INDIVIDUAL DIFFERENCES OF HOW ATTENTION IS ALLOCATED DURING READING

by

Sarah A. Orban

Submitted to the Undergraduate Faculty of

the School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Bachelor of Philosophy

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Sarah A. Orban

It was defended on

April 14, 2010

and approved by

Natasha Tokowicz, Ph.D., University of Pittsburgh, Department of Psychology

Tessa Warren, Ph.D., University of Pittsburgh, Department of Psychology

Patryk Laurent, Ph.D., Johns Hopkins University, Department of Psychological and Brain Sciences

Thesis Advisor: Erik Reichle, Ph.D., University of Pittsburgh, Department of Psychology

INDIVIDUAL DIFFERENCES OF HOW ATTENTION IS ALLOCATED DURING READING

Sarah A. Orban

University of Pittsburgh, 2010

The current study replicates a previous finding of how attention is allocated during reading and was expanded upon by controlling for individual differences between participants. An eye-tracking experiment was performed to determine how attention is allocated during reading, while individual differences between participants were recorded by measuring working memory capacity. In four tasks that increased in depth of processing, participants were instructed to correctly identify whether or not a target was presented in a series of 1-4 words. Results indicate a relationship between reaction time and working memory score in all but one task. This suggests that high-span individuals use parallel processing when detecting symbols or orthographic features, but use serial processing during tasks that require full semantic processing.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

This research was funded by the National Institute of Health grant RO1HD053639. I would like to thank my committee chair and thesis advisor Erik Reichle and committee members Tessa Warren, Natasha Tokowicz, and external examiner Patryk Laurent for their outstanding feedback and suggestions. I would also like to thank Amanda Virbitsky and Polina Vanyukov for their support and contributions.

1.0 INTRODUCTION

A current debate in the field of cognitive science deals with how attention is allocated during reading. Two models in particular offer disparate explanations of this phenomenon (see Reichle, Liversedge, Pollatsek, & Rayner, 2009). The two models of interest are the serial processing model, E-Z Reader, on one end of the spectrum (Reichle, Warren, & McConnell, 2009) and the attention gradient model, SWIFT, on the other (Engbert, Nuthmann, Richter, & Kliegl, 2005). The serial processing model proposes that, during reading, attention is focused on the lexical processing of one word at a time. Whereas, the attention gradient model assumes that allocation of attention is distributed as a gradient to support the lexical processing of more than one word at a time.

1.1 SUPPORT FOR SERIAL ALLOCATION OF ATTENTION

In a recently published article by Reichle, Vanyukov, Laurent, and Warren (2008), it has been implicated that allocation of attention during reading is done one word at a time, thus supporting the serial processing model (Reichle et al., 2009). This experiment examined the saccadic eye movements, reaction time, and accuracy of lexical processing of 1-4 simultaneously displayed words as participants performed tasks that varied in terms of their “depth” of processing (Craik & Lockhart, 1972). The experiment began with a relatively easy “shallow” task in which participants were instructed to detect an asterisk that, on some of the trials, was embedded in one of the 1-4 displayed words (i.e., asterisk detection). The "deep" processing tasks involved determining whether the letter 'q' was contained in any of the displayed words (i.e., letter-detection), whether any of the displayed words rhymed with 'blue' (i.e., rhyme-judgment), or whether any of the displayed words referred to an animal (i.e., semantic-judgment). It was hypothesized that if attention were allocated serially, response times would increase as the depth of processing increased (e.g., asterisk-detection < letter-detection < rhyme-judgment < semantic-judgment), and as the number of words being simultaneously displayed increased. Also, saccadic eye movements should be directed to the left and proceed to the right through the array of words—particularly for the deeper processing tasks. But, if attention was allocated as a gradient, then the reaction times should be less affected by the number of words being displayed in the array and the initial saccades should be centered in the array of words, with more processing in the periphery and less of a tendency to scan from left to right. The results of this experiment were more consistent with the serial-attention hypothesis because, not only were the deeper processing tasks marked by increased reaction times, but reaction times were also affected by the number of words that were displayed on the screen. The eye-tracking results further supported the serial processing model by showing that initial saccades tended to be initially directed towards the left of the array of words but then proceeded from left to right, consistent with the hypothesis that lexical processing was being completed on one word at a time.

1.2 INDIVIDUAL DIFFERENCES IN ATTENTION ALLOCATION

The current experiment extends the one by Reichle et al. (2008) in that it investigates how individual differences in working memory capacity (Carpenter, Just, & Reichle, 2000; Just & Carpenter, 1992; Reichle, Carpenter, & Just, 2000) play a role in determining where certain individuals fall on this controversial spectrum of attention allocation during reading. In other words, the present experiment is designed to investigate if individual variability in working memory capacity might cause individuals to conform to varying degrees to the behavioral profiles that have classically been interpreted as being indicative of serial versus parallel attention allocation (e.g., see Thornton & Gilden, 2007). The study replicates Reichle's et al. original design and adds an important factor, a working memory task to separate individuals with high-span working memory capacity from individuals with low-span working memory capacity.

1.2.1 WORKING MEMORY CAPACITY

Working memory capacity (*WMC*) is consistently described as reflecting the processing and storage capabilities of a multi-component limited capacity system that is responsible for active maintenance and temporary storage of information in the face of concurrent processing of information and/or distraction (Conway, Kane, Bunting, Hambrick, Wilhelm & Engle, 2005; Kofler, Rapport, Bolden, Sarver, & Raiker, 2009). However, it is important to note that researchers claim there still is no universally agreed upon definition of WMC (Barrett, Tugade, & Engle, 2004). Working memory tests, especially the reading span tests, are a popular and useful construct in measuring a wide range of complex cognitive abilities, such as reasoning, problem solving, and comprehension, and as such provide a widely used measurement tool in cognitive psychology (Conway et al., 2005). Moreover, previous studies indicate that performance on working memory span tasks are moderately correlated with verbal SAT scores with *r*s ranging from .49 to .59 (Kane, Conway, Bleckley, & Engle, 2001).

