Validating running memory span: Measurement of working memory capacity and links with fluid intelligence

JAMES M. BROADWAY AND RANDALL W. ENGLE
Georgia Institute of Technology, Atlanta, Georgia

We manipulated running memory span tasks to examine effects on recall and relations with criterion measures of working memory capacity and general fluid intelligence. The goal of the manipulations was to limit or enhance opportunities for active input processing and response preparation in advance of test. We manipulated presentation rate in Experiment 1. Recall was higher at slow than at fast rates, but correlations with criterion measures were much the same across rate conditions. In Experiment 2, we manipulated the time at which the number of items to report was made known to the participants. They were given that information in advance (precue) or at test (postcue). Recall scores and correlations with criterion measures were much the same across cuing conditions. We conclude that running memory span provides valid measurement of working memory capacity that is predictive of higher order cognition across a wide range of conditions.

Using a combination of experimental and differential approaches, psychologists have sought to isolate components of a general cognitive architecture that might explain the fact that individual differences in intellectual abilities are strongly interrelated across a wide range of domains, such as length comparison, reading comprehension, and abstract reasoning (Deary, 2000; Sternberg, 1979). Working memory (Baddeley & Hitch, 1974) has been identified as a strong candidate for this explanatory role (Conway et al., 2005; Deary, 2000; Engle, Tuholski, Laughlin, & Conway, 1999). *Working memory* is defined theoretically as the work space of the mind, a system for accessing goal-relevant information as needed to support complex cognition. This theoretical definition is supported by widespread observations (for a review, see Engle & Kane, 2004). Working memory capacity is important for understanding phenomena across a range of applied psychological disciplines, including clinical, social, educational, and neuropsychological fields, and is an important executive function or frontal lobe ability (Conway et al., 2005; Kane & Engle, 2002; Miyake et al., 2000).

Working memory is a complex system, comprised of multiple component processes for maintaining, accessing, manipulating, and coordinating information (Oberauer, Süß, Wilhelm, & Wittman, 2003; Unsworth & Engle, 2007). Therefore, researchers have sought to isolate hypothetical working memory processes specifically responsible for observed links between working memory capacity and complex cognition. Current theories differ in detail and emphasis, but there is a growing consensus that executive, attention-related processes are critical to these links

(Cowan et al., 2005; Engle & Kane, 2004; Kane et al., 2004; Lépine, Barrouillet, & Camos, 2005; Unsworth & Engle, 2007; Unsworth, Redick, Heitz, Broadway, & Engle, 2009).

In the present work, we sought to isolate the working memory processes responsible for links with general fluid intelligence by manipulating basic characteristics of a classic short-term memory task: running memory span (Pollack, Johnson, & Knaff, 1959). Our goal was to limit the contributions to performance from a hypothetical working memory process and to observe whether correlations with criterion tests of mental abilities were thereby affected (Deary, 2000).

Running Memory Span

Participants in running memory span tasks report the last n items from lists that are $m + n$ items long. In this article, the last n items that are to be reported are called *targets*. The $m + n$ items that are presented are called *inputs*. *Partial recall trials* ($m > 0$) occur when more items are shown than are to be recalled. In contrast, *whole recall trials* ($m = 0$) occur when the same number of items is shown as is to be recalled (Mukunda & Hall, 1992). Including whole recall trials in running memory span is assumed to discourage the participants from selectively ignoring early inputs on the basis of expectations that these items will not be tested (catch trials; Morris & Jones, 1990). A broad distinction between active and passive approaches to performing running memory span tasks has been influential since the pioneering work of Hockey (1973; see also Bunting, Cowan, & Saults, 2006; Elosúa

J. M. Broadway, jbroadway@gatech.edu

& Ruiz, 2008; Palladino & Jarrold, 2008; Ruiz, Elosúa, & Lechuga, 2005) and informs the organization of the present research.

A person might try to prepare responses in advance of a test, perhaps by rehearsing and grouping targets (active input processing). Alternatively, a person might wait to prepare responses until the time of test, after inputs have terminated (passive input processing). A common view of running memory span corresponds to active input processing: Performance reflects working memory updating in the form of dynamic phonological rehearsal and grouping in advance of the test (Conway et al., 2005; Engle & Kane, 2004; Friedman et al., 2006; Morris & Jones, 1990). Working memory updating, as measured by running memory span, has been identified as an important executive function that correlates strongly with fluid intelligence (Friedman et al., 2006). A contemporary alternative view of running memory span corresponds to passive input processing: Performance reflects the number of items that can be extracted from sensory memory into the focus of attention at the time of the test (Cowan, 2001; Cowan et al., 2005).

