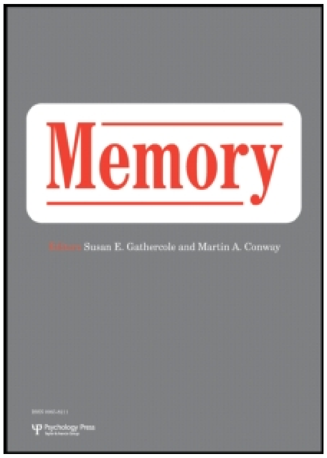

Memory
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683358

Strategy use in the reading span test: An analysis of eye movements and reported encoding strategies

Online Publication Date: 01 August 2007
To cite this Article: Kaakinen, Johanna K. and Hyönä, Jukka (2007) 'Strategy use in the reading span test: An analysis of eye movements and reported encoding strategies', Memory, 15:6, 634 - 646
To link to this article: DOI: 10.1080/09658210701457096
URL: http://dx.doi.org/10.1080/09658210701457096

MEMORY, 2007, 15 (6), 634–646

Psychology Press
Taylor & Francis Group

Strategy use in the reading span test: An analysis of eye movements and reported encoding strategies

Johanna K. Kaakinen and Jukka Hyönä
University of Turku, Finland

Strategy use in the traditional reading span test was examined by recording participants' eye movements during the task (Experiment 1) and by interviewing participants about their strategy use (Experiment 2). In Experiment 1, no differences between individuals with a low, medium, and high span were observed in how they distributed processing time between task elements. In all three groups, fixation times on words up to the to-be-remembered (TBR) word became shorter and the time spent on the TBR longer as memory load in the task increased. The results of Experiment 2, however, show that span groups differ in the use of memory encoding strategies: individuals with a low span use mainly rehearsal, whereas individuals with a high span use almost exclusively semantic elaboration. The results indicate that the use of elaborative strategies may enhance span performance but that not all individuals are necessarily able to use such strategies efficiently.

Complex working memory span tests such as the reading span (Daneman & Carpenter, 1980) and operation span (Turner & Engle, 1989) have proven to be good predictors of higher-order cognitive abilities, such as reading comprehension (e.g., Daneman & Merikle, 1996). A critical feature of the complex span tasks is that they combine a processing and a storage component. For instance, in the original version of the reading span test (Daneman & Carpenter, 1980), participants read aloud sets of two to six sentences while they try to temporarily maintain in memory the last words of the sentences. Sentences are presented one at a time so that immediately the reader has finished reading a sentence a new sentence is presented. After reading all the sentences of a particular set the participant is asked to recall the last words. There is considerable variability between individuals in how many words can be successfully recalled: even among college students, the number of successfully recalled words varies between two to five words (see e.g., Friedman & Miyake, 2004). Research shows that, for example, individuals with a high span are more effective than individuals with a low span in learning relevant information from expository text (e.g., Kaakinen, Hyönä, & Keenan, 2003).

But what is the common factor between the reading span task performance and reading comprehension? In other words, what does the reading span task actually measure? This question still remains unanswered (see e.g., Conway, Jarrold, Kane, Miyake, & Towse, 2007; Cowan, 2005; Miyake, 2001; Miyake & Shah, 1999). Depending on one's theoretical view on working memory, the reading span task (and other complex span tasks) are assumed to tap into the common pool of mental capacity that is used for processing and storage (e.g., Just & Carpenter, 1992), ability to

Address correspondence to: Johanna K. Kaakinen, Department of Psychology, University of Turku, FIN-20014 Turku, Finland. E-mail: johanna.kaakinen@utu.fi

Parts of the data were presented at the 13th Conference of the European Society for Cognitive Psychology in Granada, Spain, in September 2003; at the 46th Annual Meeting of the Psychonomic Society in Toronto, Canada, in November 2005; and at the 3rd European Working Memory Symposium in Genoa, Italy, in June 2006.

We would like to thank Miia Juntunen for her help in preparing the Finnish materials for the span task.

http://www.psypress.com/memory DOI:10.1080/09658210701457096

control attention in a task-relevant manner (e.g., Kane, Bleckley, Conway, & Engle, 2001), the costs related to the continuous switching between the processing and storage tasks incorporated in the span task (e.g., Towse, Hitch, & Hutton, 2000), the ability to inhibit irrelevant information from entering working memory (e.g., May, Hasher, & Kane, 1999), and/or the use of efficient memory encoding strategies (McNamara & Scott, 2001).

The present study examined the view that individual differences in the reading span task may derive from differences in the strategies utilised during the task (McNamara & Scott, 2001). Because the reading span task combines a processing component (reading the sentence) and a storage component (maintaining the last word), it is possible that individuals differ either in the online processing strategies utilised during reading the test sentences, strategies used for storing or maintaining the to-be-remembered (TBR) information, or some combination of these two. For instance, going through a sentence up to the TBR word as quickly as possible in order to spend more time on the TBR word would be a beneficial online processing strategy because it maximises the visual attention on the TBR information and should thus increase the likelihood of its later recall. As for the efficient memory strategies, use of mnemonic strategies such as imagery or semantic elaboration of the TBR words may enhance the memory for the TBR words. Naturally, it is also possible that if people engage in strategic online processing of the test materials, they have extra time available to make use of the mnemonic strategies.