The working memory test that was used in the current study was a reading span test (Daneman & Carpenter, 1980; implemented and administered using a automatic, web-based interface; Loboda, 2009) that requires the participant to process information, such as whether or not a sentence is logically correct, while concurrently using short-term memory storage to remember a series of letters that appear after each sentence. After each sequence of sentences appears, the participant is prompted to recall the letters in the order in which they appeared after the sentences. By definition, those participants who process the sentences of the span task more efficiently have more capacity to store the information at the end of each sentence, thereby allowing those participants with a high WMC to recall more of the letters than those with a low WMC (Kane et al., 2001).

1.3 CURRENT STUDY

Because the current tasks rely heavily upon controlled and focused attention, especially in the deeper processing tasks (e.g. rhyming- and semantic-judgment tasks), the capability to control attention in order to focus on task relevant information should be directly related to an individual's ability to perform sufficiently on a complex working memory task. In addition, those participants with high WMC should exhibit more rapid and accurate retrieval of goal-relevant information than low working memory individuals (Barrett et al., 2004). Thus, if controlled processing is affected by WMC and if attention is allocated in a serial manner, then reaction times for identifying the targets should increase as the depth of the processing task increases. However, reaction times during the shallow, asterisk detection task should be less affected by a participant's WMC because the asterisk should simply "pop out" of the display (Reichle et al., 2008) because the detection of a single simple feature (e.g., an asterisk) is dependent upon the automatic rather than controlled processing of information. This prediction may also be enhanced by the fact that those participants with low WMC often display difficulties inhibiting automatic responses (Barrett et al., 2004).

Another important prediction is that high-span working memory individuals should show faster overall reaction times, whereas low-span working memory individuals should show slower reaction times. This prediction follows logically from Reichle's et al. (2008) experiment, which only reported the average of individual ability to process these tasks; because working memory capacity and the capacity to control attention are posited to be normally distributed traits, there should be individuals who fall on the extremes of the distributions. These individuals may not only differ in their working memory span or history of reading and/or learning difficulties, but may also adopt different styles of attention allocation during reading.

For example, high-span individuals may allocate their attention during reading in a manner that is more consistent with the attention-gradient model (e.g., SWIFT; Engbert et al., 2005) because these individuals have the capacity to process several words at a time. The response times of high-span individuals might thus be less affected by the number of concurrently displayed words because these individuals possess efficient cognitive mechanisms that allow for effective processing of several words at a time. In contrast, low-span individuals, who tend to use ineffective cognitive mechanisms to guide attention, may display longer reaction times as well as a pattern of response times that is consistent with attention being directed in a serial manner, to only one word at a time during reading (e.g., as in the E-Z Reader model; Reichle et al., 2009). Although we predict that task performance accuracy will be approximately the same for both high- and low-span individuals, high-span individuals are predicted to show faster reaction times than low-span individuals. Specifically, it is predicted that reaction times for the more shallow asterisk-detection task will be less affected by working memory capacity, but that the reaction times of the deeper processing tasks (i.e. letter-detection, rhyme-judgment, and semantic-judgment) will be most affected by working memory capacity.

2.0 METHODS

2.1 PARTICIPANTS

Twenty-seven undergraduates from the University of Pittsburgh with normal or corrected-to-normal vision performed two separate tasks: reading span (which was used to measure working memory capacity) and the set of four attention-allocation tasks that were of primary interest (and that will be described below). Participants completed the experiment to fulfill partial course credit in an introductory psychology course and all participants gave informed consent that had been approved by the University of Pittsburgh's Institutional Review Board prior to their participation.

2.2 WORKING MEMORY TASK

2.2.1 Materials

Eighteen trials of a reading span task were used to measure working memory capacity. Ninety-seven sentences were presented for participants; participants were instructed to determine the logical correctness of each sentence. (One randomly selected word within half of the sentences was replaced with a word that made the sentence nonsensical.) The sentence length ranged from 10 to 15 words. Sentences were displayed to each participant in a randomized order. Each sentence was presented for a 14 s maximum limit. A random subset of twelve possible letters (i.e. F, H, J, K, L, N, P, Q, R, S, T, Y) were randomly displayed for 1000 ms after each randomly selected sentence. Participants were instructed to remember the letters for a recall test. Each of 18 trials consisted of a random set size ranging from 2-7 sentences and letters. In a set size of two, for example, a participant would decide if two sentences were logically correct and recall the two letters that were presented after each sentence at the end of each trial.

2.2.2 Procedure

Participants began the reading span task by independently reading the directions from a computer screen and completing a practice trial in order to prepare them for the main task. A random set size of 2-7 sentences was displayed and the participant was to indicate if the sentence was logically correct or incorrect. After the participant clicked the correct or incorrect response key, a letter would appear for 1000 ms, followed by another sentence, etc. The participant was instructed to remember each letter in the correct order in which it had appeared after each sentence. At the end of each trial, the twelve letters were displayed in a 3 x 4 matrix in order for the participants to select which letters had appeared after each sentence. A blank box was included if the participant did not remember some letters, but remembered other letters.

2.2.3 Equipment

The reading span task implemented to measure working memory capacity was originally developed by Daneman and Carpenter (1980) and administered via an online program (Loboda, 2009) during the experiment.

2.3 READING ATTENTION TASK

2.3.1 Experimental Design

The presentation of the four attention-allocation tasks were blocked, with each block consisting of 50 one-word trials, then 50 two-word trials, and so on. Both task and number of words per trial were blocked in this manner to encourage participants to use whatever strategies they might find most effective—including parallel processing to the extent that it might facilitate task performance. Task blocks were presented in random order. The eye movements of 27 of these participants were recorded. Participants took short breaks between each blocked trial and then the eye-tracker was recalibrated as necessary.

