

Using Working Memory Theory to Investigate the Construct Validity of Multiple-Choice Reading Comprehension Tests Such as the SAT

Meredyth Daneman and Brenda Hannon
University of Toronto

When taking multiple-choice tests of reading comprehension such as the Scholastic Assessment Test (SAT), test takers use a range of strategies that vary in the extent to which they emphasize reading the questions versus reading the passages. Researchers have challenged the construct validity of these tests because test takers can achieve better-than-chance performance even if they do not read the passages at all. By using an individual-differences approach that compares the relative power of working memory span to predict SAT performance for different test-taking strategies, the authors show that the SAT appears to be tapping reading comprehension processes as long as test takers engage in at least some reading of the passages themselves.

In this article, we show how working memory theory can be used to address questions of interest to educational researchers. In particular, we use the working memory approach as a way to investigate the construct validity of the reading comprehension portion of the revised Scholastic Assessment Test (SAT). There has been a long and persistent history of attacks on the validity of multiple-choice tests of reading comprehension such as the SAT (see Anderson, Hiebert, Scott, & Wilkinson, 1985; Cohen, 1984; Drum, Calfee, & Cook, 1981; Farr, Pritchard, & Smitten, 1990; Katz, Blackburn, & Lautenschlager, 1991; Katz, Lautenschlager, Blackburn, & Harris, 1990; Royer, 1990). One of the most serious criticisms is that test takers do not or need not read and comprehend the passages on which the test questions are based. Indeed, Katz et al. (1990) demonstrated that test takers were able to perform better than chance on as many as 72% of the multiple-choice items of the reading section of the SAT when they were not given access to the passages. On the basis of findings such as this, critics have argued that multiple-choice reading tests in general, and the reading portion of the SAT in particular, may largely be measuring factors unrelated to reading comprehension. This is a serious allegation given the widespread practice of using SAT scores in the screening and placement of college applicants in the United States. We first briefly describe how working memory theory has been applied to understanding educationally relevant

tasks. Then we describe how we have applied working memory theory to investigating the construct validity of the reading portion of the SAT.

Working Memory as a Predictor of Complex Cognition

There is already considerable evidence that working memory theory can be applied to understanding performance on educationally relevant tasks. This is not surprising given that the construct of working memory was proposed as an alternative to short-term memory largely because of concerns about the ecological relevance of the short-term memory construct (Baddeley & Hitch, 1974; Reitman, 1970). Prototypical models of short-term memory (see, e.g., Atkinson & Shiffrin, 1968; Posner & Rossman, 1965) assumed that short-term memory plays a crucial role in the performance of ecologically relevant cognitive tasks such as language comprehension, mental arithmetic, and reasoning, tasks that for their solution require that individuals temporarily store information and then operate on it. However, as soon as efforts were made to test this intuitively appealing notion, it became evident that the existing models of short-term memory were inadequate. Traditional measures of short-term memory such as word span and digit span did not predict performance on complex cognitive tasks. So the theory of short-term memory as a passive storage buffer was replaced by the theory of working memory as a dynamic system with processing and storage capabilities (see, e.g., Baddeley, 1986; Baddeley & Hitch, 1974; Just & Carpenter, 1992). Word span and digit span, measures that tap only passive short-term storage capacity or number of “slots,” were replaced by reading span (Daneman & Carpenter, 1980) and operation span (Turner & Engle, 1989), measures that tap the combined processing and temporary storage capacity of working memory during the performance of a complex cognitive task.

There is now a substantial body of evidence that measures of the combined processing and storage capacity of working memory have lived up to their promise of doing a better job at predicting performance on complex cognitive tasks than did the traditional storage measures they replaced. Measures of working memory capacity have been shown to predict performance on cognitive

Meredyth Daneman and Brenda Hannon, Department of Psychology, University of Toronto, Mississauga, Ontario, Canada.

This research was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada.

We thank Gary Buck, Anne Connell, Irene Kostin, Tom Van Essen, and the rest of the SAT team at the Educational Testing Service for allowing us to use their SAT materials and for supplying us with their item difficulty norms and details about official test administration; we thank Michelle Day for help with task development and Candice Moore and Jayme Pickett for help with data collection.

Correspondence concerning this article should be addressed to Meredyth Daneman, Department of Psychology, University of Toronto, Mississauga, Ontario L5L 1C6, Canada. Electronic mail may be sent to daneman@psych.utoronto.ca.

activities as diverse as reading, listening, writing, solving verbal and spatial reasoning problems, and programming a computer (see, e.g., Baddeley, Logie, Nimmo-Smith, & Brereton, 1985; Benton, Kraft, Glover, & Plake, 1984; Daneman & Carpenter, 1980, 1983; Daneman & Green, 1986; Gathercole & Baddeley, 1993; Jurden, 1995; Kyllonen & Christal, 1990; Kyllonen & Stephens, 1990; Masson & Miller, 1983; Shah & Miyake, 1996; Shute, 1991; for reviews, see Daneman & Merikle, 1996; Engle, 1996). These findings suggest that working memory plays a role in the performance of a range of educationally relevant complex cognitive tasks and that individuals with large working memory capacities do better on these tasks than do individuals with smaller working memory capacities. Indeed, the working memory approach has been deemed so successful that a measure of the combined processing and storage capacity of working memory has been included in the latest edition of the Wechsler Adult Intelligence Scale (WAIS-III; Wechsler, 1997), and digit span has been demoted to an optional subtest (see *WAIS-III Technical Manual*, 1997). Of particular relevance to the present study are the findings that working memory capacity is a good predictor of performance on tests of reading comprehension ability and tests of verbal reasoning ability.

