf tasks. When composite variables were created, effective strategy use did not mediate the observed span-Gf relationships. Thus, other individual differences factors beyond strategy use, such as processing speed, attention control, or integrity of semantic networks, must influence one's ability to selectively and successfully search and retrieve information in SM, and future research should explore these alternatives.

## Relationship between span performance and strategy use

Results from the current experiments are relevant to a hypothesis made about complex and simple span performance in previous research. Specifically, Turner and Engle (1989) claimed that performance on simple span tasks does not predict higher-order cognition because they introduce extraneous variance due to strategy use. That is, strategies were thought to reduce the correlation between simple span and reading comprehension, whereas the interleaving tasks (e.g., reading sentences or solving equations) on complex span tasks were thought to prevent, or at least minimize, the use of such strategies. Contrary to this hypothesis, we found that effective strategy use was more highly related to complex span performance in both experiments. Although effective strategy use predicted simple span performance at large set sizes in Experiment 2 (e.g., when 4 and 5 items were retrieved from SM), in general, strategies were less important to simple span.

But is the effectiveness of a strategy constant, or might effectiveness vary with the particular span task being performed? In the current experiments, we considered imagery, sentence generation, and grouping to be normatively effective strategies because they are associated with higher performance in complex span than are passive reading or repetition (Bailey et al., 2009; Dunlosky & Kane, 2007). However, previous research has suggested that repetition may be a relatively effective strategy in simple span tasks (Engle et al., 1992; Logie et al., 1996; Turley-Ames & Whitfield, 2003). Because most items can be maintained in PM via rehearsal (i.e., simple span tasks have no secondary task to disrupt this rehearsal process), individuals who use repetition perform relatively well. To assess this possibility, we examined whether repetition was an effective strategy by comparing performance on trials in which repetition was reported with performance on trials in which normatively effective strategies (i.e., imagery, sentence generation, and grouping) were reported. In both

experiments, performance at each set size of complex span tasks was significantly higher when individuals reported using normatively effective strategies. As expected, for the simple span tasks, we found that repetition and the normatively effective strategies produced similar performance at smaller set sizes (set sizes 3–5 in [Experiment 1](#); set sizes 2, 3, 5, & 6 in [Experiment 2](#)). By contrast, at larger set sizes, effective strategies were associated with significantly higher performance than was repetition.

Thus, when analyzed across items, repetition never outperformed normatively effective strategies on the simple span task; nevertheless, individual differences in the use of repetition may predict individual differences in span performance. To explore this issue, we examined the relationship between strategy and simple span performance when subjects reported using the repetition strategy at each set size (similar to the correlations in [Figs. 2 and 4](#)). Repetition was not related to simple span performance in either experiment. In [Experiment 1](#), the correlations ranged from 0.05 to 0.16, and in [Experiment 2](#), they ranged from -0.01 to 0.17. Moreover, simple span performance was not significantly different when subjects reported using passive reading ( $E1 = 0.83$ ;  $E2 = 0.82$ ) and repetition ( $E1 = 0.83$ ;  $E2 = 0.78$ ), but performance was significantly higher when they reported using effective strategies ( $E1 = 0.88$ ;  $E2 = 0.87$ ). We found the same pattern of results for complex span tasks in [Experiments 1 and 2](#). In general, then, strategies such as imagery, sentence generation, and grouping were more effective on both simple and complex span tasks. Most pertinent to the PM-SM issue, repetition was effective for actively maintaining items in PM, whereas more elaborate strategies allowed for the most successful retrieval of items from SM. Given the similarity of the profiles of strategy use for simple and complex span tasks, we find it unlikely that differences in strategy use alone will completely explain why strategy use did not show the same relationship with complex and simple span tasks. One potential explanation for this effect is that individuals differed in PM capacity. For instance, on simple span sets of five, some subjects may be retrieving one item from SM whereas others may be able to keep all five items active in PM. If strategies are more important for retrieval from SM than PM, then variability in PM capacity may have added noise to the correlation between effective strategy use and simple span performance. However, we found that strategies did not benefit simple span performance until set sizes eight and nine ([Experiment 2](#)), and even these significant correlations were not strong ( $r_s = 0.26$  and  $0.17$ , respectively). If the lack of a relationship between simple span and strategy use was due to variability in PM capacity, which seems to range between three and five (Cowan, 2001), then span-strategy correlations for set sizes  $\geq 6$  should have been stronger and similar to those observed for the complex span tasks ( $r_s = 0.28$ – $0.46$ ).