2.3.2 Materials

Reading attention was measured by four tasks in blocks of 200 trials per task: (1) *asterisk-detection*; (2*) letter-detection*; (3) *rhyme-judgment*; and (4*) semantic-judgment*. Words 20-100 per million in frequency (Francis & Kucera, 1982) and 4-10 letters in length were selected from the MRC Psycholinguistic Database (Coltheart, 1981) as distractors; these words were divided into four sets of 460, with the sets being rotated through each of the task conditions using a Latin-square design. Forty target words were also selected for each task; these words were divided into four sets of 10, with the sets being rotated through each of the number-of-words conditions using a Latin-square design. The mean frequency (and range) of the target words in each of the task were: asterisk-detection = 16.53 (10-40); letter-detection = 14.88 (1-143); rhyme-judgment = 89.25 (0-1791); and semantic-judgment = 10.38 (0-117). The mean length (and range) of the target words were: asterisk-detection = 6.68 (4-11); letter-detection = 7.22 (4-12); rhyme-judgment = 5.08 (4-8); and semantic-judgment = 6.15 (4-11). The mean orthographic neighborhood density (and range) of the target words (Balota et al., 2007) were: asterisk-detection = 3.03 (0-19); letter detection = 0.63 (0-6); rhyme-judgment = 2.58 (0-10); and semantic-judgment = 4.10 (0-20). Finally, the mean number of morphemes (and range) were: asterisk-detection = 1.48 (1-3); letter detection = 1.50 (1-3); rhyme-judgment = 1.15 (1-2); and semantic-judgment = 1.13 (1-3). Although pair-wise comparisons did indicate a few reliable differences in the properties of target words across tasks, these differences always worked against the predicted depth-of-processing effects. Such differences are also not unexpected because the assignment of words to conditions is by definition a quasi-experimental manipulation (e.g., see Kliegl, Nuthmann, & Engbert, 2006).

The selection of target and non-target words was exclusive; i.e., a non-target word in one task could not be a target word in another task. Target locations in 2-, 3-, and 4-word trials were equally distributed within and between subjects. Stimuli presentation was done using E-Builder software (SR Research Ltd).

2.3.3 Procedure

At the beginning of each trial, a fixation cross appeared in the center of the screen, for 350 ms. Because participants were not required to maintain word order or complete any higher-level language processing (e.g., syntactic parsing) that is necessary to understand real text, the centrally displayed fixation cross should have conducive to optimal task performance by allowing lexical processing from a viewing location that afforded both maximal visual acuity and maximal flexibility in how attention was allocated to the words that were displayed. The fixation cross was then followed by the stimuli (1-4 words displayed simultaneously), with the word(s) displayed on a single line, centered on the screen for up to 3000 ms or until a response was made. Participants used a game-pad controller to indicate their responses during the tasks. Participants were instructed to press as quickly as possible the button under their left thumb on a gaming controller after locating a target (e.g. a word containing the letter "q"). Participants were instructed to press as quickly as possible the button under their right thumb on the gaming controller if none of the displayed words contained a target. The trial sequences for the other three tasks were structured in exactly the same manner, with only the task (e.g. press the left button if any of the words rhymes with "blue") being different.

2.3.4 Equipment

Participants viewed the stimuli binocularly on a 23-in. monitor 63 cm from their eyes with approximately two letters per 1° of visual angle. An Eye-Link 1000 eye-tracker (SR

Research Ltd.) recorded the gaze location of participants' right eyes. The eye-tracker had a spatial resolution of 0.01° and sampled gaze location every millisecond.

3.0 RESULTS

3.1 BEHAVIORAL RESULTS

3.1.1 Reaction Time

Figure 1 shows the reaction times (in ms) as a function of both task type and number of words displayed per trial for the trials in which the target was correctly identified. These data were examined using a repeated-measure *Analysis of Variance* (ANOVA) with *task type* (asterisk-detection vs. letter-detection vs. rhyme-judgment vs. semantic-judgment) and *number of words* (1 vs. 2 vs. 3 vs. 4) as within-subject factors. The results of the ANOVA indicate main effects of task type [$F(3,78) = 365.62$, $MSe = 19833.66$, $p < .001$] and number of words [$F(3,78) = 122.86$, $MSe = 68680.61$, $p < .001$], and an interaction between them [$F(9,234) = 73.795$, $MSe = 6692.16$, $p < .001$].