Consider for the moment the finding that measures of the combined processing and storage capacity of working memory are good predictors of performance on tests of reading comprehension ability (Daneman & Merikle, 1996). Daneman and Merikle conducted a meta-analysis of the literature investigating the association between working memory capacity and different kinds of language comprehension tasks. The meta-analysis included data from 6,179 participants in 77 independent studies. On the predictor task side, the meta-analysis included studies that used measures of the combined processing and storage capacity of working memory, such as reading span and operation span, as well as studies that used the traditional span tests that tap predominantly storage resources, such as word span and digit span. In a typical process plus storage measure, individuals may be required to read and judge the truth value of sets of unrelated sentences (e.g., "Mammals are vertebrates that give birth to live young," "March is the first month in the year that has thirty-one days," "You can trace the languages English and German back to the same roots") and then to recall the final words of each sentence in the set (e.g., *young, days, roots*; see Daneman & Carpenter, 1980). Or they may be required to verify the stated solutions to simple arithmetic problems (e.g., " $(2 \times 3) - 2 = 4$ tree," " $(6/3) + 2 = 8$ drink," " $(4 \times 2) - 5 = 3$ chain") and then to recall the stated solutions (e.g., 4, 8, 3; see Turner & Engle, 1989) or the accompanying words (e.g., *tree, drink, chain*; see Turner & Engle, 1989). In a traditional storage span measure, individuals simply have to store and retrieve a string of random words (e.g., *cup, shoe, ball*) or digits (e.g., 8, 6, 1). On the criterion task side, the meta-analysis included studies that assessed language skill with global or standardized tests of comprehension and vocabulary knowledge and with specific tests of integration. The most common global or standardized tests were the verbal component of the SAT and the Nelson-Denny Reading Test. Specific tests of integration included tests that assessed people's ability to compute the referent for a pronoun, to make inferences, to monitor and revise inconsistencies, to acquire new word meanings from contextual cues, to abstract the main theme, and so on (see Daneman & Merikle, 1996).

The results of Daneman and Merikle's (1996) meta-analysis showed that verbal process plus storage measures such as reading span were the best predictors of comprehension, correlating .41 and .52 with global and specific tests of comprehension, respectively.¹ However, math process plus storage measures such as operation span were also significant predictors, correlating .30 and .48 with global and specific tests of comprehension, respectively,² a finding that suggests that it is an individual's efficiency at executing a variety of symbolic processes, and not simply sentence comprehension processes, that is related to comprehension ability. In addition, both the verbal and the math process plus storage measures were better predictors of comprehension than their simple word span and digit span counterparts,³ a finding that suggests that it is the combined processing and temporary storage capacity of working memory, and not simply the temporary storage capacity, that is important for comprehension.

All in all, the correlational evidence suggests that the capacity to simultaneously process and store symbolic⁴ information in working memory is an important component of success at comprehension. Moreover, this working memory capacity seems to be a sensitive predictor of individual differences in performance on global tests of comprehension that use a multiple-choice format (such as the SAT) and on specific tests of comprehension that assess comprehension by means of a variety of other non-multiple-choice formats (such as having test takers generate answers to specific questions or summarize the main theme). According to the theory, working memory span is a good predictor of comprehension because individuals who have less capacity to simultaneously process and store verbal information in working memory are at a disadvantage when it comes to integrating successively encountered ideas in a text as they have less capacity to keep the earlier read relevant information still active in working memory.

Construct Validity and the Reading Portion of the SAT

Does the reading comprehension portion of the SAT measure what it was designed to measure, namely, passage comprehension? Critics such as Katz et al. (1990) have argued no because test takers could be using a strategy that bears little if any relation to what the test was designed to measure; this is the strategy of selecting answers to questions without even reading (let alone comprehending) the passages on which the questions were based.

¹ The 95% confidence intervals (CIs) were .38 to .44 for global tests of comprehension and .49 to .55 for specific tests of comprehension.

² The CIs were .25 to .35 for global tests of comprehension and .43 to .53 for specific tests of comprehension.

³ Word span correlated .28 (CI = .23-.33) and .40 (CI = .34-.46) with the global and specific tests of comprehension, respectively, and digit span correlated .14 (CI = .10-.18) and .30 (CI = .25-.35) with the global and specific tests of comprehension, respectively.

⁴ There is some evidence to suggest that when it comes to predicting comprehension skills, the predictive power of the process plus storage measures of working memory are limited to measures that tap symbolic processes (e.g., words, sentences, digits). Daneman and Merikle's (1996) meta-analysis investigated only the predictive power of verbal and math process plus storage measures. A number of recent studies have shown that spatial process plus storage measures do not predict comprehension ability (see, e.g., Shah & Miyake, 1996).

Indeed, Katz et al.'s claim was based on their finding that college students could perform at better-than-chance levels on the multiple-choice reading questions on pre-1994 versions of the SAT even though they were not given access to the passages (see also Powers & Leung, 1995, for a similar result with the revised version of the SAT). Given that the Educational Testing Service (ETS) developed the SAT as a measure of the ability to read and comprehend short English prose passages (Donlon, 1984), Katz et al. reasoned that test takers "must answer a group of multiple-choice items based on what is stated or implied in a passage, and not choose answers on the basis of personal opinion or prior knowledge, or because a particular choice is known to be true" (p. 122). Findings of better-than-chance performance on a passageless task led Katz et al. to conclude that the test "substantially measures factors unrelated to reading comprehension" (p. 122).

A problem with Katz et al.'s (1990) conclusions is that even though students can perform at levels exceeding chance in the absence of the passages, it is unlikely that they resort to such a strategy when taking the test with the passages available. As Powers and Leung (1995) pointed out, the no-passage strategy "is neither efficient nor effective" (p. 125). Powers and Leung found that students used nearly as much time answering the questions without reading the passages as is allowed on the exam for reading the SAT passages and answering the questions. Although students performed better than chance in the absence of the passages, they did not perform substantially better than chance and certainly not at levels that are competitive for college admission. Thus, it is unlikely that students would adopt a strategy of ignoring the available passages when they are taking a test "that will dramatically affect their futures" (Freedle & Kostin, 1994, p. 109). Nevertheless, students may use a range of strategies that vary in the extent to which they emphasize reading the questions versus reading the passages. A goal of the present study was to approach the construct validity issue by investigating strategies that students might plausibly be using in the conventional SAT setting.

A study by Farr et al. (1990) provides some evidence concerning the strategies used by college students when completing a multiple-choice reading comprehension test. A group of 26 college students completed a portion of the Iowa Silent Reading Test (Farr, 1972), a typical standardized reading comprehension test consisting of passages followed by multiple-choice questions. Half the students were asked to explain to the researcher exactly what they were thinking and doing as they read the test passages and answered the multiple-choice questions. The other half completed the test without interruption and recounted their strategies afterwards. From the concurrent and retrospective protocols provided by the students, Farr et al. identified four overall strategies (or general approaches) used to complete the reading comprehension test.