Previous results on the strategic online processing of the span test materials are somewhat inconsistent, possibly because of methodological differences between the different studies (Carpenter & Just, 1989; Engle, Cantor, & Carullo, 1992; Friedman & Miyake, 2004). Carpenter and Just (1989) registered eye movements of six individuals with low span and six individuals with a high span during the original version of the task and found that participants who scored high in the span task spent less time reading the words within the sentence up to the TBR word and more time gazing at the TBR word than participants who scored low in the task. The difference was observed only under memory load, i.e., when reading sentences while maintaining one or more sentence-final words in memory. In other words, individuals who were successful in the span task adopted a selective processing strategy when the task became more taxing and allocated less time to task-irrelevant information and more time to TBR information. Friedman and Miyake (2004) recorded sentence reading times during the original version of the span task and reported that sentence reading time was negatively correlated with the span score, indicating that individuals with a high span were faster in reading the test sentences. Moreover, individuals with a low span slowed down their reading more as a function of increasing memory load than individuals with a high span. However, because Friedman and Miyake (2004) only report total reading times of the test sentences, the time used to view the TBR word is included in the measure. It is thus impossible to say whether individuals with a high span were using a selective processing strategy (cf. Carpenter & Just, 1989) or if they simply were faster in reading the whole sentence. In another study, Engle et al. (1992) used a self-paced moving-window technique and a modified reading span task (Turner & Engle, 1989), and found that under memory load individuals with a high span spent longer viewing the first word of the sentence as well as the TBR word than the individuals with a low span. They did not find group differences in the reading time of the words in the middle of the sentences.

As for the memory encoding strategies utilised during the reading span task, Friedman and Miyake (2004) analysed strategy use in the original version of the reading span test and found that the use of mental imagery was beneficial for the test performance. Using a slightly modified version of the original reading span task (Turner & Engle, 1989), McNamara and Scott (2001) found that individuals with a high span use elaborative encoding strategies, such as chaining and imagery, whereas individuals with a low span use simple rehearsal. Turley-Ames and Whitfield (2003) and Dunlosky and Kane (in press) report strategies used in another complex span task, the operation span test (Turner & Engle, 1989). Both studies suggest that higher span scores are related to the use of elaborative encoding strategies.

In sum, it is clear from the previous studies that an elaborative encoding strategy is beneficial in the span task. On the other hand, the role of strategic online processing of the span task materials is still unclear. Because the protocol of administering the reading span task varied across the studies it is difficult to compare the results: for instance, in some studies an experimenter-paced version of the test was used (Carpenter & Just, 1989; Friedman & Miyake, 2004), whereas in another study participants were allowed to pace the testing themselves (Engle et al., 1992). Moreover, the exact version of the reading span task varied in the different studies. In some studies the original version (Daneman & Carpenter, 1980) was used (Carpenter & Just, 1989; Friedman & Miyake, 2004), whereas other studies used a modified version (Turner & Engle, 1989), in which the TBR words are presented separately from the sentence that is being read aloud (Engle et al., 1992; McNamara & Scott, 2001). The administration method has a significant impact on the span task performance because it may limit the strategy use (Friedman & Miyake, 2004) and thus the results from different studies are not comparable. An exception is a study by Friedman and Miyake (2004), who examined the online processing and memory encoding strategies in the same task. However, they report only overall sentence reading times, which do not allow the analysis of the allocation of processing time between the different parts of the test sentences (beginning of the sentence versus the TBR word). What is needed is a study in which online processing and memory encoding strategies are examined in the same task using a time-sensitive measurement method that reveals the processing time allocated to the different regions of the test sentences. In the present study, eye tracking was used to obtain a detailed view of the time-course of processing of the reading span test materials (Experiment 1), while the memory encoding strategies in the same task were examined with interviews (Experiment 2).

The present study examined online processing strategies (Experiment 1) and memory encoding strategies (Experiment 2) during the original version of the reading span task (Daneman & Carpenter, 1980). In accordance with the original task and some recent recommendations (Conway et al., 2005; Friedman & Miyake, 2004), we implemented an experimenter-paced procedure in both experiments. In Experiment 1, eye tracking was utilised to obtain a detailed view about the time-course of processing of the test materials. Eye tracking is a highly time-sensitive method for examining online reading processes as they occur in natural reading (without the need for, e.g., word-by-word presentation). Eye fixation times have been shown to be sensitive measures of processing difficulty also during reading aloud (Hyönä & Olson, 1995) and thus it was possible to examine the reading span task performance in standard testing conditions. If strategic online processing is related to good span task performance, then we should observe differences between individuals with low and high spans in how they distribute processing time between different regions of the test sentences. More specifically, if a selective processing strategy is beneficial for the span task performance (Carpenter & Just, 1989), then individuals with a high span should show shorter eye fixation times in the task-irrelevant regions of the test materials (beginning and middle parts of the test sentences) and longer viewing times on the TBR word than individuals who score low in the span test. The group differences should be especially pronounced in high memory load conditions (Carpenter & Just, 1989; Engle et al., 1992; Friedman & Miyake, 2004).

In Experiment 2, we collected general reports (see Dunlosky & Kane, in press) about the memory encoding strategies utilised during the task. If individual differences in the reading span task are related to the use of memory encoding strategies, then individuals with a high span should report more strategy use and/or more use of efficient strategies than individuals with a low span. Efficient strategies include either semantic elaboration or visual imagery of the TBR words, whereas simple phonological rehearsal of the TBR words is regarded as a less efficient strategy (Dunlosky & Kane, in press; Friedman & Miyake, 2004; McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003).

EXPERIMENT 1

In Experiment 1 we examined participants' eye movement patterns during the reading span task. Participants performed an experimenter-paced version (see Conway et al., 2005; Friedman & Miyake, 2004) of the original reading span task (Daneman & Carpenter, 1980) while their eye movements were recorded. Eye fixation measures (gaze duration and total fixation time) were computed separately for words in the different regions of the test sentences: the first word of the sentence, the words in the middle of the sentence, and the TBR word (Engle et al., 1992).

If strategic online processing is related to good span task performance, then individuals with low and high spans should differ in how they distribute processing time between different regions of test sentences. More specifically, if a selective processing strategy is beneficial for the span task performance (Carpenter & Just, 1989), then individuals with a high span should show shorter eye fixation times in the task-irrelevant regions of the test sentences (beginning and middle region) and longer viewing times on the TBR word than individuals with a low span. The group differences should be especially pronounced under increased levels of memory load (Carpenter & Just, 1989; Engle et al., 1992; Friedman & Miyake, 2004). When the participant is reading the first test sentence there is no memory load involved, in the sense that there are no final words to be remembered. During the reading of the second sentence in the trial, the reader has to maintain the last word of the first sentence in memory—thus, there is a memory load of one word. During the reading of the third sentence of the trial, the reader has to maintain two words in memory, and so on. In order to examine the allocation of processing time between the different parts of the test sentences under different memory loads, the eye fixation times were analysed as a function of sentence region and memory load. Because we used the experimenter-paced version of the reading span test (see e.g., Conway et al., 2005; Friedman & Miyake, 2004), which minimises the time available to devise and implement memory encoding strategies, this experiment is a stringent test of the importance of strategic online processing in span task performance.