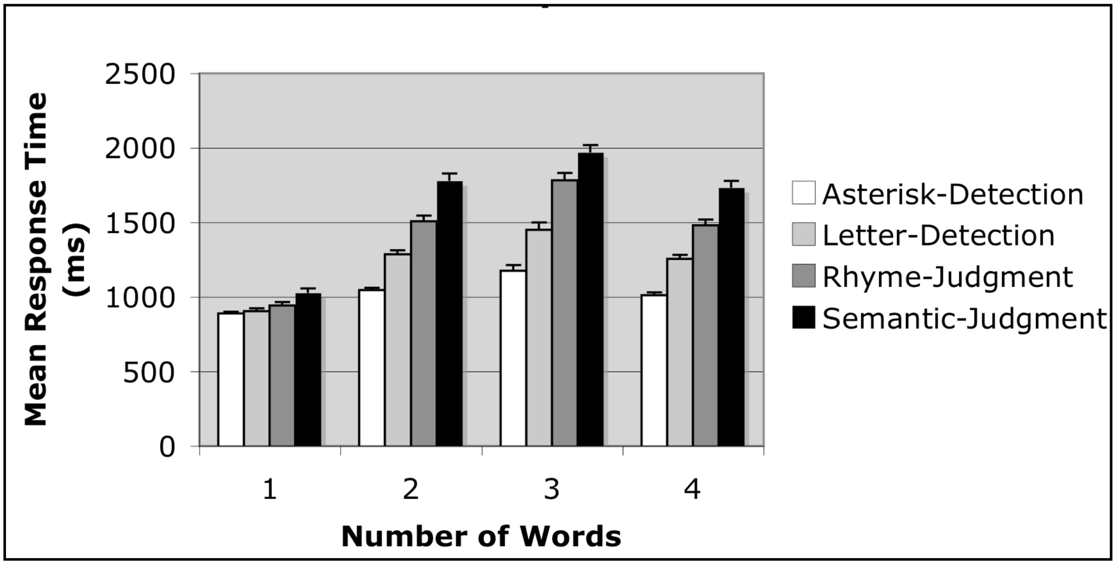

Figure 1. Reaction times (ms) as a function of task type and number of words

Further post-hoc analyses using pair-wise comparisons were performed to examine the precise nature of Task Type x Number of Words interaction. These comparisons were performed using a Bonferroni adjustment for multiple comparisons. All of the pair-wise comparisons for the asterisk-detection task to examine the effect of number of displayed words (e.g., 1 vs. 2, 1 vs. 3, etc.) were reliable [all $t(26)$s > 3.00, ps < .006], with the exception that the pair-wise comparison between the 2-word versus 4-word condition was not reliable [$t(26) = .14$, $p = .890$]. The pair-wise comparisons for the letter-detection task were reliable [all $t(26) > 3.52$, $p < .008$] except for the comparison between the 2-word versus 4-word condition [$t(26) = .033$, $p = .974$]. Likewise the pair-wise comparisons for the rhyme-judgment task were reliable [all $t(26)$s > 5.31, ps < .008] except for the pair-wise comparison between the 2-word versus 4-word condition [$t(26) = .040$, $p = .969$]. And finally, the pair-wise comparisons for the semantic-judgment task were reliable [all $t(26) > 4.32$, $p < .008$] except for the comparison between the 2-word versus 4-word conditions [$t(26) = .636$, $p = .530$].

3.1.2 Response Accuracy

Figure 2 shows the response accuracies (i.e., the total percentage of hits for both yes and no trials) as a function of task type and number of words displayed. Overall, the participants performed all four tasks very accurately (at least 92% correct in all four tasks). A repeated-measure ANOVA using task type and number of words as within-subject factors indicated significant main effects for task type and number of words [$F(3, 78) = 17.41$, $MSe = 13.35$, $p < .001$; and $F(3, 78) = 23.09$, $MSe = 7.08$, $p < .001$], and for an interaction between them [$F(9, 234) = 2.686$, $MSe = 7.34$, $p < .005$].

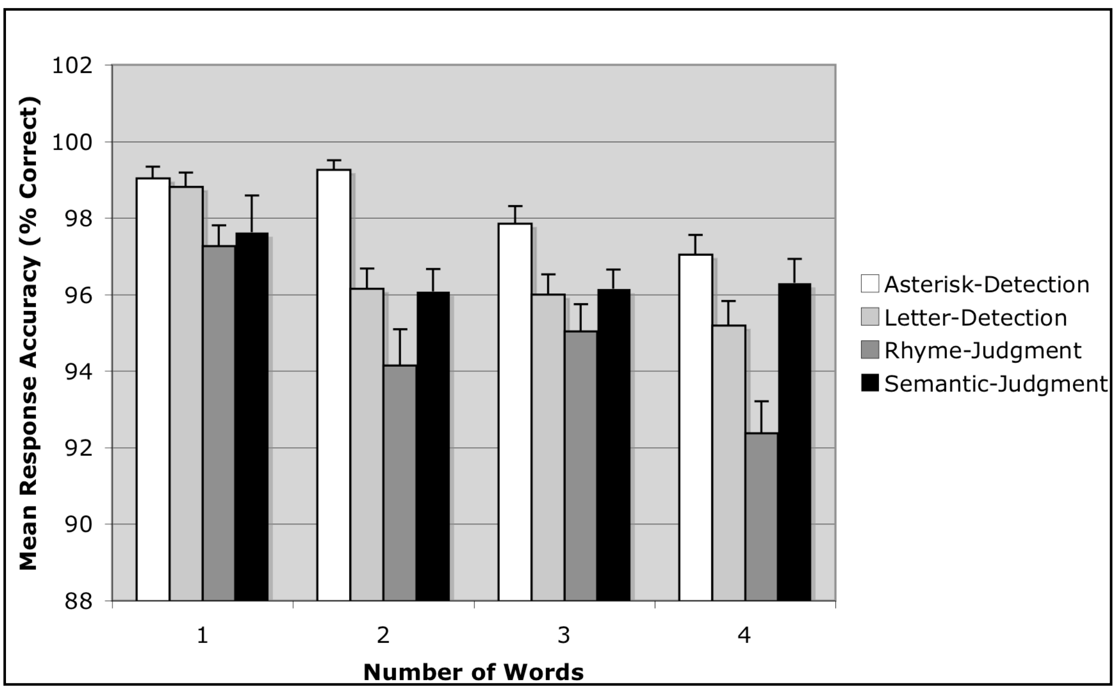

Figure 2. Response accuracies (%) as a function of task type and number of words

Pair-wise comparisons of each task were again performed to determine the effect of number of words displayed on response accuracy. Pair-wise comparisons of the asterisk-detection task were reliable between the 1-word versus 4-word, 2-word versus 3-word, and 2-word versus 4-word conditions [all $t(26)$s > 2.89, ps < .001], but not for the comparisons between the 1-word versus 2-word, 1-word versus 3-word, and 3-word versus 4-word conditions [all $t(26)$s < 2.53, ps > .133]. Pair-wise comparisons of the letter-detection task were reliable for the 1-word versus 2-word, 1-word versus 3-word, and 1-word versus 4-word conditions [all $t(26)$s > 2.89, ps < .001], but not for the comparisons between the 2-word versus 3-word, 2-word versus 4-word, and 3-word versus 4-word conditions [all $t(26)$s < 1.34, ps > .191]. Pair-wise comparisons of the rhyme-judgment task were reliable between the 1-word versus 2-word, 1-word versus 3-word, 1-word versus 4-word conditions, and 3-word versus 4-word conditions [all $t(26)$s > 2.85, ps < .001], but not for the comparisons between the 2-word versus 3-word or 2-word versus 4-word conditions [all $t(26)$s < 1.75, ps > .092]. All of the pair-wise comparisons of the semantic judgment task were unreliable [all $t(26)$s < 1.62, ps > .118].

3.2 EYE-TRACKING RESULTS

3.2.1 Mean Number of Fixations

Figure 3 shows the mean number of fixations that were computed until the target was correctly identified, as a function of both the task being performed and number of words displayed. Because of problems with calibration, only the eye-movements of 19 participants were included in the analyses. A repeated-measure ANOVA using task type and number of words as within-subject factors indicated significant main effects for task type and number of words [$F(3, 54) = 2.99$, $MSe = 6.17$, $p < .04$; and $F(3, 54) = 78.74$, $MSe = .81$, $p < .001$], and for an interaction between them [$F(9, 162) = .3.97$, $MSe = .46$, $p < .001$].