The first and second strategies identified by Farr et al. (1990) were *passage-first strategies* in that students consulted the passage before turning to the questions. The two strategies differed in the extent of the initial interaction with the passage. Strategy 1 was to read the entire passage before proceeding to the questions; then each question was read and was followed by a search of the passage for the correct answer. Strategy 2 was to read the passage only partially before proceeding to the questions; then each question was read and was followed by a search of the passage for the correct answer. The third and fourth strategies were *question-first strategies* in that students consulted the questions before reading

the passage. The two strategies differed in the extent of the initial interaction with the questions. Strategy 3 was to read through all the questions and then read the entire passage; this was followed by a rereading of each question, followed by a search of the passage for the correct answer. Strategy 4 was to read the first question and then search the passage for the correct answer, read the second question and search the passage for the correct answer, and so on. Although over 50% of students initially adopted the strategy of reading the entire passage before considering the questions (Strategy 1), 30% of them shifted strategies during the course of taking the test. For those who shifted strategies, the tendency was to move away from reading the entire passage before attempting to answer the questions. The students moved either to a partial reading of the passage before turning to the questions (Strategy 2) or to reading the questions before reading any of the passage (Strategies 3 and 4). Thus, it appears that as the test progressed, students tended to adopt a strategy of getting to the questions as quickly as possible. However, an important finding by Farr et al. is that students always used the questions to direct their search of the passage for the relevant information to answer the questions. In no case did students attempt to answer a question from information in the question alone, that is, without consulting at least some portion of the passage itself.

Although Farr et al.'s (1990) descriptive study was useful for identifying four strategies used by test takers on multiple-choice reading comprehension tests, two important questions remain unanswered. Does the particular strategy adopted by a test taker affect test success? Does the particular strategy adopted by a test taker affect the construct validity of the test? With respect to the first question, Farr et al. did not analyze comprehension success as a function of the strategy selected. They simply reported mean performance as a function of whether students gave concurrent verbal protocols (72.6%) versus retrospective protocols (72.4%), as well as the range of performance for the whole group (50–93.8%). So one knows that individuals varied widely in their performance on the test but not whether individual differences in performance were at all related to strategy selection. In the present study, we investigated performance success as a function of test-taking strategy by manipulating strategy use directly. In Experiment 1, an exploratory study, students were instructed to use four different strategies when completing different parts of the reading comprehension portions of two forms of the SAT. In Experiment 2, they were instructed to use two different strategies. By manipulating test-taking strategy, we could investigate whether it contributed to test success independent of any individual differences in test takers' abilities.

With respect to the construct validity question, the four strategies identified by Farr et al. (1990) all involved consulting or searching local parts of the passage to find information to answer a specific question. However, there were differences among the four strategies in how much global reading of a passage took place before the test taker initiated question-directed search of the passage. The amount of global passage reading ranged from reading the entire passage (Strategies 1 and 3) to reading part of the passage (Strategy 2) to not reading the passage in a global, non-question-directed sense at all (Strategy 4). To the extent that a reading comprehension test like the SAT is designed to assess passage comprehension, the construct validity of the test could vary as a function of how much passage reading the test taker

actually engages in. In other words, the test could have greater construct validity when test takers use Strategies 1 and 3 than when they use Strategies 2 and 4 because the former two strategies engage more reading comprehension processes than do the latter two, and the test could be least valid when test takers use Strategy 4 because this is the strategy that engages the least amount of passage reading.

In the present study, we applied working memory theory to investigating the extent to which the four different strategies tapped reading comprehension processes. The logic was as follows. Measures of the combined processing and storage capacity of working memory such as reading span (Daneman & Carpenter, 1980) and, to a lesser extent, operation span (Turner & Engle, 1989) are good predictors of performance on a range of reading comprehension tasks regardless of whether the tasks use a multiple-choice testing format or some other testing format (see Daneman & Merikle, 1996). Thus, investigating the strength of the correlations between working memory span measures and SAT performance as a function of test-taking strategy provides a way to determine whether the various strategies differ in the extent to which they draw on working-memory-demanding comprehension processes, and, consequently, the extent to which the SAT is a valid reflection of passage comprehension ability. If the working memory measures are more highly correlated with SAT performance when test takers use some strategies than when they use other strategies, such a finding would suggest that the validity of the SAT may vary as a function of strategy use. If the correlations are high and similar across the board, this would suggest that the validity of the SAT is not affected by the overall strategy adopted by the test taker. Consequently, in both Experiments 1 and 2, our test takers were administered the reading span and operation span tests of working memory capacity in conjunction with the reading comprehension portions of the SAT.⁵ Although the primary goal of the present study was to investigate strategies that students might plausibly be using in the conventional SAT setting, we included in Experiment 3 a condition in which students were required to answer the SAT questions without being allowed to consult the passages at all. The rationale for including a no-passage condition is explained later.

Can Epistemic Knowledge Compensate for Limited Working Memory Capacity?

In Experiment 2, we included an additional measure of individual differences, namely, a measure of the test takers' epistemic knowledge (see Schommer, 1990; Schommer, Crouse, & Rhodes, 1992). Although working memory span measures are excellent predictors of reading comprehension performance, they are by no means perfect predictors. A recent study by Rukavina and Daneman (1996) showed that epistemic knowledge (or knowledge about knowledge and learning) is also related to reading comprehension success. Rukavina and Daneman measured epistemic knowledge with two subtests of Schommer's (1990) epistemological questionnaire, one that taps individuals' beliefs about the degree to which knowledge is simple or complex and a second that taps individuals' beliefs about the degree to which integration is important to learning. Students used a 5-point scale to respond to items such as "Most words have one clear meaning. You will get confused if you try integrating new ideas in a text with knowledge

you already know about a topic." On the basis of their responses, they were classified as having mature or naive epistemic beliefs. Rukavina and Daneman found that students with mature epistemic beliefs did better than students with naive beliefs on a task that required them to acquire knowledge about scientific theories from written texts. There were two additional findings of interest in the study. First, working memory capacity (as measured by reading span) and epistemic knowledge were uncorrelated in the sample and hence appeared to be making independent contributions to comprehending the scientific texts. Second, there was some suggestion in the data that students appeared to be able to compensate for a deficit in one resource as long as they had sufficient of the other. In other words, students who had small working memory spans but mature epistemic beliefs appeared to be compensating for a deficit in working memory span because they performed almost as well as students who had both large working memory spans and mature epistemic beliefs on some of the learning outcome measures. Similarly, students who had naive epistemic beliefs but large working memory spans performed almost as well as students who had both large working memory spans and mature epistemic beliefs. In other words, it was only the students with both small working memory spans and naive epistemic beliefs who were particularly penalized. A goal of Experiment 2 was to see whether we could replicate and extend Rukavina and Daneman's findings by showing that mature epistemic knowledge is related to comprehension success and can compensate for a small working memory capacity in the context of completing the reading comprehension portion of the revised SAT.