It is widely supposed that running memory span elicits different strategies from participants as a function of variables such as presentation rate (Bunting et al., 2006; Hockey, 1973; Postle, 2003). Consistent with this perceived sensitivity to conditions, running memory span has been adapted to measure widely differing processes such as working memory updating and maintenance in the focus of attention. Whether active or passive input processing approaches were induced, strong relations to higher order cognition were observed in separate studies (Cowan et al., 2005; Friedman et al., 2006). We thought that evidence for the separation of these two important working memory constructs might be obtained within a single sample, by examining responses in conditions favorable for their distinct measurement. Alternatively, boundaries might be established for the range of conditions under which running memory span can yield valid measurement of working memory capacity.

Overview of Experiments

We arranged conditions in two experiments that either supported or limited active input processing, in order to distinguish this approach from passive input processing. In addition, our parametric exploration of running memory span enabled us to assess the range of conditions under which the task can yield valid measurement of working memory capacity. To process inputs actively and prepare responses in advance of test, it would help to have relatively long, empty intervals between successive inputs. In Experiment 1, we examined the effects of presentation rate. Participants encoded inputs at a fast rate or at a slow rate (one item every 500 msec or every 2,000 msec). To process inputs actively, it would help to know in advance how many items to report. In Experiment 2, participants learned how many items to report before inputs were presented or else at the time of test (precue or postcue).

EXPERIMENT 1

Predictions

We expected to dissociate active and passive input processing approaches to running memory span and to observe different relations with reference abilities of working memory capacity and general fluid intelligence. We expected that responses across running span conditions would not identify the same individual differences in these reference abilities. Alternatively, we expected to find boundaries on the range of conditions under which running memory span yields valid working memory capacity measurement.

Method

Participants. A total of 89 individuals (57 women and 32 men) participated in Experiment 1. The participants in both experiments were recruited from the Atlanta community through advertisements or through the undergraduate research pool; were between the ages of 18 and 35 years, inclusive (Experiment 1, $M = 23.6$ years); reported normal or corrected-to-normal hearing and vision; were fluent in English; and were compensated with a check or partial course credit.

Procedure. All tasks were programmed in E-Prime experimental software (Schneider, Eschman, & Zuccolotto, 2002) and administered to participants seated at a comfortable distance from a personal computer. The participants performed all tasks alone in a sound-attenuated room. A trained research assistant entered the room to begin each program when summoned by the participant's press of an intercom button. The research assistant stayed with the participant during practice trials to ensure that task requirements were understood and left at the start of real trials. The participants were made aware that they would be monitored by closed-circuit television when the research assistant was absent from the room.

Individuals participated in Experiment 1 after first visiting the lab to perform three complex-span tasks—operation span (OSpan), reading span (RSpan), and symmetry span (SymSpan)—and one general fluid intelligence task—a subset of twelve problems from *Raven's standard progressive matrices* (RavSetA; Raven, Raven, & Court, 1998)—in a 60-min session. In the OSpan task (Unsworth, Heitz, Schrock, & Engle, 2005), the participants solved simple math equations between presentations of to-be-remembered letters (F, H, J, K, L, N, P, Q, R, S, T, or Y). After between three and seven of these paired events (set size randomly determined), a cue prompted the participants to report the letters in order. The participants responded by clicking on the cells of a 4 × 3 grid displaying the 12 letters of the pool of items from which the to-be-remembered letters were sampled across trials. The RSpan task was identical to the OSpan task, except that the participants judged whether a sentence made sense in between encoding the to-be-remembered letters. The participants in the SymSpan task (Unsworth et al., 2009) judged whether black-and-white images were symmetrical in between encoding the location in a 4 × 4 grid in which a red square appeared. After between two and five of these paired events (randomly determined), a cue prompted the participants to report the square locations in order. Responses were made by clicking on the cells of a blank 4 × 4 grid in the correct order. There were three trials for each set size in each complex-span task.

The responses in the span tasks were constrained to forward order. To maintain the correct serial position of recalled items with respect to their position in the exposed list, the participants were instructed to click on a blank option for each item that could not be recalled (i.e., strict serial position scoring was applied to the response sequences). One point was assigned for each item correctly chosen in the correct serial position. For example, if four items—J, K, L, and T—were to be reported, responding "J blank L T" would receive 3 points, but responding "J L T" would receive 1 point. The maximum score possible in the OSpan and RSpan tasks was 75; the maximum score possible in the SymSpan task was 42.

For the RavSetA task (Raven et al., 1998), the participants had 5 min to complete 12 spatial-reasoning problems. Each problem presented a rectangular matrix of geometric figures with a missing element. The participants selected from an array of choices at the bottom of the screen the figure that would complete the overall pattern of the matrix.