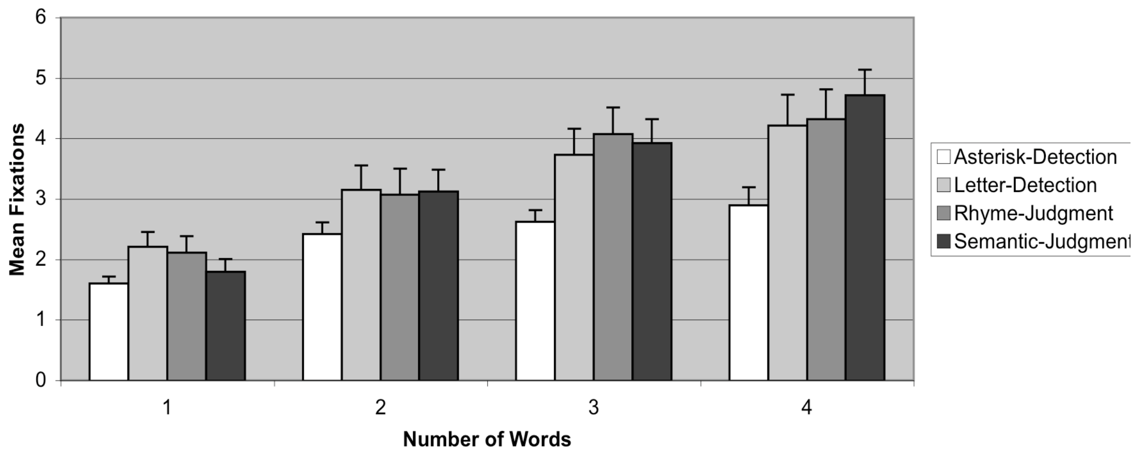

Figure 3. Mean number of fixations as a function of task type and number of words

Pair-wise comparisons of each task were performed to determine the effect of number of words displayed on total number of fixations. Pair-wise comparisons of the asterisk-detection task were reliable between all conditions [all $t(18)$s > -6.59, ps < .001], except for the comparisons between the 2-word versus 3-word, 2-word versus 4-word, and 3-word versus 4-word conditions [all $t(18)$s < -1.65, ps > .038]. All pair-wise comparisons of the letter-detection task were reliable between all conditions [all $t(18)$s > -5.90, ps < .008], except for the comparison between the 3-word versus 4-word condition [$t(18) = -1.93$, $p = .070$]. Similarly, pair-wise comparisons of the rhyme-judgment task were reliable for all conditions [all $t(18)$s > -7.58, ps < .002], except for the comparison between the 3-word versus 4-word conditions [$t(18) = -1.24$, $p = .230$]. And finally, pair-wise comparisons of the semantic-judgment task were reliable for all conditions [all $t(18)$s > -9.52, ps < .001].

3.2.2 Mean Fixation Duration

Figure 4 shows the mean fixation duration of correctly identified targets, as a function of both task type and number of words displayed. A repeated-measure ANOVA using task type and number of words as within-subject factors indicated significant main effects for task type and number of words [$F(3, 54) = 12.06$, $MSe = 3539.37$, $p < .001$; and $F(3, 54) = 109.93$, $MSe = 3466.50$, $p < .001$], and a significant interaction between them [$F(9, 162) = 7.873$, $MSe = 1508.52$, $p < .001$].

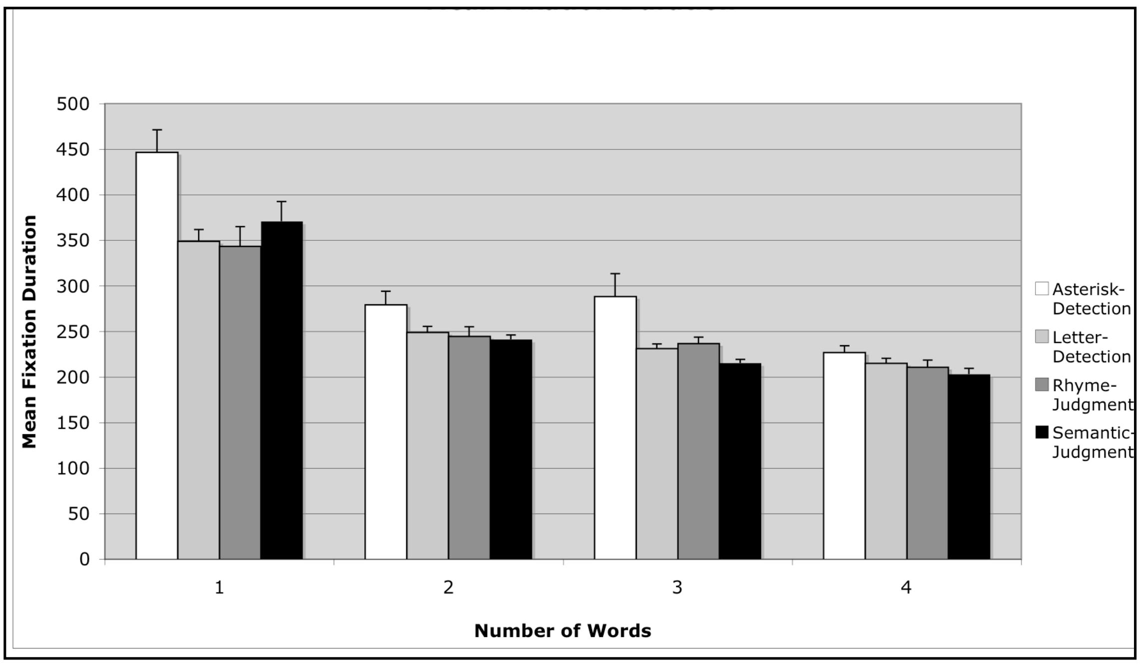

Figure 4. Mean fixation duration as a function of both task type and number of words