Experiment 1

Farr et al. (1990) identified four general strategies that test takers spontaneously use when completing a multiple-choice reading comprehension test. However, Farr et al. did not analyze test success as a function of the strategy used, nor did they investigate the extent to which test-taking strategy affected the validity of the test as a measure of passage comprehension. Both issues were investigated in Experiment 1.

To investigate these issues, we manipulated test-taking strategy directly. Students were required to use four different strategies when taking the reading comprehension portion of the revised SAT. Two forms of the revised SAT were used. Each form had four passages and 40 questions, and each student was instructed to change strategies after every two passages. This meant that for two

⁵ One of the reviewers of an earlier version of this article suggested that our approach seemed circular because historically the correlations between working memory span tasks and the reading comprehension portion of the SAT were used to validate the working memory span tasks, and here we use the same correlations to validate the reading comprehension portion of the SAT. The reviewer is certainly correct in pointing out that working memory measures were originally validated against standardized comprehension tests like the SAT. However, the evidence that has accumulated over the past 20 years overwhelmingly supports a strong relation between working memory span measures and comprehension ability (Daneman & Merikle, 1996). It now seems appropriate to take advantage of this well-established relationship and to start using working memory span tests as useful tools for investigating various applied issues that go beyond zero-order correlations, such as reading and test-taking strategies.

of the eight passages, students were required to read the entire passage before turning to the questions (passage-first; entire passage). For another two passages, they were required to read half the passage and then to turn to the questions (passage-first; half passage). For another two passages, students were required to read all the questions first, followed by the entire passage, and then to turn to the questions (question-first; entire passage). For two passages, students were required to read the first question, then consult the passage to answer it, read the second question, then consult the passage to answer it, and so on (question-first; none of the passage). By manipulating test-taking strategy in this manner, we could examine whether it influenced SAT test performance independent of an individual test taker's abilities. In addition, we examined the degree to which each strategy might be engaging reading comprehension processes by investigating how well performance with that strategy correlated with two working memory span measures, reading span and operation span. If a particular strategy engaged reading comprehension processes fully, we would predict a good correlation between SAT performance and reading span and a slightly weaker but still significant correlation between SAT performance and operation span.⁶ If a particular strategy did not engage reading comprehension processes sufficiently, we would predict poor correlations between SAT performance and both measures of working memory capacity.

Method

Participants. The participants were 48 University of Toronto students. All students were fluent in English and were tested individually in two sessions. During the first session, participants were administered two tests of the combined processing and storage capacity of working memory, reading span (Daneman & Carpenter, 1980) and operation span (Turner & Engle, 1989), as well as another standardized multiple-choice test of reading comprehension, the Nelson-Denny (Brown, Bennett, & Hanna, 1981). During the second session, participants were administered two forms of the critical reading section of the revised SAT. Although Canadian university students have had some experience taking standardized tests of reading comprehension, most have not taken any form of the SAT because it is not a requirement for admission to Canadian universities.

Reading span test. As one of our measures of working memory span, we used a variant of Daneman and Carpenter's (1980) reading span test, which was designed to measure the combined processing and storage capacity of working memory during reading. In the version we used (see also Hannon & Daneman, 2001), participants were required to read aloud sets of unrelated sentences (e.g., "Torrential rains swept over the tiny deserted island," "His mouth was twisted into an inhuman smile," "The umbrella grabbed its bat and stepped up to the plate") and make judgments about the sensibility of each sentence (e.g., respond yes after reading the first sentence, yes after reading the second sentence, and no after reading the third). Then, at the end of the set, they were required to recall the final word of each sentence in the set (e.g., *island*, *smile*, *plate*).

Sentences 8 to 12 words in length, each ending with a different word, were presented one at a time on a computer screen. After responding yes or no to indicate whether or not the sentence made sense, the participant pressed a key, and the next sentence appeared. The procedure was repeated until a blank screen indicated that the trial was over, and the participant had to recall the last word of each of the sentences in the set. Participants were allowed to recall the words in any order but were encouraged not to recall the last word in the set first. Sentences were arranged in five sets of 2, 3, 4, 5, and 6 sentences each. Participants were presented with increasingly longer sets until all 100 sentences had been presented. Reading span was the total number of *sentence final words* out of 100 that the participant

could recall. The reading span task was presented on a 486 IBM-PC compatible system, using the Micro Experimental Laboratory software developed by Schneider (1988).

Operation span test. As our second measure of working memory span, we used Turner and Engle's (1989) operation span test, which was designed to measure the combined processing and storage capacity of working memory during the performance of simple mathematical computations. In this task, participants were given sets of equations and accompanying words (e.g., " $(4 \times 2) - 1 = 7$ girl," " $(1 \times 6) - 5 = 1$ paper," " $(9/3) + 3 = 2$ truth"). For each equation-word pair, participants were required to read aloud the math equation and the word, verify whether or not the statement was true (e.g., respond true for the first equation, true for the second, and false for the third). Then, at the end of the set, they were required to recall all the words in the set (e.g., *girl*, *paper*, *truth*).

The equations each required participants to perform two operations; the first operation involved multiplying or dividing two integers that were in parentheses, and the second operation involved adding an integer to, or subtracting an integer from, the product or quotient from the first operation. The stated solution was correct approximately half the time. When the equation was false, the stated solution was always incorrect by 4. All of the integers used in the equations were from 1 to 9 (see also Turner & Engle, 1989). The words consisted of 100 high-frequency nouns four to six letters in length.

The equation-word pairs were presented one at a time on the computer screen, using the same hardware and software as was used for the reading span task. After responding true or false to indicate whether or not the stated solution to the equation was correct, the participant pressed a key, and the next equation-word pair appeared. The procedure was repeated until a blank screen indicated that the trial was over, and then the participant had to recall the words. Equation-word pairs were arranged in five sets of 2, 3, 4, 5, and 6 equation-word pairs each. Participants were presented with increasingly longer sets until all 100 equation-word pairs had been presented. Operation span was the total number of words recalled, the maximum being 100.