The participants returned to the lab to perform running memory span tasks and an additional general fluid intelligence test—the Shipley abstraction series (Zachary, 1986)—in a second 60-min session. The participants had 5 min to complete 20 incomplete alphanumeric series presented individually on the screen, one after the other. The participants were required to type in the letter(s), number(s), or word that would complete the series. One point was assigned for each correct response in the RavSetA and Shipley tasks.

In the running memory span tasks, the participants were instructed to report the last n letters (target length = 3, 4, 5, 6, 7, or 8 letters) from lists constructed from the same set of letters as in the OSpan and RSpan tasks. The letters were presented sequentially in the center of the screen in black 18-pt font against a gray background. There were six trials for each target length in each running memory span task. Three trials for each target length presented lists in which $m = 0$ distractors preceded the targets (whole recall trials). One trial for each target length presented lists in which $m = 1, 2,$ or 3 distractors preceded the targets (partial recall trials). The target lengths were blocked (randomly ordered), and the number of distractors preceding the targets was randomized within blocks. The participants were informed at the start of a block how many letters to report from each list in that block.

The participants experienced both input-rate conditions (one letter every 500 msec vs. one every 2,000 msec) successively, with rate order counterbalanced across participants. Stimulus duration was 300 msec; the interstimulus interval was 200 msec and 1,700 msec for the 500- and 2,000-msec input-rate conditions, respectively. After the inputs terminated for a trial, the participants made their responses by clicking the cells of a 3 × 4 grid displaying all letters from the set of possible letters. The response screen reminded the participants how many letters to report. As in the complex-span tasks, strict serial position scoring was applied to the response sequences. One point was assigned for each item correctly chosen in correct serial position, with respect to the set of the last n targets, not the whole $m + n$ input sequence. For example, if the last four items were to be reported from an input sequence that was five items long (e.g., J L K T R), a response of "L K T R" would receive 4 points, but a response of "J L K T" would receive 0 points. The maximum possible score for partial recall trials, where targets were preceded by distractors, was 99.

Results

The data were screened for outliers and influential data points by inspecting scatterplots and histograms and by computing Mahalanobis distance and Cook's D for each participant (Stevens, 2002). One individual was excluded as a result. Working memory updating is assumed to be unnecessary in whole recall trials, where targets are not preceded by distractors (Morris & Jones, 1990). Therefore, the data from the whole recall trials were ignored in the following analyses. The following analyses included data from the running memory span partial recall trials, where targets were preceded by distractors.

The running memory span scores were higher when the presentation rate was slow than when it was fast [$F(1,87) = 53.32, p = .001, \eta_p^2 = .38$]. However, correlations between running memory span and criterion measures were not different across input-rate conditions (Table 1). The correlation between running memory span conditions did not approach unity but did approach the limits of the reliability estimates and was as high as the intercorrelations among the complex-span working memory measures. The correlation between running memory span conditions ($r = .776$) and a 95% confidence interval constructed around it (.734, .818) overlapped with the average intercorrelation among complex-span measures of working memory capacity ($r = .778$) and a 95% confidence interval constructed around it (.736, .820). This result suggests that working memory capacity was measured consistently by running memory span across presentation rate conditions. No differences between correlated correlations (i.e., between performances of either running memory span task and a given criterion task) were statistically significant (Meng, Rosenthal, & Rubin, 1992). Notably, performance of running memory span in either condition was as strongly related to general fluid intelligence as was performance of complex-span working memory measures. A single principal component with eigenvalue > 1 was extracted, accounting for 63.88% of total variance.

Discussion

We did not find evidence for separate working memory strategies by manipulating presentation rate in running memory span. The recall scores were indeed higher when the inputs were presented at a slow rate than at a fast rate, but, more importantly, the relations to criterion measures of working memory capacity and general fluid intelligence were not changed. Since the measurement of work-

Table 1
Descriptive Statistics and Correlations for Experiment 1

Measure	M	SD	Correlation						
			1	2	3	4	5	6	7
1. OSpan	53.63	16.95	(.848)						
2. RSpan	48.36	17.76	.819	(.869)					
3. SymSpan	23.55	9.48	.717	.798	(.843)				
4. RavSetA	7.83	2.29	.386	.468	.459	(.665)			
5. Shipley	14.06	3.09	.415	.464	.567	.618	(.849)		
6. 2,000	43.49	19.13	.545	.608	.560	.449	.615	(.864)	
7. 500	33.91	17.24	.492	.579	.612	.523	.615	.776	(.817)

Note—Coefficient alpha reliability estimates are on the diagonal. All correlations were statistically significant ($p < .01$). $N = 88$. OSpan, operation-span task; RSpan, reading-span task; SymSpan, symmetry-span task; RavSetA, Raven's standard progressive matrices; Shipley, Shipley abstraction series; 2,000, running memory span task with a presentation rate of one input item every 2,000 msec; 500, running memory span task with a presentation rate of one input item every 500 msec.