Instead, one must consider the differences between complex and simple span tasks that could lead to the differential contribution of strategy use. One difference between the two, which is relevant to Unsworth and Engle's (2006, 2007) hypothesis, is the proportion of to-be-remembered items retrieved from SM. For large, simple span sets, some items are ostensibly retrieved from SM and others from PM. If a subject uses imagery (or another effective strategy) on a simple span set of eight, the strategy may facilitate retrieval of the four items from SM, but it may not be as beneficial to retrieving the four items from PM. In this example, effective strategies only help with retrieval of approximately 50% of the items on a simple span task (and far fewer than 50% of simple-span trials consist of such long lists). Conversely, for all complex span set sizes, all but one item is presumably retrieved from SM. If a subject uses imagery on a complex span set size of five, it may facilitate retrieval of the four items from SM. In this example, effective strategies help with retrieval of 80% of the items. Note that in these examples, the *number* of items retrieved from SM is equated on simple and complex span tasks, but the *proportion* of items is higher on complex span tasks. Thus, effective strategy use may be more important in predicting complex span versus simple span performance because complex span tasks consist of a higher proportion of items being retrieved from SM.

Preliminary evidence for this idea is found in [Fig. 4](#). For complex span tasks, as the proportion of items retrieved from SM increases, the correlation between strategy use and span performance also increases (albeit not linearly). Even more compelling is evidence from simple span tasks. That is, when a higher proportion of items are retrieved from PM than from SM (i.e., simple span set sizes 2–7), the correlation between effective strategy use and simple span performance is low and non-significant ( $r_s = 0.01$ – $0.11$ ). However, when a higher proportion of items are retrieved from SM than from PM (i.e., simple span set sizes 8 & 9), correlations are larger and significant ( $r_s = 0.23$ – $0.17$ ).

Finally, in accordance with Turner and Engle's claim, our results do suggest that strategies introduce some extraneous variance. After controlling for rates of effective strategy use, span-Gf correlations became stronger (although not significantly so). The effect of strategy use suppressing span-cognition relationships also has been reported by Turley-Ames and Whitfield (2003), who found that the correlation between span and reading comprehension was stronger after experimentally removing variance due to individual differences in strategy use (see also Friedman & Miyake, 2004). Thus, because strategies somewhat can reduce span-cognition correlations, it is important to measure and control for strategy production on span tasks. To measure strategy production, we recommend using the current set-by-set reports because they (1) are reliable and valid, (2) are easy

to collect and to objectively analyze and (3) have greater specificity than more global-report measures (for further details, see Dunlosky & Kane, 2007).

## Conclusion

In general, results of the current studies indicate that simple span tasks can predict Gf as well as do complex span tasks do (as in Unsworth & Engle, 2006). Further, they explain why performance on complex and simple span tasks is more similar after equating the number of

items retrieved from SM. That is, both complex and simple span performance are significantly higher on trials in which effective strategies are used as compared to the trials in which less effective strategies are used. However, individual differences in strategy use (1) were more highly related to complex versus simple span performance and (2) were not responsible for the span-Gf relationships. These findings suggest that strategy use may be one factor that increases the efficiency of the search through SM during a complex span task, but may not have a commensurate effect on SM search during simple span or Gf tasks.

## Appendix

**Table 6** Proportion of reported strategy use for a given strategy

Span Task	Read	Repetition	Imagery	Sentence	Grouping	Other
Experiment 1						
OSPAN	.36 (.03)	.29 (.03)	.06 (.01)	.12 (.02)	.12 (.02)	.05 (.01)
RSPAN	.24 (.03)	.37 (.04)	.10 (.02)	.13 (.03)	.13 (.02)	.03 (.02)
LSPAN	.19 (.03)	.50 (.04)	.16 (.03)	.02 (.01)	.10 (.02)	.03 (.01)
WSPAN	.43 (.03)	.33 (.03)	.07 (.02)	.07 (.02)	.08 (.01)	.02 (.01)
Experiment 2						
OSPAN	.41 (.03)	.34 (.03)	.07 (.01)	.09 (.01)	.07 (.01)	.02 (.01)
RSPAN	.33 (.03)	.39 (.03)	.10 (.02)	.07 (.01)	.08 (.01)	.03 (.01)
LSPAN	.25 (.02)	.48 (.03)	.09 (.02)	.03 (.01)	.12 (.01)	.03 (.01)
WSPAN	.37 (.03)	.38 (.03)	.09 (.02)	.05 (.01)	.10 (.02)	.01 (.01)

Note. Standard errors of the means are reported in parentheses.