Nelson-Denny test of reading comprehension. Participants were administered the Nelson-Denny test of reading comprehension (Brown et al., 1981). The Nelson-Denny consists of eight prose passages and 36 multiple-choice questions. Participants were given 20 min to read the passages and answer the questions.

SAT strategy task. Students were instructed to use four different strategies to complete the SAT test of reading comprehension. There were eight passages in total, and students changed strategy after every two passages.

⁶ It would be difficult, if not impossible, to predict what the magnitude of a good working memory/SAT correlation should be on the basis of previous findings. Although numerous studies have reported the correlations between working memory span measures and verbal SAT scores (see Daneman & Merikle, 1996), they do not provide an adequate baseline of comparison for several reasons. As far as we know, these correlations have all been based on composite verbal SAT scores taken from university records or from students' self-reports rather than from data collected in the study itself and on just the reading comprehension portion of the SAT as was done here. Consequently, the reported SAT scores have included performance on nonreading sections as well as reading sections. In the case of pre-1994 SAT scores (which would pertain to most of the published studies to date), the reading comprehension questions constituted a smaller proportion of the overall verbal SAT score than in the revised SAT used in the present study. Since 1994, the verbal reasoning portion of the SAT has placed a greater emphasis on critical reading, and vocabulary is measured in context (as part of the reading comprehension portion) rather than with discrete antonym items. Indeed, the proportion of reading questions in the revised SAT has increased by almost 60% (see Powers & Leung, 1995).

The materials consisted of two forms of the critical reading portion of the revised SAT. Each form had four passages and 40 multiple-choice questions. The following is a brief description of the topics of each passage and the number of questions associated with each one: (a) an excerpt from the autobiography of a Hispanic American writer (12 questions); (b) an excerpt from an essay on women and writing by a contemporary American poet (5 questions); (c) an account of the emergence of genetics, the science of inherited traits (10 questions); (d) two passages reflecting two views of the values and integrity of journalism (13 questions); (e) an excerpt about 19th century Bohemia (12 questions); (f) an excerpt about Chinese American women writers (7 questions); (g) a discussion of various ways that living creatures have been classified over the years (8 questions); and (h) two passages presenting two different perspectives of the United States prairie (13 questions).

The multiple-choice questions all had five alternatives and tested information stated or implied in the accompanying passages. Some examples of question stems are (a) "The passage serves primarily to . . ."; (b) "In creating the impression of the prairie for the reader, the author of Passage 1 makes use of . . ."; (c) "The author implies that 'a good reader' is one who . . ."; and (d) "In line 20, 'tamed' most nearly means . . ."

All participants saw the eight passages in the same fixed order and changed strategies after every two passages. However, the order in which participants used the four different strategies was completely counterbalanced. There were 24 possible orderings in which the four strategies could be assigned to the four pairs of passages, and there were 48 participants; consequently, 2 participants were assigned to each of the 24 possible orders; in each case, one of the participants had scored above average on the reading span test administered in the first session, and the other participant had scored below average on the reading span test.

Participants were told that we were interested in investigating whether different reading and test-taking strategies affect performance on a test of reading comprehension. They were told that they would be informed which strategy to use before beginning each passage, and they were requested to use that strategy and not deviate from it even if they thought it was a bad one. Then, before commencing a particular strategy, participants were given specific instructions for that strategy. The experimenter went through the instructions for the strategy twice, using a sample passage and questions as props. The experimenter also carefully monitored the test-taking session to ensure that each participant followed the instructions at all times.

The four strategies were based on the ones identified by Farr et al. (1990). For Strategy 1 (passage-first; entire passage), participants read the entire passage first. Only after completing the passage were participants given the multiple-choice questions; they read the first question and then searched the passage for the answer and recorded their answer on the answer sheet provided. Participants then read the second question, consulted the passage to answer it, and so on. For Strategy 2 (passage-first; half passage), participants were given a printed sheet containing only half the passage and were told to read it to get a sense of what the passage was about. After reading it, they were given a copy of the entire passage and the multiple-choice questions, and they were required to go immediately to the first question, read it, and then consult the intact passage for the answer, record their answer, and repeat the procedure with the rest of the questions. For Strategy 3 (questions-first; entire passage), participants read the complete set of questions for the passage first. Following that, they were given the entire passage, which they were required to read. Then they reread the first question and searched the passage for the answer, recorded their answer, reread the second question, searched the passage, and so on. For Strategy 4 (question-first; none of the passage), participants read the first question and then searched the passage for the answer, recorded their answer, read the second question, searched the passage for the answer, and so on. In other words, they were allowed to use the passage only to search for specific answers and were not allowed to read the passage from start to finish as in Strategies 1 and 3.⁷

Regardless of the strategy being used, participants were allowed as much time as needed to complete the questions for any given passage. However, we did attempt to get some indication of the time our participants spent on the task by having the experimenter time how long each participant took to complete the questions for a given passage. For Strategies 1 and 2, the experimenter started the timer as soon as the participant began to read the passage (Strategy 1) or half passage (Strategy 2). For Strategy 3, the experimenter started the timer as soon as the participant began to read the set of questions. For Strategy 4, the experimenter started the timer as soon as the participant began to read the first question. For all four strategies, the experimenter stopped the timer as soon as the participant recorded his or her answer for the final question. Mean time per question was calculated by dividing the total time needed to complete the questions for a given passage by the number of questions for that passage.