Pair-wise comparisons of each task were performed to determine the effect of number of words displayed on mean fixation duration. Pair-wise comparisons of the asterisk-detection task were reliable between the 1-word versus 2-word [$t(25) = 8.37$, $p < .001$], 1-word versus 3-word, 1-word versus 4-word, and 2-word versus 4-word conditions [all $t(23)$s > 4.44, ps < .001], but not for the comparisons between the 2-word versus 3-word or 3-word versus 4-word conditions [$t(23)$s > 2.01, ps $> .050$]. Pair-wise comparisons of the letter-detection task were reliable between the 1-word versus 2-word, 1-word versus 3-word, 1-word versus 4-word [$t(22) = 8.41$, $p < .001$], and 2-word versus 4-word conditions [all $t(21)$s > 6.02, ps $< .001$], but not for the comparisons between the 2-word versus 3-word or 3-word versus 4-word conditions [$t(23)$s > 2.12, ps $> .010$]. Pair-wise comparisons of the rhyme-judgment task were reliable for the 1-word versus 2-word [$t(23) = 5.11$, $p < .001$], 1-word versus 3-word, and 1-word versus 4-word conditions [$t(22)$s > 5.52, ps $< .001$], but not for the comparisons between the 2-word versus 3-word [$t(23) = .604$, $p = .552$], 2-word versus 4-word, or 3-word versus 4-word conditions [$t(22)$s < 2.86, ps $> .009$]. And finally, pair-wise comparisons of the semantic-judgment task were reliable for all conditions [all $t(23)$s > 5.23, ps $< .001$], except for the comparison between the 3-word versus 4-word condition [$t(23) = 2.04$, $p = .053$].

3.3 WORKING MEMORY RESULTS

3.3.1 Scoring

The working memory score was calculated using a partial-credit unit (PCU) scoring system. The partial-credit scoring is used to give credit to correct items recalled regardless if they are recalled in the correct serial order and unit scoring is used to give credit to all items equally as a proportion of correctly recalled items per item, regardless of size. Thus, PCU calculation is the mean proportion of elements within an item that are recalled correctly. A recent study (Conway et al., 2005) suggested empirical results favor partial-credit scoring and unit-weighted scoring is preferred because it follows established and sound procedures from psychometrics.

3.3.2 Results

Twenty-five subjects working memory data was used to calculate regression between PCU score and slopes of response times in each of the four attention-allocation tasks. In other words, the correlation was calculated between working memory span and the slope of the response times (as a function of the number of words that were displayed) in each of the four tasks. A moderate correlation was found in the asterisk-detection, letter-detection, and the rhyme-judgment tasks (*r*s = -0.24, -0.34, and -0.24, respectively) between PCU scores and reaction time slopes, see figures 5-7. A weak correlation was found in the semantic-judgment task between PCU scores and reaction time slopes ($r = -0.05$), see figure 8. Despite these trends, however, *t*-tests indicated that the *rs* were not reliable: asterisk-detection: $t(23) = -1.19$, $p = .248$; letter-detection: $t(23) = -1.73$, $p = .096$; rhyme-judgment: $t(23) = -1.19$, $p = .248$; and semantic-judgment: $t(23) = -.24$, $p = .812$.

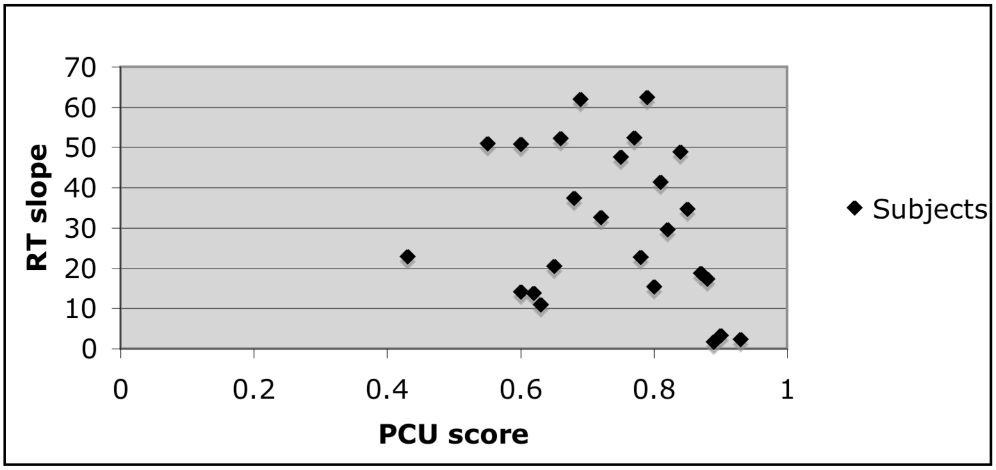

Figure 5. PCU score and reaction-time function slopes of the asterisk-detection task

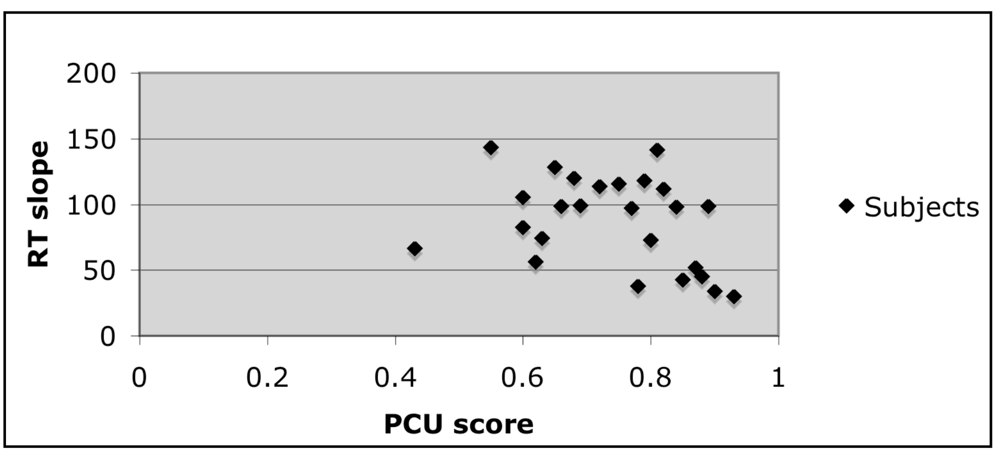

Figure 6. PCU score and reaction-time function slopes of the letter-detection task