Method

Participants. A total of 56 students of the University of Turku participated in the study to fulfil a course requirement. The mean age of the participants was 24.02 years ($SD = 4.14$).

Apparatus. Eye movements were collected by the EyeLink eye tracker manufactured by SR Research Ltd. (Mississauga, Ontario, Canada). The eye tracker is an infrared video-based tracking system combined with hyperacuity image processing. There are two cameras mounted on a headband (one for each eye), including two infrared LEDs for illuminating each eye. The headband weighs 450 g in total. The cameras sample pupil location and pupil size at the rate of 250 Hz. Registration can be done either monocularly or binocularly. We performed it for the selected eye (usually the right eye) by placing the camera and the two infrared lights 4–6 cm away from the eye. The resolution of eye position is 15 s of arc and the spatial accuracy better than 0.5°. Head position with respect to the computer screen is tracked with the help of a head-tracking camera mounted on the centre of the headband at the level of the forehead. Four LEDs are attached to the corners of the computer screen, which are viewed by the head-tracking camera, once the participant sits directly facing the screen. Possible head motion is detected as movements of the four LEDs and is compensated for online from the eye position records.

Reading span test. We used a Finnish version of the original reading span test (Daneman & Carpenter, 1980). The test materials consisted of 81 test sentences and 6 practice sentences, which were all five to seven words long. All final words of the sentences were nouns, which were matched for mean length and frequency (Laine & Virtanen, 1999). None of the final words was repeated in the other test sentences.

In the test, participants read aloud sets of unrelated sentences presented on a computer monitor. The participants were instructed to start reading each sentence as soon as it appeared on the screen. The experimenter controlled the sentence presentation time: immediately after the participant had finished vocalising the sentence, the experimenter pressed a button, the sentence disappeared, and a new sentence came up on the screen. After the participant had read the sentences of a particular trial, a blank screen appeared and the participant had to recall the last word of each sentence of the trial in the order in which they were presented. Participants were instructed that if they could not retain the correct order of the words, they should recall all words they could, but that they should not start the recall with the last word of the last sentence (see e.g., Friedman & Miyake, 2004).

The test started with a practice session of three two-sentence trials. In other words, in one trial participant read aloud two different sentences and then recalled the sentence-final words. When the practice session ended, participants were notified that the actual experiment would start and the testing started with another set of three two-sentence trials. There always were three trials in each set size (i.e., three trials in the set size of two sentences, three trials in the set size of three sentences, three trials in the set size of four sentences, etc.). After the three trials of the particular set size, a decision was made whether the participant would continue the test or not. If the participant successfully recalled all sentence-final words at least in one of the three trials (e.g., in set size two, if the participant correctly recalled the final words of both sentences at least in one trial), the testing was continued with the next set size. However, if the participant failed to recall the sentence-final words in all three trials, the testing was terminated. The test sentences were presented in a fixed order; the maximum number of sentences in a trial was seven.

In accordance with recent recommendations for scoring the span test performance (Conway et al., 2005; Friedman & Miyake, 2005), the span score for each participant was defined as the total number of correctly recalled words in the test (maximum score was 81). A word was scored as correctly recalled even if the participant failed to retain the correct order (Friedman & Miyake, 2004).

Procedure. Each participant was tested individually. In the beginning of the experiment, the eye tracker was set up and calibrated. Participants were instructed to read aloud the sentences presented one at a time on a computer screen. They were told that when a blank screen appeared, they should repeat all final words of the sentences in that particular set of sentences. It was stressed that they should start reading the sentence aloud immediately when it appeared on the screen. If the experimenter noticed during the test that the participant was pausing before starting to read the next sentence, the participant was reminded that there should be no delay. Participants were informed that the number of sentences would gradually increase, and they were warned whenever the set size increased during the task. The testing lasted for about 20 minutes.

Results

Words in each experimental sentence were categorised into three groups: the first word of the sentence, words in the middle of the sentence, and the sentence-final (TBR) word. Average per-word fixation time measures were computed separately for each sentence region across all three trials of the particular set size, irrespective to whether the participant correctly recalled sentence-final word or not. Two fixation time measures were computed: (1) gaze duration corresponds to the first-pass fixation time on a word, and (2) total fixation time is the summed durations of all fixations landing on a word, including first-pass fixation time and possible rereadings. Because the words varied in length across the different regions, the fixation time on each word was divided by the number of characters in the word.[1]

Participants were divided into three groups based on their performance in the reading span task: a tertile split of the total number of words correctly recalled was used to form a low-span group (scores 17–21, $N = 19$), a medium-span group (scores 22–29, $N = 17$), and a high-span group (scores 30–73, $N = 20$). These cut points coincide closely with those used in previous studies examining the relationship between working memory span and reading (e.g., Kaakinen et al., 2003). The means of total number of recalled words were 19.89 ($SD = 1.41$), 22.82 ($SD = .81$), and 43.95 ($SD = 15.55$) for low-, medium-, and high-span groups, respectively. Pairwise t tests using a correction to the degrees