ing memory capacity and the prediction of general fluid intelligence did not differ much across the fast and slow conditions, it can be inferred that the participants took essentially the same approach to performing the task. This suggests that a slow presentation rate was not sufficient to induce the participants to rehearse to-be-remembered items in running memory span (cf. Bunting et al., 2006). It appears that the participants across the running span conditions were discouraged from taking an active approach, probably because of uncertainty during inputs regarding the number of items that would ultimately be shown and thus the identity of items that would be defined as targets at test (Cowan et al., 2005; Elosúa & Ruiz, 2008; Palladino & Jarrold, 2008).

On a more positive note, the present results demonstrate that running memory span yields consistent measurement of working memory capacity across widely differing conditions. Notably, most of the variance predicted in general fluid intelligence was shared across types of working memory measure, complex spans, and running memory spans, suggesting that a common working memory capacity was measured. This is consistent with previous reports (Cowan et al., 2005; Tucker-Drob & Salthouse, 2009).

EXPERIMENT 2

We were not able to distinguish individual differences in working memory strategies by manipulating the time between inputs in a running memory span task. However, against the results of our Experiment 1 there stands a larger body of research supporting the idea that participants spontaneously rehearse and group to-be-remembered items unless strong experimental controls are imposed (Watkins, 1989). Therefore, we wished to give active input processing another chance to be distinguished. We took a different approach in Experiment 2.

First, we gave the participants even more time between inputs. Presentation rate was constant across conditions and relatively slow (one item every 2,500 msec). Second, we examined the effects of cuing (precue or postcue). We assumed that informing the participants how many items to report before inputs (precue) would support active input processing. We assumed that withholding that information until test (postcue) would discourage active input processing. Note that as in the postcue (as well as precue) condition, the participants were left free to rehearse and group an arbitrary set of items of their own choosing, with or without knowledge of the experimenter-defined target set (Henson, 1998), but we assumed that this strategy would be more effective in the precue condition than in the postcue condition, precisely because of the reported experimenter-defined target set.

In Experiment 1, the number of $m + n$ inputs was not strictly independent of the number of n targets, because the number of inputs always exceeded the number of targets according to an incremental scheme (i.e., $m = 0, 1, 2,$ or 3 distractors, always preceding the targets, although m was randomly determined on each trial). Learning this relation between the number of $m + n$ inputs and number of n tar-

gets could provide the participants with information that could be used to prepare responses even in a postcued version of running memory span. For example, after viewing seven inputs in a particular trial, a participant could safely infer that recall of three items would not be required at test, because no target set was ever preceded by more than three distractors. Thus, a participant could infer that recall of more than three items would be required on such trials. In addition, this relation between the number of inputs and the number of targets entails that on a large proportion of trials, the participants might be able to memorize the entire input sequence and later selectively report the defined target set. In Experiment 2, we controlled these alternative approaches to performing running memory span by including an input-length condition (between subjects) in which the number of $m + n$ inputs was independent of the number of n targets, and the input sequences uniformly exceeded typical span limits.

Predictions

Again, we expected to dissociate active and passive input-processing approaches to running memory span and to observe different relations with reference abilities of working memory capacity and general fluid intelligence. We expected that responses across the running span conditions would not identify the same individual differences in reference abilities. Alternatively, we expected to find boundaries on the range of conditions under which running memory span yields valid working memory capacity measurement.

Method

Participants. A total of 294 individuals participated in this experiment (153 women and 141 men). The participants were between 18 and 35 years old, inclusive ($M = 23.9$ years). None of these participants had participated in the previous experiment.

Procedure. The participants performed running memory span tasks in a 90-min session after first performing two complex-span tasks and three general fluid intelligence tasks. The complex-span tasks were the same OSpan and SymSpan tasks used in Experiment 1. We changed the general fluid intelligence tasks from those in Experiment 1 in order to generalize to other measures and because one had shown notably weak reliability in Experiment 1 (the RavSetA task).

The participants had 10 min to complete the 18 odd-numbered problems from Raven's Advanced Progressive Matrices (Raven et al., 1998). One point was assigned for each correct response within the time limit. The *letter-sets* task (Ekstrom, French, Harman, & Derman, 1976) required the participants to identify a group of letters that violated a pattern defined by four other letter groups. The participants had 5 min to complete 20 problems, and 1 point was assigned for each correct response. The *number-series* task (Thurstone, 1938) required the participants to choose a number from a set of alternatives that would complete a pattern defined by a numeric series displayed above. The participants had 5 min to complete 15 problems, and 1 point was assigned for each correct response.