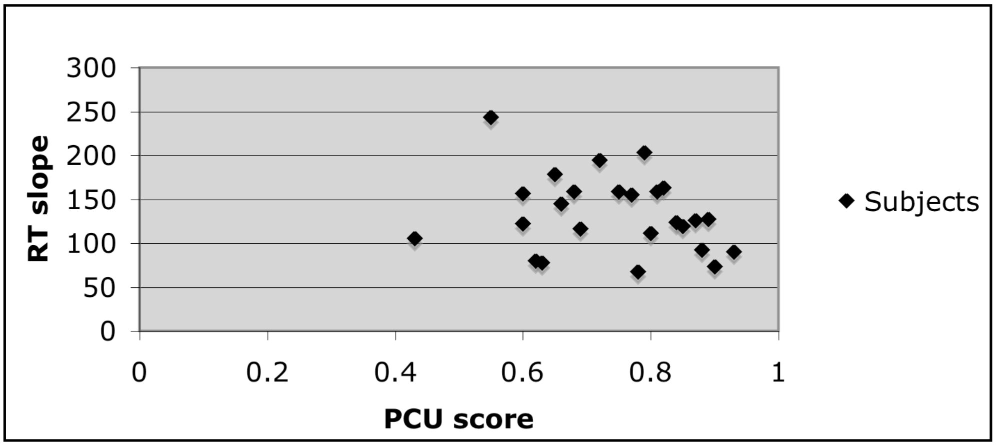

Figure 7. PCU score and reaction-time function slopes of the rhyme-judgment task

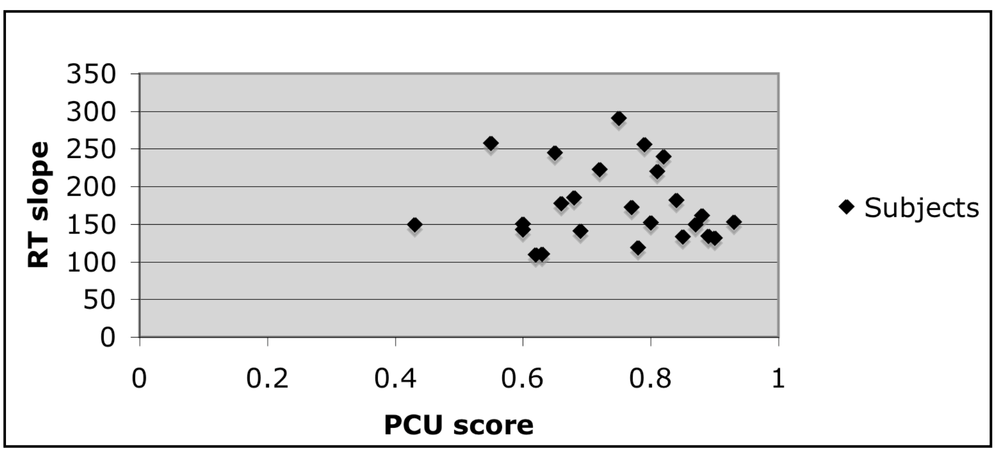

Figure 8. PCU score and reaction-time function slopes of the semantic-judgment task

4.0 DISCUSSION

4.1 STUDY REPLICATION

The results of this study generally replicate Reichle's et al. (2008) initial findings, albeit with some subtle differences. Reaction times did, in fact, increase as a function of number of words displayed and the complexity of the processing task and this supports the serial processing of attention model and Reichle's et al. (2008) initial findings. Reaction times, again, were shown to be faster during the "shallow" asterisk-detection task, indicating more of a tendency for parallel processing in that task. This phenomenon has previously been attributed to the asterisk seeming to "pop" out of the display (Reichle et al., 2008) and may allow for attention to be allocated to more than one word at a time. This study somewhat differs in that the results indicate that the reaction times of the "deeper" processing tasks were slower, with the slowest reaction times present in the semantic-judgment task, followed by rhyme-judgment task, and finally letter-detection task. Also, reaction times as a function of number of words displayed revealed to be longest for the 3-word condition, followed by the 4-word, 2-word, and finally the 1-word conditions. This may be caused by a practicing effect as the number of words increase through each task. All pair-wise comparisons of the conditions in reaction times revealed reliable differences, except the comparison between the 2-word versus 4-word, in each task.

Overall, participants were accurate in their responses in all four tasks. Results indicate that response accuracy decreased as a function of processing task and number of words displayed, but was not lower than 92% in any of the conditions. Further analysis using pair-wise comparisons of accuracy revealed reliable differences in the asterisk detection, letter-detection, and rhyme-judgment tasks between the 1-4 word conditions. In the semantic judgment task no differences were found between each of the 1-4 word conditions, indicating that participants performed just as accurately in the 1-word condition as in the 2-word, 3-word, and 4-word conditions.

The mean number of fixations was consistent with Reichle's et al. (2008) original study. The number of fixations was lowest for the asterisk-detection task when compared to the more "deeper" processing task and increased as a function of number of words displayed. As the mean number of fixations relates to task type, results showed as the level of the processing task increased, the number of fixations simultaneously increased. The interaction between task-type and number of words indicates that participants are likely to make less fixations during the "shallow" asterisk detection task because they are able to allocate attention in a manner reflecting parallel processing. During the "deeper" processing task, participants must use serial attention allocation as this is reflected in more fixations as the number of words increase. Pair-wise comparisons reveal reliable differences in each task between most conditions. Mean fixation durations were longest for the 1-word trials, replicating Reichle's et al. (2008) previous study and is attributed the more words appearing on the screen, the less amount of time a participant is able to fixate on a target word. Fixation durations were the longest for the asterisk-detection task, than the other three tasks and this might due to the participants having more time to fixate on the asterisk-detection task because it is a relatively "shallow" task. Pair-wise comparisons of mean fixation duration revealed reliable differences between most conditions.

4.2 WORKING MEMORY CAPACITY AND ATTENTION ALLOCATION

The relationship between working memory capacity and how attention is allocated during reading can be summarized as follows. In this study, high working memory capacity is attributed to a high PCU score and low working memory to a low PCU score. The slope of the reaction time functions was calculated by dividing the range (i.e. the difference between the 4-word and 1-word conditions) of the reaction times of the conditions involving 1-4 words by the total number of possible words per task [i.e., slope = (maximum RT – minimum RT) / 4]. In this study, the asterisk-detection function had the shallowest slope and the slopes of the reaction time functions of the other 3 tasks become increasingly steeper (i.e. letter-detection > rhyme-judgment > semantic-judgment).