[1] The present study focused on how participants divided processing time between different regions of the reading span test materials, not on the component processes related to reading, such as word recognition. Thus, some modifications were made to the standard procedures typical for psycholinguistic research. First, in the present study a skipped word was assigned a gaze duration value of 0 ms. In psycholinguistic eye-movement research it is generally preferable to analyse skip rates separately and to treat skips as missing data in fixation time analyses. However, because we were interested in how participants divide their processing time between the different regions of the reading span test sentences, a composite measure of fixation time is more informative than separate measures of skip rates and gaze durations. Due to the same reason, data from all words that were in the middle of the sentence, including function words, were included when computing the average fixation time for the middle of the sentence. Thus, the words in the different sentence regions were not matched for frequency or length, which is why one should not directly compare the fixation times between the different sentence regions (both word frequency and length have a significant influence on eye fixation times during reading; see Hyönä & Olson, 1995; Rayner, 1998). Moreover, the first word and the TBR were not matched for these factors (the mean length and lemma frequency per million were 9.00, $SD = 3.41$ and 595, $SD = 1178$ for the first word in set-size 4; 8.75, $SD = 2.45$ and 56, $SD = 6$ for the TBR word in set-size 4; 8.13, $SD = 2.47$ and 701, $SD = 1041$ for the first word in set-size 5; and 7.73, $SD = 1.49$ and 56, $SD = 8$ for the TBR word in set-size 5). It is also typical that the sentence-final words attract longer gaze durations due to the sentence wrap-up effect (Rayner, Kambe, & Duffy, 2000). Thus, the effects of sentence region are only interesting if they emerge in interactions with the other variables of interest (memory load and reading span group).

of freedom due to unequal variances confirmed that all group means differed from each other, smallest $t(19) = 6.07$, $SE = 3.48$, $p < .001$, for the difference between medium and high span groups.

The data were analysed for set sizes four (i.e., there were four words to be recalled) and five (i.e., there were five words to be recalled). In set-size four, participants with a low and medium span were performing at their maximum performance level and they did not proceed to set-size five. Comparisons between the span groups at this level reveal whether the relatively poor performance of participants with a low and medium span was a consequence of different type of processing of the test materials, compared to that of participants with a high span. For set-size five data were available only for the high-span group. These data were analysed to examine whether processing strategies of participants with a high span changed when they were performing close to their maximum performance level.

Set-size 4. Fixation time measures were analysed with mixed 3 (Sentence region: beginning, middle, TBR) × 4 (Memory load: 0, 1, 2, or 3 words) × 3 (Span: low, medium, high) ANOVAs. Sentence region and load were within-participant factors and span a between-participants factor. Greenhouse-Geisser corrections to the degrees of freedom were made when the sphericity assumption was not met. An effect size estimate (partial eta squared) is reported for the main effects, interactions, and contrasts. The analysis of *gaze duration* (see Figure 1) showed that the span groups did not differ in their mean gaze duration, $F(2, 53) = .36$, $MSE = 1465.66$, $p = .69$, $\eta_p^2 = .01$, nor were there interactions between span and sentence region, $F(3, 80) = 1.19$, $MSE = 1508.64$, $p = .32$, $\eta_p^2 = .04$, span and memory load, $F(6, 159) = 1.02$, $MSE = 709.29$, $p = .42$, $\eta_p^2 = .04$, or a three-way interaction, $F(8, 210) = 1.03$, $MSE = 772.78$, $p = .42$, $\eta_p^2 = .04$.

The analysis revealed a significant main effect of sentence region, $F(2, 80) = 350.37$, $MSE = 1508.64$, $p < .001$, $\eta_p^2 = .87$. More importantly, a significant Sentence Region × Load interaction emerged, $F(4, 210) = 10.61$, $MSE = 772.78$, $p < .001$, $\eta_p^2 = .17$. In order to examine whether increased memory load resulted in a linear increase or decrease in gaze duration in the different sentence regions, we tested for linear contrast effects across the four memory load conditions separately for each sentence region[2]. Gaze duration in the sentence beginning decreased linearly from 54 ms/char in the no-load condition to 39 ms/char in the 3-word-load condition, $F(1, 55) = 8.49$, $MSE = 775.81$, $p = .005$, $\eta_p^2 = .13$. A linear decrease from 52 ms/char (no-load condition) to 41 ms/char (3-word-load condition) was also observed for the middle region, $F(1, 55) = 25.59$, $MSE = 123.24$, $p < .001$, $\eta_p^2 = .32$. In contrast, for the TBR word there was a linear *increase* from 106 ms/char (no-load) to 132 ms/char (3-word-load) $F(1, 55) = 12.43$, $MSE = 1221.27$, $p = .001$, $\eta_p^2 = .18$. The main effect of memory load was not significant, $F(3, 6) = .33$, $MSE = 709.29$, $p = .80$, $\eta_p^2 = .01$.

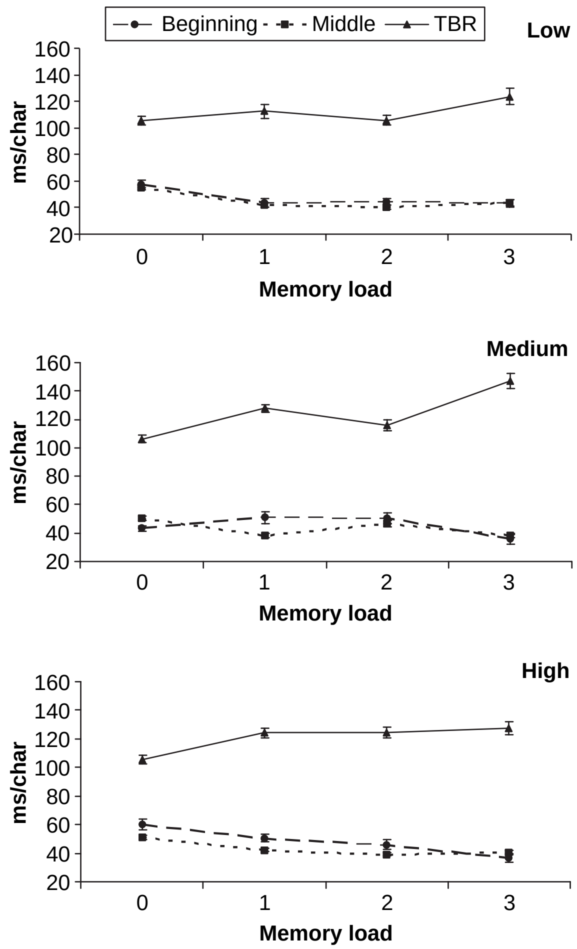

Figure 1. Mean gaze duration (in ms/char) as a function of sentence region and memory load separately for individuals with a low, medium, and high span at set-size 4. Error bars represent standard errors.