In all running memory span tasks in Experiment 2, the participants were instructed to report the last n (target length = 3, 4, 5, or 6) letters from lists constructed from the same set of possible letters as were those in Experiment 1. The stimuli were presented visually, as in Experiment 1. There were three trials for each target length. The stimuli were presented for 300 msec, with an interstimulus interval of 2,200 msec in all conditions. Thus, the input rate in all running

Table 2
Descriptive Statistics and Correlations for Experiment 2,
Dependent Input-Length Condition

Measure	M	SD	Correlation						
			1	2	3	4	5	6	7
1. OSpan	56.48	14.92	(.834)						
2. SymSpan	26.22	8.57	.588	(.773)					
3. RAPM	8.45	3.68	.501	.563	(.792)				
4. LetSet	9.75	3.21	.373	.415	.536	(.728)			
5. NumSer	8.44	3.26	.481	.568	.671	.610	(.794)		
6. Precue	22.16	8.65	.566	.518	.585	.502	.615	(.758)	
7. Postcue	21.62	9.17	.591	.462	.521	.408	.524	.785	(.807)

Note—Coefficient alpha reliability estimates are on the diagonal. All correlations were statistically significant ($p < .01$). $N = 151$. OSpan, operation-span task; SymSpan, symmetry-span task; RAPM, Raven's advanced progressive matrices; LetSet, letter-set task; NumSer, number-series task; Precue, running memory span in the precue condition; Postcue, running memory span in the postcue condition.

Table 3
Descriptive Statistics and Correlations for Experiment 2,
Independent Input-Length Condition

Measure	M	SD	Correlation						
			1	2	3	4	5	6	7
1. Ospan	56.32	15.20	(.845)						
2. SymSpan	26.36	8.49	.573	(.745)					
3. RAPM	8.53	3.45	.424	.617	(.760)				
4. LetSet	9.71	3.46	.428	.536	.433	(.768)			
5. NumSer	8.69	2.95	.507	.639	.606	.616	(.746)		
6. Precue	32.80	11.16	.569	.560	.386	.496	.452	(.769)	
7. Postcue	31.55	12.06	.563	.563	.528	.400	.497	.684	(.810)

Note—Coefficient alpha reliability estimates are on the diagonal. All correlations were statistically significant ($p < .01$). $N = 143$. OSpan, operation-span task; SymSpan, symmetry-span task; RAPM, Raven's advanced progressive matrices; LetSet, letter-set task; NumSer, number-series task; Precue, running memory span in the precue condition; Postcue, running memory span in the postcue condition.

span tasks in Experiment 2 was one item every 2,500 msec, slower than the slow rate in Experiment 1.

The participants were instructed how many targets to report in advance of the input sequence (precue), as in Experiment 1, or else they had to wait until the recall screen to learn this information (postcue). In the precue condition, target lengths were blocked (randomly ordered), and the number of distractors preceding the targets was randomized within blocks, as in Experiment 1. The participants were informed at the start of a block how many letters to report from each list in that block, as in Experiment 1. The response screen reminded the participants how many letters to report. In the postcue condition, target lengths, as well as the number of distractors preceding the targets, were fully randomized. The response screen gave the participants their only source of information regarding how many letters to report. Cuing conditions were experienced successively by all of the participants, and the order of cuing was counterbalanced across participants.

A between-subjects manipulation of input length varied the dependence of input length on target sequence lengths. This also provided a condition in which the total number of inputs always exceeded typical span limits, to prevent the participants from memorizing input sequences entirely and later selectively reporting targets. Approximately half of the participants ($n = 151$) were assigned to the dependent input-length condition and received sequences in which the n targets were preceded by $m = 0$, 1, or 2 distractors, similar to the running span tasks in Experiment 1. We refer to this input-length condition as *dependent* because the number of $m + n$ inputs depended on the number of n targets, as in the running span tasks in Experiment 1. For each target length, one trial presented lists in which $m = 0$, 1, or 2 distractors preceded targets, determined randomly. Approximately half of the participants ($n = 143$) were

assigned to the independent input-length condition and received sequences that were 9, 10, or 11 items in length, regardless of the number of targets. We refer to this input-length condition as *independent* because the number of $m + n$ inputs did not depend on the number of n targets, unlike the other input-length condition (and the running span tasks in Experiment 1). For each target length, one trial presented lists in which the total number of inputs was 9, 10, or 11, determined randomly.

The participants were assigned to the input-length conditions in alternation in the order in which they visited the lab. The response format and scoring were the same as those in Experiment 1. Only data from the partial recall trials were included. This affected the precue condition but not the postcue condition; therefore, the maximum possible scores are different. The maximum score possible in the dependent input-length condition was 36. The maximum score possible in the independent input-length condition (i.e., all trials) was 54.