Our initial hypotheses were not supported by results found in the current study. Our results indicate that performance in the working memory task (i.e. PCU score) is moderately (see Cohen, 1988 for guidelines on effect sizes) negatively correlated with the slope calculated from the reaction time of the asterisk-detection, letter-detection, and rhyme-judgment. This indicates that high-span participants are more likely to display faster reaction times during the asterisk-detection, letter-detection, and rhyme-judgment tasks, than low-span individuals. Shallow reaction-time function slopes are indicative of faster reaction times, supporting parallel processing of words during reading, whereas steeper reaction-time function slopes are indicative of slower reaction times, supporting serial processing during reading. But, at the point when the high-span participant encounters the semantic-judgment task, they perform just as well as the low-span individuals. All participants would have to approach the semantic-judgment task in a serial manner by processing one word at a time, searching for the meanings of those words, and then responding to whatever the trial indicated as a target. This suggests that working memory span has little affect on how subjects process words for meaning because, in performing this task, attention has to be allocated in a serial manner by all subjects.

In contrast, during the asterisk- and letter-detection tasks, as high-span individuals search through the words to locate a specific symbols or orthographic features (i.e. "*" or "q"), they may be able to allocate attention to more than one word at a time because it is unnecessary to process the whole word when searching for these simple features. It is less clear why performance in the rhyme-judgment task is more similar in regards to the PCU score and the asterisk- and letter-detection tasks, but not with the semantic-judgment task. It may be that high-span participants are searching words for features at the end of words that commonly rhyme with blue (i.e. "–ue", "-oo", -"ew", etc.). Like the asterisk- and letter-detection tasks, high-span individuals can allocate attention to processing more than one word at a time because, again, they are looking for specific orthographic features and do not have to process the whole word. This is less true for the low-span participants, who display steeper slopes and lower PCU scores. Low-span participants take longer, on average, to locate a target in each of the all four processing tasks, which suggests attention being allocated in a serial manner, one word at a time.

4.3 CAVEATS

It must be noted that there are a few important caveats to this study. Studies involving individual differences must have a large enough sample to indicate reliable differences between groups. Twenty-five participants did not allow for our individual differences results to become significant. Given our observed effect sizes, we would expect to observe statistically significant effects if our sample size were doubled. Also, seven participant's eye-movement data was not recorded for some trials and this might be one reason why our data deviated from Reichle's et al. (2008) study. And finally, our participants were limited to undergraduate students of at least a moderate, if not high intelligence. Selecting participants from the general population may have allowed our working memory results have stronger relationships to the reaction times of the reading tasks (i.e., our study suffers from the "restriction of range" problem in that our subjects had a fairly limited range of PCU scores). Despite these limitations, however, our results our highly suggestive in that performance in the task that a priori was considered to be the most similar to word identification during reading (i.e., the semantic judgment task) was not affected by working memory span—presumably because this task requires the strictly serially allocation of attention and consequently is less modulated by between-participant differences in working memory span. Working memory span did seem to affect the participants' capacity to detect asterisks, however—presumably because this task can be performed by allocating attention (at least to some degree) in parallel to multiple words, and this capacity is modulated by the working memory resources that are available to a given participant. Of course, we will need to collect additional data to confirm this prediction.

4.4 FUTURE DIRECTIONS

Future research in this area may look at the importance of the "yes" and "no" responses of correct trials on reaction time, accuracy, and eye-movement data to be used in computational modeling. Additional studies in this area could expand the subject pool and control for certain demographic information, such as age, gender, education level, and/or number of books read. Individual differences in these areas may similarly be related to how attention is allocated during reading.

REFERENCES

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchinson, K. A., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behavior Research Methods, 39,* 445-459.

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin, 130*(4), 553-557.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33A,* 497-505.

Carpenter, P. A., Just, M. A., & Reichle E. D. (2000). Working memory and executive function: evidence from neuroimaging. *Current Opinion in Neurobiology, 10*(2), 195-199.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (second ed.). Lawrence Erlbaum Associates.

Francis, W., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton Mifflin.

Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122-149.

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 users guide. *Psychonomic Bulletin & Review, 12*(5), 769-786.

Daneman, M., & Carpenter, P. A. (1980). Individual Differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19,* 450-466.

Engbert, R., Nuthmann, Richter, R, & Kliegl. (2005). SWIFT: a dynamic model of saccade generation during reading. *Psychological Review, 112,* 777-813.

Kane, M. J., Conway, A. R. A., Bleckley, M. K., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130*(2), 169-183.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mid during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General, 135,* 12-35.

Kofler, M. J., Rapport, M.D., Bolden, J., Sarver, D.E., & Raiker, J.S. (2010). ADHD and working memory: The impact of central executive deficits and exceeding storage/rehearsal capacity on observed inattentive behavior. *Journal of Abnormal Child Psychology: An official publication of the International Society for Research in Child and Adolescent Psychopathology, 38*(2), 149-161.

Loboda, T. (2009) Psychological tests: Reading span test. Retrieved November 1, 2009, from http://www.sis.pitt.edu/~tloboda/res/research/psych-tests/rspan/index.html.

Reichle, E. D., Carpenter, P. A., & Just, M. A. (2000). The neural bases of strategy and skill in sentence-picture verification. *Cognitive Psychology, 40*(4), 261-295.

Reichle, E. D., Liversedge, S. P., Pollatsek, A., & Rayner, K. (2009). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences, 13*(3), 115-119.

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher-level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, *16*, 1-21.

Reichle, E. D., Vanyukov, P. M., Laurent, P. A., & Warren, T. (2008). Serial or parallel? Using depth-of-processing to examine attention allocation during reading. *Vision Research, 48*, 1831-1836.

Thornton, T. L. & Gilden, D. L. (2007). Parallel and serial processes in visual search. *Psychological Review, 114*(1), 71-103.