[2] Due to multiple levels of memory load, it would have been possible to also test for other types of contrasts (e.g., quadratic or cubic). However, as we had no a priori predictions for such effects, they were not tested.

In the *total fixation time* (see Figure 2), there was no evidence of span group differences, $F(2, 53) = .34$, $MSE = 2456.08$, $p = .72$, $\eta_p^2 = .01$, nor of an interaction between memory load and span, $F(6, 159) = 1.26$, $MSE = 245.73$, $p = .28$, $\eta_p^2 = .05$, sentence region and span, $F(4, 95) = .33$, $MSE = 1242.27$, $p = .53$, $\eta_p^2 = .03$, or memory load, sentence region and span, $F(8, 217) = 1.36$, $MSE = 217.45$, $p = .21$, $\eta_p^2 = .05$.

There was a significant main effect of sentence region, $F(2, 95) = 296.86$, $MSE = 1242.27$, $p < .001$, $\eta_p^2 = .85$. Moreover, the total time spent on reading the sentence increased as the memory load increased, as indicated by a significant main effect of memory load, $F(3, 159) = 16.41$, $MSE = 245.73$, $p < .001$, $\eta_p^2 = .24$. This main effect was modified by a two-way interaction between sentence region and memory load, $F(4, 217) = 23.27$, $MSE = 557.11$, $p < .001$, $\eta_p^2 = .31$. In order to examine whether increased memory load resulted in a linear increase or decrease in total fixation time for different sentence regions, we tested for linear contrast effects across the four memory load conditions separately for each sentence region. For the first word of the sentence the total fixation time decreased linearly from 79 ms/char in the no-load condition to 65 ms/char in the 3-word load condition, $F(1, 55) = 5.07$, $MSE = 518.66$, p = .028, $\eta_p^2 = .08$. In the middle region total fixation time did not monotonically increase or decrease, $F(1,55) = 1.64$, $MSE = 74.86$, $p = .206$, $\eta_p^2 = .03$, whereas on the TBR word the total fixation time linearly increased from 111 ms/char in the no-load condition to 157 ms/char in the 3-word load condition, $F(1, 55) = 75.52$, $MSE = 671.05$, $p < .001$, $\eta_p^2 = .58$.

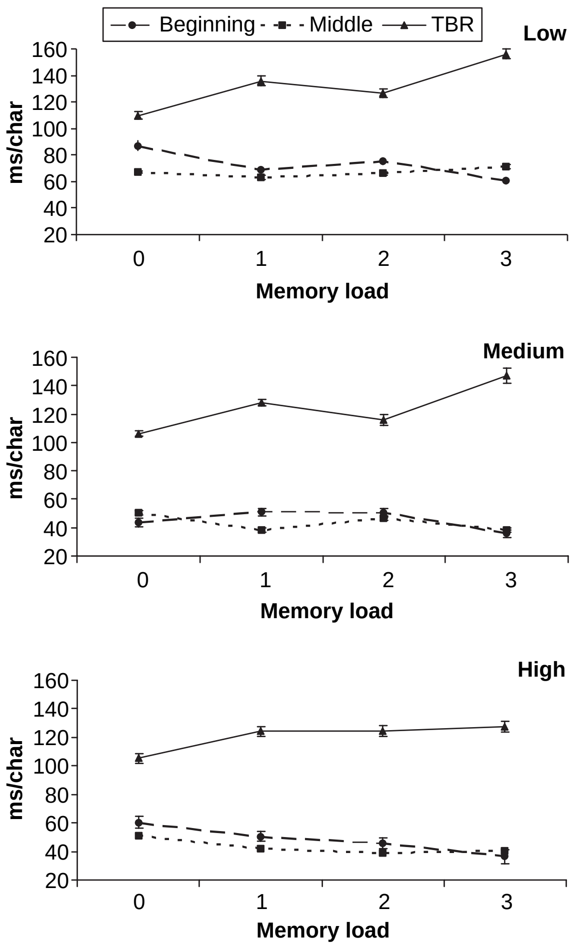

Figure 2. Mean total fixation time (in ms/char) as a function of sentence region and memory load separately for individuals with a low, medium, and high span at set-size 4. Error bars represent standard errors.

In sum, the analyses of gaze duration and total fixation time indicated that: (1) there were no span group differences in online processing of the test sentences and (2) as memory load increased, participants speeded up the reading of the irrelevant parts of the test sentences (beginning and middle region) and invested more time on the TBR word.

Set-size 5. Next, we analysed the high-span group's performance in set-size 5 with repeated measures 3 (Sentence region: beginning word, middle, TBR) × 5 (Memory load: 0, 1, 2, 3 or 4 words) ANOVAs. Greenhouse-Geisser corrections to the degrees of freedom were made when the sphericity assumption was not met. The effect size estimate (partial eta squared) is reported for the main effects, interactions, and contrasts.

In *gaze duration*, the main effect of sentence region was significant, $F(1, 24) = 83.94$, $MSE = 3665.03$, $p < .001$, $\eta_p^2 = .82$ (see Figure 3). The main effect of memory load, $F(4, 76) = 3.09$, $MSE = 639.35$, $p = .021$, $\eta_p^2 = .14$, indicates that the gaze durations changed as the memory load increased. Looking at the means, it can be seen that gaze duration decreased from 81 ms/char in the no-load condition to 70 ms/char in the 4-word-load condition. The Sentence Region × Memory load interaction did not reach significance, $F(4, 84) = 1.33$, $MSE = 838.76$, $p = .264$, $\eta_p^2 = .07$. However, in order to examine whether increased memory load differently affected processing of the different sentence regions, we tested for linear contrast effects across all five memory load conditions separately for each sentence region. The gaze durations linearly decreased from