Results

No outliers or influential points were identified while screening the data. There were no effects of cuing on performance, and the relations to criterion measures of working memory capacity and general fluid intelligence were much the same across conditions (Tables 2 and 3). Percentile scores for the running memory span distributions are reported in Table 4. Within the dependent condition, the total numbers of correct responses were not statistically different as a function of cuing [$F(1,150) = 1.265$, $p = .262$, $\eta_p^2 = .008$]. Within the independent condition, the total numbers of correct responses were not statistically

Table 4
Percentile Scores for Running Memory Span Tasks
in Experiment 2

Percentiles	Dependent Input-Length Condition		Independent Input-Length Condition	
	Precue	Postcue	Precue	Postcue
10	10	8	15	12
20	14	13	25	21
30	18	16	28	28
40	20	19	31	31
50	23	23	34	34
60	26	25	37	36
70	28	28	40	38
80	30	30	41	41
90	33	33	47	47

Note—Precue, running memory span in the precue condition; Postcue, running memory span in the postcue condition.

different as a function of cuing [$F(1,142) = 2.635$, $p = .107$, $\eta_p^2 = .018$].

The strength of the correlation between the pre- and postcue responses did not differ across input-length conditions. A 95% confidence interval constructed around the difference between the uncorrelated correlations (Zou, 2007) between the precue and postcue responses in dependent versus independent conditions included zero ($-.006$, .208).

The correlation between performances across running memory span conditions was stronger than the correlations among criterion working memory capacity measures. Within the dependent input-length condition, comparing the correlation between performance of running memory span across cuing conditions ($r = .785$) and the 95% confidence interval constructed around it (.75, .82) with the correlation between criterion working memory capacity measures ($r = .58$) and the 95% confidence interval constructed around it (.52, .63) indicates that responses across cuing conditions were more strongly related to each other than responses across the criterion working memory capacity measures. This suggests that working memory capacity was measured consistently by running memory span across cuing conditions within the dependent input-length condition.

Within the independent input-length condition, comparing the correlation between performance of running memory span across cuing conditions ($r = .684$) and the 95% confidence interval constructed around it (.64, .73) with the correlation between criterion working memory capacity measures ($r = .573$) and the 95% confidence interval constructed around it (.52, .63) indicates that responses across cuing conditions were more strongly related to each other than responses across the criterion working memory capacity measures. This suggests that working memory capacity was measured consistently by running memory span across cuing conditions within the independent input-length condition.

Within either input-length condition, the same individual differences in criterion abilities were related to the same individual differences in running memory span, whether the responses were pre- or postcuing. No differences between correlated correlations (i.e., between performances

of either running memory span task and a given criterion task) were statistically significant. Notably, performance of running memory span in either cuing condition was as strongly related to general fluid intelligence as was performance of complex-span measures. In the dependent input-length condition, a single principal component with eigenvalue > 1 was extracted, accounting for 60.95% of total variance. In the independent input-length condition, a single principal component with eigenvalue > 1 was extracted, accounting for 59.62% of total variance.

Discussion

We manipulated the availability of information that participants could use for active input processing. The results again failed to provide evidence for active response preparation during inputs, although the participants experienced relatively long empty intervals between inputs in both the precue condition and the postcue condition and advance information in the precue conditions. Performance levels were about the same whether the number of targets was known in advance of the test or learned at the time of the test. Furthermore, and more important, the rank orderings of individuals by criterion measures were unchanged. Responses were strongly intercorrelated across running memory span conditions and were correlated with reference ability measures to the same extent. Since measurement of working memory capacity and prediction of general fluid intelligence were not different across the pre- and postcue conditions, it can be inferred that the participants took essentially the same approach to performing the task.

Which approach to running memory span—active or passive—appears more likely to account for the similar performance levels and equally strong relations with criterion measures across cuing conditions? It is not plausible that preparing responses in advance of test could have been equally effective across cuing conditions. It is more plausible that waiting until the inputs terminated could have been equally effective across cuing conditions. Because presentation rate was relatively slow across cuing conditions, the results indicate (as in Experiment 1) that a slow presentation rate was not sufficient to induce the participants to rehearse and group to-be-remembered items in running memory span. Again, it appears that the participants across conditions were discouraged from taking an active approach, probably because of uncertainty during the inputs regarding the number of items that would be shown and, thus, the identity of items that would be defined as targets at test.

On a more positive note, the present results demonstrate again that running memory span yields consistent measurement of working memory capacity across widely differing conditions. Indeed, the performance levels and correlations with reference abilities were so similar across the pre- and postcue versions that these might serve as parallel test forms. Most of the variance predicted in general fluid intelligence was shared across the complex-span and running memory span tasks, suggesting that a common working memory capacity was measured, as in Experiment 1 and consistent with previous reports (Cowan et al., 2005; Tucker-Drob & Salthouse, 2009).