Results and Discussion

SAT performance. As Table 1 shows, overall performance on the SAT was 67% ($SD = 14.82$). Although we conducted the test under somewhat different testing conditions than those used by ETS (e.g., we imposed test-taking strategies on our test takers; however, we did not impose time constraints or subtract marks for wrong answers), there were a number of indications that our overall data resemble those found under official testing conditions. First, performance on the two forms of the test were highly correlated (.81), a finding that is consistent with the alternate-form reliability figures reported by ETS. Second, mean performance on individual passages in our data correlated very highly (.96) with the item difficulty statistics (equated delta) provided to us by ETS for these two forms. Third, even though we did not impose time constraints on our test takers, the actual time they spent per question rarely exceeded the average time available per question during official testing. The average time available per question (including time to read the passage) is estimated to be 65 s for the pre-1994 versions (see Katz et al., 1990) and 69.23 s for the revised version used here.⁸ The average time per question spent by our test takers was 60.60 s ($SD = 18.40$), which was in fact less than the average time per question available under official testing. The average times per question for the four strategies were 61.71 s ($SD = 16.43$) for Strategy 1, 56.93 s ($SD = 19.29$) for Strategy 2, 73.12 s ($SD = 17.01$) for Strategy 3, and 52.68 s ($SD = 14.46$) for Strategy 4. Thus, only in the case of Strategy 3, the strategy that required the most reading (advance reading of questions and passage), did the average time spent per question by our test takers slightly exceed the average time per question allowed in the revised version of the SAT used in this study (73.12 s vs. 69.23 s). Note also that the average time our test takers spent per question increased systematically with the amount of reading required by a

⁷ There were 2 participants who did not conform to the instructions for Strategy 4; that is, they tried to read the passage in detail before attempting to answer specific questions. These participants were excluded from the study.

⁸ It is a little tricky to calculate the average amount of time available per question under official SAT testing conditions because a timed session usually includes sentence completion and analogy items as well as reading comprehension items. Our calculation of 69.23 s per reading item for the revised SAT is based on one timed session that includes reading comprehension items only. In this session, test takers are given 15 min to read a passage and answer 13 questions about it.

Table 1
Means and Standard Deviations for All Tasks
Used in Experiment 1

Test and task	M	SD	Range
Working memory span			
Reading span (maximum = 100)	58.31	13.28	35–95
Operation span (maximum = 100)	74.52	13.12	47–100
Nelson-Denny reading comprehension (maximum = 36)	23.90	6.23	9–36
SAT strategy task (% correct)			
Strategy 1: passage-first; entire passage	69.72	17.34	28.57–100.00
Strategy 2: passage-first; half passage	64.25	17.59	26.32–91.30
Strategy 3: questions-first; entire passage	69.20	20.14	19.05–100.00
Strategy 4: questions-first; none of the passage	65.77	18.52	23.81–100.00
Overall SAT performance	67.23	14.82	30.08–92.40

Note. SAT = Scholastic Assessment Test.

strategy (Strategy 3 > Strategy 1 > Strategy 2 > Strategy 4). A one-way analysis of variance (ANOVA) on the test-taking times for the four strategies was highly significant, $F(3, 141) = 39.47$, $MSE = 94.61$, $p < .0001$, and all pairwise t tests were significant, all $ps < .03$. These findings lend support to our contention that participants were following strategy instructions. Finally, mean performance on the SAT was quite high (67%), suggesting that our participants took the task seriously and were motivated to perform well even though their performance in our study did not have the kinds of consequences for their future that performance on an official SAT would have.

SAT performance as a function of test-taking strategy. Did the particular strategy used by a test taker influence performance level on the SAT? As Table 1 shows, the means seem to reflect a small advantage for the two strategies that involve reading the entire passage before answering the questions over the strategies that involve reading half the passage or none of the passage before answering the questions. However, none of the differences among the four strategies was statistically significant (all $ps > .05$), although the differences between Strategies 1 and 2 (70% vs. 64%) and between Strategies 3 and 2 (69% vs. 64%) approached significance (both $ps < .06$).

There are two possible interpretations of these null/marginal findings. One possibility is that there are no differences in SAT performance as a function of the four overall strategies; as long as test takers can search the passages as much as they need to in answering the specific questions, this is all the reading and comprehension required to produce a mean performance level of at least 64%. This performance level is of course well above chance level (20%). It is also well above the 36% to 38% performance levels that have been reported for test takers who were required to complete the questions without any access to the passages (see Katz et al., 1990; Powers & Leung, 1995).⁹ However, the second possible interpretation of these marginal results is that there are strategy effects on SAT performance but that we did not find significant and reliable effects because of insufficient power in our design. Our test takers used a particular strategy for only two passages (the equivalent of half of one SAT test).¹⁰ Two passages

may not have provided sufficient data for us to pick up reliable effects. Consequently, in Experiment 2, we included only two of the four strategies in our design. This meant that each student had the opportunity to use the same strategy for an entire SAT form (i.e., for four passages and 40 questions).

Working memory span measures. Table 1 shows mean performance on the two working memory measures, reading span and operation span. Our versions of reading span and operation span were highly correlated with one another, $r(46) = .76$, $p < .001$, and both were correlated with performance on the Nelson-Denny test of reading comprehension, with reading span being the slightly better predictor of comprehension of the two (reading span correlated .65 with the Nelson-Denny whereas operation span correlated .53 with the Nelson-Denny). These findings are all consistent with previous findings in the literature (see Daneman & Merikle, 1996; Turner & Engle, 1989) and show that our versions of reading span and operation span are good predictors of performance on a typical standardized test of reading comprehension ability when test-taking strategy is uncontrolled.

Working memory span and SAT performance as a function of test-taking strategy. Does the construct validity of the SAT vary as a function of test-taking strategy? Table 2 shows how reading span and operation span correlated with performance on the SAT as a function of test-taking strategy. As the table shows, there were some differences in the predictive power of reading span as a function of test-taking strategy. Reading span correlated best with Strategy 3 performance (.53) and poorest with Strategy 4 performance (.30).¹¹ Strategy 3 was the strategy involving the greatest amount of reading in advance because test takers read the full set of questions and the entire passage before beginning to answer the specific questions. Strategy 4 was the strategy that involved the least amount of reading in advance because test takers read neither the full set of questions nor any of the passage before beginning to find answers to specific questions. Although these results are suggestive, one should be cautious about drawing any firm conclusion from them in light of our earlier discussed concerns about the potential problems associated with having our test takers use a given strategy for only two SAT passages.

Indeed, another indication that performance based on only two SAT passages may lack reliability is the finding of relatively low correlations between SAT performance and performance on our other multiple-choice test of reading comprehension, the Nelson-Denny. As Table 2 shows, the correlations were rather low, even for SAT performance involving a lot of passage reading (for Strategy 1, SAT performance correlated .55 with the Nelson-

⁹ In one experiment, Katz et al. (1990) reported means of 46% to 47% for no-passage conditions; however, they acknowledged that their samples were small and selective, with mean verbal SAT scores higher and standard deviations smaller than are typically found in more representative groups.