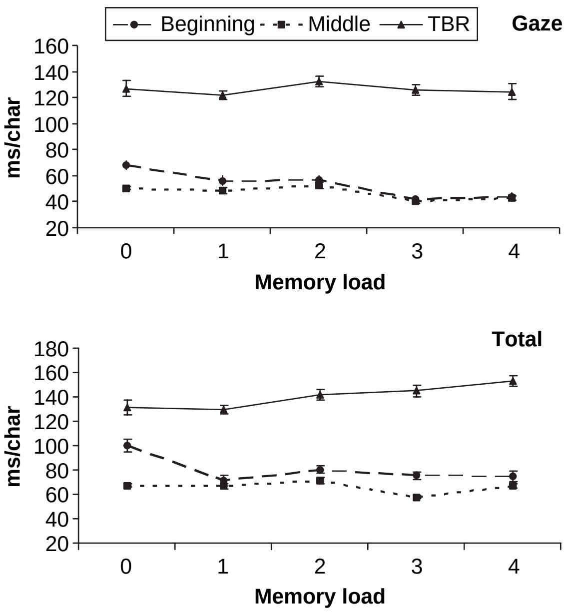

Figure 3. Mean gaze duration and total fixation time (in ms/char) as a function of sentence region and memory load at set-size 5 (only individuals with a high span). Error bars represent standard errors.

68 ms/char in the no-load condition to 44 ms/char in the 4-word-load condition for the sentence-initial word, $F(1, 19) = 13.38$, $MSE = 596.12$, $p = .002$, $\eta_p^2 = .41$. In the middle region, the gaze durations linearly decreased from 50 ms/char in the no-load condition to 43 ms/char in the 4-word-load condition, $F(1, 19) = 11.28$, $MSE = 100.44$, $p = .003$, $\eta_p^2 = .37$, whereas there was practically no change in gaze duration for the TBR word, $F(1, 19) = .000$, $MSE = 1200.74$, $p = .989$, $\eta_p^2 = .00$.

There was a significant main effect of sentence region in the *total fixation time,* $F(2, 38) = 70.37$, $MSE = 2191.49$, $p < .001$, $\eta_p^2 = .79$ (see Figure 3). Memory load also had a significant influence on the total fixation time, $F(4, 76) = 3.40$, $MSE = 343.28$, $p = .013$, $\eta_p^2 = .15$. More importantly, the main effect of memory load was qualified by a significant Sentence Region × Memory load interaction, $F(4, 83) = 4.60$, $MSE = 761.78$, $p = .002$, $\eta_p^2 = .20$. In order to examine the nature of the change as a result of increased memory load in different sentence regions, we tested for linear contrast effects across all five memory load conditions separately for each sentence region. The total fixation time decreased linearly from 100 ms/char in the no-load condition to 75 ms/char in the 4-word-load condition in the beginning of the sentence, $F(1, 19) = 10.33$, $MSE = 414.77$, $p = .005$, $\eta_p^2 = .35$. The linear contrast did not reach significance for the middle region, $F(1, 19) = 3.50$, $MSE = 61.41$, $p = .077$, $\eta_p^2 = .16$, whereas on the TBR word the total fixation time linearly increased from 131 ms/char in the no-load condition to 153 ms/char in the 4-word-load condition, $F(1, 19) = 11.93$, $MSE = 565.30$, $p = .003$, $\eta_p^2 = .39$.

In sum, the analyses of gaze duration and total fixation time for the high-span group performing with set-size 5 resemble the results obtained for all span groups in set-size 4: Participants allocated less processing time to the irrelevant parts of the test sentences (beginning and middle region) and more time to the TBR word as the memory load increased.

Discussion

We failed to observe differences between individuals with a low and high reading span in how they distributed processing time between different sentence regions in the reading span task. Specifically, we did not find evidence supporting the view that individual differences in the span task derive from the knowledge and use of different online processing strategies utilised during the task (McNamara & Scott, 2001). In some previous studies such differences have been found (Carpenter & Just, 1989; Engle et al., 1992; Friedman & Miyake, 2004). However, Engle et al. (1992) used a different version of the task and also a different administration method, whereas Friedman and Miyake (2004) only report global sentence reading times, which makes it difficult to compare our results directly with theirs. On the other hand, the study of Carpenter and Just (1989) resembles our experiment quite closely. It should be noted that Carpenter and Just report data from only 6 individuals with low and 6 individuals with a high span, whereas we report data from 19 individuals with low and 20 individuals with a high span. In order to more closely replicate their procedure, we also conducted analyses for only the extreme groups (individuals with a low and high span) but none of the interactions involving span group was significant: largest $F(4, 149) = 1.46$, $MSE = 581.07$, $p = .216$, $\eta_p^2 = .04$ for the three-way interaction in the total fixation time. Thus, even though we had more power for detecting group differences than Carpenter and Just, we failed to replicate their findings. The span groups in the present study did differ significantly in the span task performance (as measured by the number of correctly recalled words), and similar differences in span performance have successfully been related to individual differences in reading tasks (Kaakinen et al., 2003), so the lack of a span group effect in Experiment 1 is not caused by a lack of genuine differences in the working memory capacity.

Consistently with the previous studies (Carpenter & Just, 1989; Engle et al., 1992; Friedman & Miyake, 2004), we found that memory load had a significant impact on processing time devoted to the test materials. In all analyses of the two processing time measures (gaze duration and total fixation time) and two set sizes (four and five), the processing time allocated to the beginning or middle regions of the sentences (i.e., irrelevant to the memory task) decreased as the memory load increased. In contrast, in three analyses out of four, the processing time directed to the TBR words increased together with the memory load. In other words, participants adjusted their online processing strategies to cope with the increasing task demands and adopted a selective processing strategy. It should be noted that the experimenter-paced administration method used in the present study poses limitations to the time available to process the test materials. Because it reduces variability in strategy use, it produces higher correlations with, for example, reading comprehension measures than self-paced administration (Conway et al., 2005; Friedman & Miyake, 2004). However, our results imply that experimenter pacing does not completely eliminate strategy use but may indeed reduce variability in what type of strategies are implemented. In the present experiment there were no individual differences related to working memory span with respect to what type of processing strategies were utilised. All participants selectively allocated more processing resources to task-relevant information and less time to task-irrelevant information.