GENERAL DISCUSSION

Active input processing in running memory span means that participants prepare responses in advance of test, perhaps by rehearsing and grouping a changing set of targets. This approach corresponds to a common account of running memory span as involving working memory updating (Morris & Jones, 1990). We had reasoned that if the participants were given relatively long empty intervals between inputs, they would be induced to take an active approach, perhaps by rehearsing and grouping (Bunting et al., 2006). *Passive input processing* means that participants wait until after inputs have terminated and prepare responses at the time of test. We had reasoned that if the participants were given very little time between inputs, they would be forced to rely on a passive approach.

The results indicate that a fast presentation rate is not necessary to discourage participants from active input processing in running memory span. Also, a slow presentation rate is not sufficient to induce them to active input processing. The scores were indeed higher when the inputs were presented more slowly than when they were presented more quickly. But the scores were nearly identical whether the participants knew how many targets to report in advance or learned that information at the time of test. More importantly, however, the rank ordering of individuals in terms of working memory capacity and general fluid intelligence did not change in response to any manipulation at all. If the participants had taken qualitatively different approaches across conditions, it should have been evident in differing relations to criterion variables, as well as in performance levels.

These results were surprising. It is widely accepted that participants in serial recall tasks will rehearse if provided with relatively long, empty intervals between items (Oberauer & Lewandowsky, 2008; Tan & Ward, 2008; Watkins, 1989), and good evidence for rehearsal in a slow-paced running memory span task has been reported previously (Bunting et al., 2006). However, the present data suggest that even with relatively long, empty intervals, the participants did not rehearse or group inputs in a way that differentially affected working memory capacity measurement. Whether rehearsal was strongly controlled with a fast presentation rate or weakly controlled with a slow presentation rate, the relations with criterion measures were not changed.

Overall, the results suggest that participants in running memory span are generally discouraged from taking an active approach, probably because of uncertainty regarding the number of items that will be shown and thus the identity of items that will be defined as targets at test (Cowan et al., 2005; Elosúa & Ruiz, 2008; Palladino & Jarrold, 2008). This has implications for the common interpretation of running memory span as a measure of working memory updating (Morris & Jones, 1990; Postle, 2003). Unless explicit instructions to rehearse are given and enforced (as in Friedman et al., 2006), it should not be assumed that participants engage in working memory updating in running memory span. In general, participants are more likely to wait until inputs have terminated and the identity of target items has been fixed before they expend great effort to prepare responses.

On a more positive note, the present results demonstrate that running memory span yields consistent measurement of working memory capacity across widely differing conditions. The present results, obtained from within the same samples in single experiments, are consistent with previous research, where running span was administered under widely differing conditions, but across different samples and separate studies, each reporting strong relations with criterion measures of working memory and fluid intelligence (Cowan et al., 2005; Friedman et al., 2006; Tucker-Drob & Salthouse, 2009).

If the uniformly strong correlations with criterion measures of working memory capacity and general fluid intelligence observed here cannot be ascribed to working memory updating, to what psychological process might they be ascribed? The focus of attention proposed by Cowan (2001; Cowan et al., 2005) must be considered a strong candidate. However, this mechanism is usually described as operating on rapidly decaying sensory memory traces. We feel that unless this assumption is relaxed, the long interitem intervals in the present experiments argue against focus of attention as the sole contributor to relations with criterion measures. We suggest that the control of attention, as well as its scope (Cowan et al., 2005; Engle & Kane, 2004), and effortful retrieval (Unsworth & Engle, 2007) would be of major importance for performing running memory span across any condition, to combat interference from distractors presented before targets. In this defining aspect (distractors before targets), running memory span tasks are similar to complex-span tasks, which might account in part for their common relations with higher order cognition.

Summary and Conclusions

We identified the same individual differences in working memory capacity across a wide range of running memory span conditions. In some conditions, corresponding in general to standard administration as a measure of working memory updating (Morris & Jones, 1990), active input processing was assumed to be feasible. In other conditions, it was much less feasible. The relations with criterion measures of working memory capacity and general fluid intelligence were much the same in either case.

Overall, the results indicate that running memory span is robust as a measure of working memory capacity and as a predictor of general fluid intelligence but may not be as sensitive to details of administration as has previously been supposed. Finally, the results add to growing evidence that running memory span tasks function similarly to complex-span tasks as measures of working memory capacity that are strongly predictive of general fluid intelligence.

ing this article should be sent to J. M. Broadway, School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332 (e-mail: jbroadway@ gatech.edu).