¹⁰ In some cases, this meant that test takers used a particular strategy on as few as 17 of the 40 questions of a particular form of the test; in other cases, test takers used the same strategy on as many as 23 of the 40 questions. Also, not only are there different numbers of questions for the first two passages than for the second two passages of one form of the revised SAT but also there are different numbers of each kind of question (e.g., testing main idea, vocabulary in context, specific details) in the two halves of the form as well.

¹¹ $r(45) = 1.88$, $p > .05$ (two-tailed), $p < .05$ (one-tailed).

Table 2
Correlations Between Working Memory Span Measures and SAT Performance in Experiment 1

Measure	SAT performance			
	Strategy 1	Strategy 2	Strategy 3	Strategy 4
Reading span	.40**	.37**	.53***	.30*
Operation span	.39**	.39**	.42**	.42**
Nelson-Denny	.55***	.57***	.50***	.45**

Note. SAT = Scholastic Assessment Test.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Denny; for Strategy 3, it correlated .50). Given the high similarity between the SAT and Nelson-Denny in structure and format, we would have expected higher correlations had the SAT scores based on two passages been more reliable.

All in all, the Experiment 1 findings suggest that it would be preferable to increase the number of passages assigned to a given test-taking strategy. Consequently, in Experiment 2, we included only two test-taking strategies, Strategy 3, which involves the greatest amount of advance reading, and Strategy 4, which involves the least.

Experiment 2

In Experiment 2, we examined whether test-taking strategy influenced test success and the construct validity of the SAT, but this time we included only two of the four strategies identified by Farr et al. (1990). One was the strategy in which test takers read all the questions followed by the entire passage before beginning to answer specific questions (Strategy 3), and the other was the strategy of beginning to answer specific questions immediately without having read through the set of questions or any of the passage beforehand (Strategy 4). By including only two strategies, we were able to have students use each strategy to complete an entire SAT test (four passages and 40 questions).

An additional goal of Experiment 2 was to investigate whether students could compensate for a small working memory capacity if they had good metacognitive awareness about the complexity of knowledge and the importance of integration to learning. Rukavina and Daneman (1996) found some evidence that students with mature epistemic beliefs about knowledge and integration were able to compensate for a limited working memory capacity in a task that required them to integrate and acquire knowledge about two competing scientific theories from written text, and so they performed almost as well on some learning outcome measures as did students with large working memory spans. A possible explanation is that having mature awareness about knowledge and integration led these students to engage in a number of strategies to facilitate knowledge acquisition and integration, strategies such as backtracking and rereading relevant information in the science learning task. This kind of strategic reading and rereading would have had the effect of compensating for their limited capacities to hold information in temporary storage during reading. We hypothesized that a similar compensatory process might be involved in completing the SAT test of reading comprehension and that the compensation would be particularly evident in the situation that allowed for the most strategic use of advance knowledge about the

questions and the text to guide the process of reading and searching for correct answers. In other words, we hypothesized that students with large working memory capacities would be at an advantage on the reading comprehension task in comparison to students with small working memory capacities. However, if students had small working memory capacities and mature epistemic beliefs, they might perform almost as well as the students who had large working memory capacities in the advance reading situation (Strategy 3) because they would have the metacognitive awareness to use the advance reading situation to their advantage.

To investigate these issues, we included a measure of epistemic knowledge (Schommer, 1990), as well as our two measures of working memory capacity, reading span and operation span. The measure of epistemic knowledge was identical to that used by Rukavina and Daneman (1996). It consisted of two subtests of Schommer's (1990) epistemological questionnaire, one that taps people's beliefs about the degree to which knowledge is simple or complex and a second that taps people's beliefs about the degree to which integration is important to learning.

Method

Participants. The participants were 48 University of Toronto students who had not participated in Experiment 1. All students were fluent speakers of English and were tested individually in two sessions. During the first session, participants were administered the reading span test, the operation span test, and the Nelson-Denny. During the second session, they were administered the same two forms of the critical reading section of the revised SAT used in Experiment 1, followed by the epistemic knowledge measure. Below, we describe only those measures and procedural features that differed from Experiment 1.

Epistemic knowledge measure. Using a 5-point scale (1 = *strongly disagree*; 5 = *strongly agree*), students responded to 12 items selected from two subsets of Schommer's (1990) epistemological questionnaire. These two subsets, referred to as "seek single answers" and "avoid integration," are measures of students' beliefs about knowledge, whether they believe knowledge is simple or complex and whether they believe knowledge consists of facts or integrated ideas. Two sample items from the battery are (a) "The best thing about science courses is that most problems have only one solution" and (b) "I try my best to combine information across chapters or even across classes."¹² Students' scores were out of a maximum of 60—the higher the score, the more naive the student's epistemic knowledge.

SAT strategy task. Participants completed both forms of the SAT. Half the participants were instructed to use Strategy 3 (advance reading of questions and passage) for Passages 1 to 4, followed by Strategy 4 (no advance reading of questions or passage) for Passages 5 to 8. The other half were instructed to use Strategy 4 for Passages 1 to 4 and Strategy 3 for Passages 5 to 8.

¹² Notice that Item (a) has a positive valence, so that higher ratings correspond to naive epistemic knowledge. By contrast, Item (b) has a negative valence, so that higher ratings correspond to sophisticated or mature epistemic knowledge. Of the 12 test items, 9 had a positive valence, and 3 had a negative valence. These two subsets of Schommer's (1990) epistemological questionnaire seek single answers and avoid integration; each loaded highly on the same factor identified as "Knowledge is discrete and unambiguous" (Simple Knowledge), .47 and .42, respectively (Schommer, 1990; Schommer et al., 1992).

Results and Discussion

SAT performance as a function of test-taking strategy. Table 3 shows that there was an effect of test-taking strategy on test success. As the table shows, students performed better on the SAT test when they engaged in advance reading of the questions and passage ($M = 71\%$) than when they engaged in no advance reading ($M = 63\%$), $t(47) = 4.68, p < .0001$. This highly significant finding of a strategy effect suggests that the null effects in Experiment 1 may indeed have been a result of too few passages per strategy. Thus, it would seem that test takers do better on the SAT test if they do advance preparation than if they are more question driven and turn immediately to search for answers to specific questions without having read the complete passage and set of questions beforehand. However, does the question-driven strategy entail so little reading comprehension that it interferes with the construct validity of the test? To answer this question, we examined whether engaging in this strategy reduced the working memory span/SAT correlation.