EXPERIMENT 2

The results of Experiment 1 suggest that individual differences in the reading span task performance are not a consequence of different online processing strategies utilised during the span task. Instead, all participants seemed to adopt a selective processing strategy and invested more processing time to the TBR word at the expense of the sentence up to the TBR word, as the memory load increased. It is possible that even though individuals with low and high spans use a similar online processing strategy, they do differ in the use of memory encoding strategies. Previous research suggests that individuals with a high span make more use of elaborative encoding strategies, such as chaining or imagery, than individuals with a low span (Dunlosky & Kane, in press; Friedman & Miyake, 2004; McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003). In Experiment 2, we collected general reports (see Dunlosky & Kane, in press) of the strategy use during the reading span task. We expected that individuals with a high span would report encoding strategies that involve elaborating the TBR information by semantic chaining, imagery, or associations. Individuals with a low span, on the other hand, were expected to report either less strategy use or use of less efficient encoding strategies, such as rehearsal.

Method

Participants. A total of 53 University of Turku students volunteered in the study. The mean age of the participants was 24.21 years ($SD = 5.59$). None of the participants in Experiment 2 had participated in Experiment 1.

Procedure. Participants were tested with the reading span task using the same procedure and scoring method as in Experiment 1. After finishing the task, they were interviewed about the strategies they used in the task, to obtain general reports of strategy use (see Dunlosky & Kane, in press). First, they were asked, “Did you use some strategy in trying to remember the final words?” If participants indicated that they had used some strategy, they were asked to further specify what kind of a strategy it had been. Participants were not given options from which to choose but were allowed to freely describe as many strategies as they liked. The experimenter wrote down the orally given answers.

Results

Most of the participants reported the use of one or more strategies in the reading span task. Only one participant responded that he had not tried to use any strategy. The transcribed responses were categorised into three categories: (1) Semantic elaboration, which included chaining of the final words (e.g., trying to form meaningful sentences out of the TBR words), associating TBR words with either personal experiences or information in long-term memory (e.g. the word was related to a TV show the participant had watched the previous night), or imagery, which meant that the participant tried to visualise the TBR words and then tried to form mental scenes that included the meaning of all the TBR words; (2) silent rehearsal of the TBR words while reading the sentences, and (3) selective processing of the sentence (e.g., " I tried not to pay attention to the meaning of the sentence and just concentrated on the last word").

The 53 participants produced a total of 80 responses. The most often reported strategy was semantic elaboration (51% of all responses), followed by rehearsal (33%) and selective processing (15%). Using the same cut points as in Experiment 1 for the total number of correctly recalled words, the participants were divided into low- ($N = 16$), medium- ($N = 16$), and high-span ($N = 21$) groups. The means of total number of recalled words were 19.63 ($SD = 2.33$), 22.31 ($SD = .48$), and 43.48 ($SD = 12.68$) for low-, medium-, and high-span groups, respectively. The participant (classified as low span) who reported no strategy use was excluded from the subsequent analyses. The percentages of reported strategies within each span group are presented in Figure 4.

The span groups differed in what kind of memory encoding strategies they used in the task, $\chi^2(4) = 14.22$, $N = 79$, $p = .007$. The individuals with low reading span most often reported simple rehearsal of the TBR words, whereas semantic elaboration and selective processing were less often reported. As for the individuals with medium span, the most frequently reported strategy was semantic elaboration, followed by

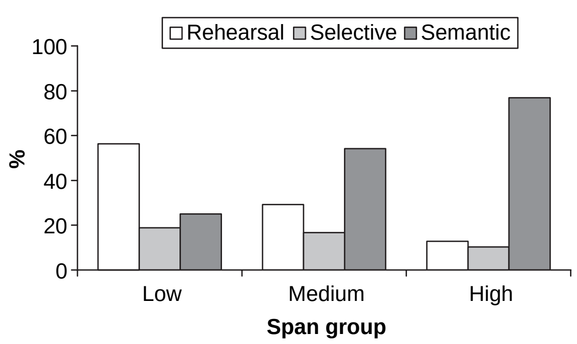

Figure 4. Reported strategy use (% of the strategies reported by the group) separately for individuals with a low, medium, and high span.

rehearsal and selective processing. Individuals with a high span reported mostly semantic elaboration: about 77% of the strategies were elaborative. Rehearsal and selective processing were relatively rare in this group.

Discussion

The results showed that during the reading span task, individuals with a low span mainly used inefficient memory encoding strategies such as rehearsal, whereas individuals with a high span implemented more efficient elaborative memory encoding strategies. In other words, the probability of using elaborative encoding strategies increased with memory span, whereas the probability of using less efficient strategies (rehearsal) decreased with memory span. These results are in line with previous studies (Dunlosky & Kane, in press; Friedman & Miyake, 2004; McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003) and support the view that individual differences in the reading span task derive (at least partly) from the use of efficient memory encoding strategies (McNamara & Scott, 2001).

It should be noted that all except for one participant reported the use of some strategy—thus, the differences in the span task performance were not caused by some individuals not using any strategy at all and others being more strategic. Rather, what seemed to be crucial was what type of a memory encoding strategy was adopted. Elaborative strategies such as imagery and verbal chaining of the TBR words were more beneficial in the span task than simply rehearsing the TBR words.

But why did individuals with a low span tend adopt an inefficient rehearsal strategy and individuals with a high span a more efficient elaborative strategy? Were the individuals with a low span not aware of the possibility of using elaborative strategies? A closer look at the data reveals that about 25% of the individuals with a low span actually did try to use an elaborative encoding strategy. Despite the use of a "good" strategy, they were not very successful in the task (i.e., their performance level was still in the lowest tertile). This finding is in line with the results of a strategy training study by Turley-Ames and Whitfield (2003) and suggests that even though strategy use is an important factor in the reading span task performance, there may be more to it than simple knowledge about what is a good strategy. Turley-Ames and Whitfield (2003) examined the impact of strategy instruction on the operation span task performance and found that individuals with a low span only benefited from rehearsal strategy instruction—instructions to use semantic strategies were not helpful for them. Turley-Ames and Whitfield argue that semantic strategies require cognitive resources and that because individuals with a low span have more limited resources available, they cannot make use of semantic strategies. Thus, it seems that even though the use of elaborative memory encoding strategies is beneficial in the span task, one must have enough cognitive resources to implement them efficiently. This argument is discussed in more detail in the General Discussion.