REFERENCES

BADDELEY, A. D., & HITCH, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). New York: Academic Press.

BUNTING, M., COWAN, N., & SAULTS, J. S. (2006). How does running memory span work? *Quarterly Journal of Experimental Psychology*, **59**, 1691-1700.

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.

COWAN, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences*, **24**, 87-114.

COWAN, N., ELLIOTT, E. M., SAULTS, J. S., MOREY, C. C., MATTOX, S., HISMJATULLINA, A., & CONWAY, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, **51**, 42-100.

DEARY, I. J. (2000). *Looking down on intelligence.* Oxford: Oxford University Press.

EKSTROM, R. B., FRENCH, J. W., HARMAN, H. H., & DERMAN, D. (1976). *Kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

ELOSÚA, M. R., & RUIZ, R. M. (2008). Absence of hardly pursued updating in a running memory task. *Psychological Research*, **72**, 451-460.

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* (Vol. 44, pp. 145-199). New York: Elsevier.

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.

FRIEDMAN, N. P., MIYAKE, A., CORLEY, R. P., YOUNG, S. E., DEFRIES, J. C., & HEWITT, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, **17**, 172-179.

HENSON, R. N. A. (1998). Short-term memory for serial order: The start–end model. *Cognitive Psychology*, **36**, 73-137.

HOCKEY, R. (1973). Rate of presentation in running memory and direct manipulation of input-processing strategies. *Quarterly Journal of Experimental Psychology*, **25**, 104-111.

KANE, M. J., & ENGLE, R. W. (2002). The role of prefrontal cortex in working memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, **9**, 637-671.

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.

LÉPINE, R., BARROUILLET, P., & CAMOS, V. (2005). What makes working memory spans so predictive of high-level cognition? *Psychonomic Bulletin & Review*, **12**, 165-170.

MENG, X. L., ROSENTHAL, R., & RUBIN, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, **111**, 172-175.

MIYAKE, A. FRIEDMAN, N. P., EMERSON, M. J., WITZKI, A. H., HOWERTER, A., & WAGER, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, **41**, 49-100.

MORRIS, N., & JONES, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, **81**, 111-121.

MUKUNDA, K. V., & HALL, V. C. (1992). Does performance on memory for order correlate with performance on standardized measures of ability? A meta-analysis. *Intelligence*, **16**, 81-97.

OBERAUER, K., & LEWANDOWSKY, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review*, **115**, 544-576.

OBERAUER, K., SÜß, H.-M., WILHELM, O., & WITTMAN, W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, **31**, 167-193.

PALLADINO, P., & JARROLD, C. (2008). Do updating tasks involve updating? Evidence from comparisons with immediate serial recall. *Quarterly Journal of Experimental Psychology*, **61**, 392-399.

POLLACK, I., JOHNSON, L. B., & KNAFF, P. R. (1959). Running memory span. *Journal of Experimental Psychology*, **57**, 137-146.

POSTLE, B. R. (2003). Context in verbal short-term memory. *Memory & Cognition*, **31**, 1198-1207.

RAVEN, J. E., RAVEN, J. C., & COURT, J. H. (1998). *Progressive matrices.* Oxford: Oxford Psychologists Press.

RUIZ, M., ELOSÚA, M. R., & LECHUGA, M. T. (2005). Old-fashioned responses in an updating memory task. *Quarterly Journal of Experimental Psychology*, **58A**, 887-908.

SCHNEIDER, W., ESCHMAN, A., & ZUCCOLOTTO, A. (2002). E-Prime user's guide. Pittsburgh: Psychology Software Tools.

STERNBERG, R. J. (1979). The nature of mental abilities. *American Psychologist*, **34**, 214-230.

STEVENS, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Erlbaum.

TAN, L., & WARD, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review*, **15**, 535-542.

THURSTONE, T. G. (1938). *Primary mental abilities.* Chicago: University of Chicago Press.

TUCKER-DROB, E. M., & SALTHOUSE, T. A. (2009). Confirmatory factor analysis and multidimensional scaling for construct validation of cognitive abilities. *International Journal of Behavioral Development*, **33**, 277-285.

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., HEITZ, R. P., SCHROCK, J. C., & ENGLE, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, **37**, 498-505.

UNSWORTH, N., REDICK, T. S., HEITZ, R. P., BROADWAY, J. M., & ENGLE, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, **17**, 635-654.

WATKINS, M. J. (1989). Willful and nonwillful determinants of memory. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 59-71). Hillsdale, NJ: Erlbaum.

ZACHARY, R. A. (1986). *Shipley Institute of Living Scale: Revised manual.* Los Angeles: Western Psychological Services.

ZOU, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods*, **12**, 399-413.