## References

- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131*, 30–60.
- Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77*, 1698–1716.
- Baddeley, A. D., & Ecob, J. R. (1970). Simultaneous acoustic and semantic coding in short-term memory. *Nature, 227*, 288–289.
- Bailey, H., Dunlosky, J., & Kane, M. J. (2008). Why does working memory span predict complex cognition? Testing the strategy-affordance hypothesis. *Memory & Cognition, 36*, 1383–1390.
- Bailey, H., Dunlosky, J., & Hertzog, C. (2009). Does differential strategy use account for age-related deficits in working-memory performance? *Psychology and Aging, 24*, 82–92.
- Bayliss, D. M., Jarrold, C., Baddeley, A. D., & Leigh, E. (2005). Differential constraints on the working memory and reading abilities of individuals with learning difficulties and typically developing children. *Journal of Experimental Child Psychology, 92*, 76–99.
- Bethell-Fox, C. E., Lohman, D. F., & Snow, R. E. (1984). Adaptive reasoning: Componential and eye movement analysis of geometric analogy performance. *Intelligence, 8*, 205–238.
- Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114*, 539–576.
- Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55*, 627–652.
- Cantor, J., Engle, R., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? *Intelligence, 15*, 229–246.
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97*, 404–431.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Colom, R., Rebollo, R., Abad, F. J., & Shih, P. C. (2006). Complex span tasks, simple span tasks, and cognitive abilities: A reanalysis of key studies. *Memory & Cognition, 34*, 158–171.
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D., & Minkoff, S. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163–183.
- 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. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, 24, 87–185.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.
- Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, 3, 422–433.
- Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *The Quarterly Journal of Experimental Psychology*, 60, 1227–1245.
- Eksstrom, R. B., French, J. W., Harman, H. 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., & 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 prefrontal 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.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128, 309–331.
- Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, 208, 1181–1182.
- 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.
- Kaakinen, J. K., & Hyönä, J. (2007). Strategy use in the reading span test: An analysis of eye movements and reported encoding strategies. *Memory*, 15, 634–646.
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comments on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, 131, 66–71.
- 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.
- La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 1118–1133.
- Logie, R. H., Della Sala, S., Laiacona, M., Chalmers, P., & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory & Cognition*, 24, 305–321.
- Marschark, M., Richman, C. L., Yuille, J. C., & Hunt, R. R. (1987). The role of imagery in memory: On shared and distinctive information. *Psychological Bulletin*, 102, 28–41.
- McCabe, D. P. (2008). The role of covert retrieval in working memory span tasks: Evidence from delayed recall tests. *Journal of Memory and Language*, 58, 480–494.
- McCabe, D. P. (in press). The influence of complex working memory span task administration methods on prediction of higher-level cognition and metacognitive-control of response times. *Memory & Cognition*.
- McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition*, 29, 10–17.
- Mitchum, A. L., & Kelley, C. M. (in press). Solve the problem first: Constructive solution strategies can influence the accuracy of retrospective confidence judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Mogle, J. A., Lovett, B. J., Stawski, R. S., & Sliwinski, M. J. (2008). What's so special about working memory?: An examination of the relationships among working memory, secondary memory, and fluid intelligence. *Psychological Science*, 19, 1071–1077.
- Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review*, 115, 544–576.
- Raven, J. C., Raven, J. E., & Court, J. H. (1998). *Progressive matrices*. Oxford, England: Oxford Psychologists Press.
- 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, 597–614.
- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, 125, 4–27.
- St. Clair-Thompson, H., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology*, 30, 203–219.
- Stanovich, K. E., & Cunningham, A. E. (1993). Where does knowledge come from? Specific associations between print exposure and information acquisition. *Journal of Educational Psychology*, 85, 211–229.
- Tehan, G., Hendry, L., & Kocinski, D. (2001). Word length and phonological similarity in simple, complex, and delayed serial recall tasks: Implications for working memory. *Memory*, 9, 333–348.
- Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language*, 49, 446–468.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127–154.
- Unsworth, N., Brewer, G. A., & Spillers, G. J. (2009). There's more to the working memory capacity-fluid intelligence relationship than just secondary memory. *Psychonomic Bulletin & Review*, 16, 931–937.
- Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language*, 54, 68–80.
- Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, 133, 1038–1066.
- Unsworth, N., & Spillers, G. J. (2010). Variation in working memory capacity and episodic recall: The contributions of strategic encoding and contextual retrieval. *Psychonomic Bulletin & Review*, 17, 200–205.
- Vigneau, F., Caissie, A. F., & Bors, D. A. (2005). Eye movement analysis demonstrates strategic influence on intelligence. *Intelligence*, 34, 261–272.