GENERAL DISCUSSION

The present study examined the view that individual differences in the reading span task derive from strategy use during the task (McNamara & Scott, 2001). Specifically, we examined the hypotheses that individuals with low and high spans differ either in strategies used for online processing of the test materials (Experiment 1) or in memory encoding strategies utilised to encode and maintain the TBR words (Experiment 2). The results demonstrate that when performing the reading span task individuals with low and high spans do not differ in how they distribute processing time within the sentences. On the other hand, when interviewed for the memory encoding strategies used in the task, individuals with a low span mainly reported the use of silent rehearsal of the TBR words, while individuals with a high span mostly reported of the use of elaborative encoding strategies. In sum, the likelihood of effective elaborative encoding strategies increased and the likelihood of less effective simple rehearsal decreased as the individual's memory span increased. Thus, what is crucial in the individual differences in the reading span task is not the strategic allocation of processing time on the test materials, but rather what types of memory strategies are utilised.

The present results suggest that individuals with a low span are not completely ignorant about the availability of efficient elaborative strategies. About 25% of the individuals with a low span did try to use an elaborative encoding strategy but with no apparent success. An interesting anecdote is that one of the participants with a low span who reported trying to form sentences of the TBR words commented that he was not very successful in using this strategy because there was not enough time. Moreover, results of Experiment 2 showed that individuals with a high span used mainly elaborative memory encoding strategies (with apparent success), while the results of Experiment 1 suggest that individuals with a high span did not need extra time for elaborative processing. Even though a null result should always be considered with caution, these findings together with previous results (Turley-Ames & Whitfield, 2003) indicate that individuals with a high span are highly efficient in making use of elaborative strategies. Thus, span task performance seems to be correlated with the ability to effectively use elaborative encoding strategies (cf. Turley-Ames & Whitfield, 2003).

According to the *long-term working memory* (LT-WM) theory (Ericsson & Delaney, 1999; Ericsson & Kintsch, 1995), the crucial difference between the experts and novices in a task is that the experts have more organised knowledge structures, which allows them to rapidly access information in the knowledge base and to build links between new information by using their existing knowledge structures. Our results suggest that individuals with a high span, the experts in the task, are able to encode the task-relevant information to memory without spending extra time (in comparison to other participants) on it by elaborating on the task-relevant information—a strategy that involves retrieving information from long-term memory and linking TBR information with it.

Another possible explanation is that one needs a sufficient amount of attentional capacity to be able to strategically make use of long-term memory (LTM). For instance, Rosen and Engle (1997) compared individuals with low and high spans in a verbal fluency task and found no span group differences in the richness of the knowledge base. However, the span groups differed in how strategic they were in retrieval: for individuals with a low span, retrieval of category exemplars seemed to be automatic in nature, thus not taxing the attentional system, whereas individuals with a high span engaged in strategic retrieval of elements from LTM. It is thus possible that in addition to tapping into strategy knowledge and use (McNamara & Scott, 2001), the reading span task may also tap into the attentional capacity required for activating and maintaining the activation of LTM representations (e.g., Barrouillet, Bernandin, & Camos, 2004; Lépine, Barrouillet, & Camos, 2005). According to this view, high-span participants have more attentional capacity to activate LTM representations, which is required for building elaborative links between the TBR elements. They also have attentional capacity to maintain the activation of LTM representations while reading the sentences, which decreases the decay rate of the memory traces and results in a successful recall of the TBR information.

The present exploration of the strategy use in the reading span task does not completely resolve the issue of what the reading span task actually measures, and why it is such a good predictor of complex cognitive skills such as reading comprehension, but it does shed additional light on the question by highlighting the importance of strategy use and especially the ability to make use of strategies involving LTM encoding and retrieval (see Kaakinen et al., 2003). We conclude that future studies are necessary to define the intriguing relationships between working memory, attentional control, and long-term memory.

Manuscript received 6 October 2007
Manuscript accepted 14 May 2007

REFERENCES

Barrouillet, P., Bernandin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*, 83–100.

Carpenter, P. A., & Just, M. A. (1989). The role of working memory in language comprehension. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 31–68). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Conway, A. R. A., Jarrold, C., Kane, M. J., Miyake, A., & Towse, J. N. (2007). *Variation in working memory*. New York: Oxford University Press.

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 and Review*, *12*, 769–786.

Cowan, N. (2005). *Working memory capacity*. Hove, UK: Psychology Press.

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. (in press). The contributions of strategy use to working memory span: A comparison of strategy-assessment methods. *Quarterly Journal of Experimental Psychology*.

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.

Ericsson, K. A., & Delaney, P. F. (1999). Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). Cambridge, UK: Cambridge University Press.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.

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.

Friedman, N. P., & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods*, *37*, 581–590.

Hyönä, J., & Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1430–1440.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.

Kaakinen, J. K., Hyönä, J., & Keenan, J. M. (2003). How prior knowledge, WMC, and relevance of information affect eye fixations in expository text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 447–457.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology: General*, *130*, 169–183.

Laine, M., & Virtanen, P. (1999). *WordMill Lexical Search Program*. Turku, Finland: Centre for Cognitive Neuroscience, University of Turku.

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

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

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

Miyake, A. (2001). Individual differences in working memory: Introduction to the special section.

Journal of Experimental Psychology: General, *130*, 163–168.
Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press.
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.
Rayner, K., Kambe, G., & Duffy, S. A. (2000). Clause wrap-up effects on eye movements during reading. *Quarterly Journal of Experimental Psychology*, *53A*, 1061–1080.
Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, *126*, 211–227.
Towse, J. N., Hitch, G. J., & Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory & Cognition*, *28*, 341–348.
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 task dependent? *Journal of Memory and Language*, *28*, 127–154.