Working memory span and SAT performance as a function of test-taking strategy. Our two working memory span measures were again highly correlated with one another, $r(46) = .76, p < .001$, and both were correlated with performance on the Nelson-Denny test of reading comprehension, with reading span being the slightly better predictor of comprehension of the two (reading span correlated .58 with the Nelson-Denny, whereas operation span correlated .52). As in Experiment 1, these findings show that our versions of reading span and operation span are good predictors of performance on a typical standardized test of reading comprehension ability when test-taking strategy is uncontrolled.

Table 4 shows how our two working memory span measures correlated with SAT performance. As Table 4 shows, there was absolutely no evidence that engaging in a question-driven strategy reduced the working memory span/SAT correlation. In fact, the correlations showed the typical pattern, namely, significant correlations for both reading span and operation span, with reading span the better predictor than operation span. If anything, both working memory measures were slightly better predictors of performance on the question-driven strategy than on the strategy that involved advance reading of the questions and passage. For example, the correlation between reading span and SAT performance with advance reading was an impressive .60, but the correlation between

Table 3
Means and Standard Deviations for All Tasks
Used in Experiment 2

Test and task	M	SD	Range
Working memory span			
Reading span (maximum = 100)	59.33	12.16	39–96
Operation span (maximum = 100)	77.94	13.70	43–99
Epistemic knowledge (maximum = 60)	33.46	6.36	18–47
Nelson-Denny reading comprehension (maximum = 36)	23.60	6.47	11–33
SAT strategy task (% correct)			
Strategy 3: advance reading	70.57	14.83	35.00–95.00
Strategy 4: no advance reading	63.02	17.12	25.00–100.00
Overall SAT performance	66.80	15.08	30.00–96.25

Note. SAT = Scholastic Assessment Test.

Table 4
Correlations Between Individual Differences Measures and SAT Performance in Experiment 2

Measure	SAT performance	
	Strategy 3 (advance reading)	Strategy 4 (no advance reading)
Reading span	.60***	.64***
Operation span	.46***	.53***
Epistemic knowledge	-.53***	-.42**
Nelson-Denny	.74***	.76***

Note. SAT = Scholastic Assessment Test.

** $p < .01$. *** $p < .001$.

reading span and SAT performance with no advance reading was an equally impressive and slightly higher .64 (see Table 4). A similar pattern was evident also in the Nelson-Denny/SAT correlations: The Nelson-Denny test of reading comprehension was as highly correlated with SAT performance when there was no advance reading (.76) as when there was (.74). All in all, there is no evidence to suggest that the SAT test was tapping something substantially different when test takers did not do any advance reading of the passages but simply directed their reading to answering specific questions. We take these findings to indicate that sufficient reading comprehension processes were engaged even in this condition, and so the construct validity of the SAT test as a test of passage comprehension was not compromised.

Of course, there is an alternative explanation for the pattern of correlations seen in Experiment 2. It could be something other than skill at reading comprehension processes that is accounting for the high correlations throughout. We know, for example, that reading span and operation span are good predictors of a host of verbal and nonverbal cognitive processes besides reading comprehension (see Daneman & Merikle, 1996; Engle, 1996). So an alternative explanation for the results is that some other skill or skills tapped by all these measures is contributing to the high correlations. If this were the case, then we could not conclude that we have established that the SAT is measuring passage comprehension with and without advance reading. All that we could conclude is that the people who are good at one task tend to be good at other tasks as well. We argue against this interpretation of the Experiment 2 data on the basis of the findings of Experiment 3, which investigated the pattern of correlations when students were not allowed to consult the SAT passages at all.

Can mature epistemic beliefs compensate for a limited working memory capacity? So far, we have considered how a student's working memory capacity is related to reading comprehension independent of that student's epistemic beliefs about knowledge and integration. An additional goal of Experiment 2 was to investigate whether mature epistemic beliefs about the complexity of knowledge and the importance of integration to learning also contribute to reading comprehension success and whether they can compensate for a limited working memory. Table 3 provides the descriptive statistics for our measure of epistemic knowledge, and Table 4 shows how it correlated with SAT performance.

As Table 4 shows, epistemic knowledge was a significant predictor of performance on both SAT tasks, correlating $-.53$ when

there was advance reading and $-.42$ when there was no advance reading. These correlations show that students with mature epistemic knowledge (those with lower scores on the measure) performed better on the SAT reading comprehension test than did students with naive epistemic beliefs, a finding that is consistent with Rukavina and Daneman's (1996) findings for a science reading task. In Rukavina and Daneman's study, epistemic knowledge and working memory span were uncorrelated; in our study, epistemic knowledge correlated moderately with working memory span ($-.36$ with reading span and $-.17$ with operation span).

There was also some evidence to suggest that mature epistemic beliefs could compensate somewhat for a small working memory capacity in the situation that allowed students to make strategic use of advance knowledge about the questions and the text. This compensation is illustrated in Figure 1, which plots SAT performance for four different groups on the basis of whether participants scored above or below the group mean for reading span and above or below the group mean for epistemic knowledge, respectively: students with large working memory spans and mature epistemic beliefs, large working memory spans and naive epistemic beliefs, small working memory spans and mature epistemic beliefs, and small working memory spans and naive epistemic

beliefs. As Figure 1(a) shows, in the case of the advance reading task, students with both small working memory spans and naive epistemic beliefs performed particularly poorly; they were correct on only 57% of the questions, which means that they answered 22% fewer questions correctly than their counterparts who had large working memory spans and naive epistemic beliefs and 25% fewer correctly than students who had both large spans and mature epistemic beliefs. On the other hand, students who also had small working memory spans but had mature epistemic beliefs were correct on 73% of the questions, which means that they answered only 9% fewer than their counterparts who had both large working memory spans and mature epistemic beliefs, a finding that suggests that their mature epistemic beliefs enabled them to compensate to some degree for their below-average working memory spans. Put another way, test takers with large working memory spans performed well on the advance reading SAT task regardless of whether they had mature epistemic beliefs or naive beliefs (82% vs. 79%). On the other hand, test takers with small working memory spans did much more poorly if they also had naive epistemic beliefs (57%) but were able to perform quite well if they had mature epistemic beliefs (73%). An ANOVA confirmed this description of the results insofar as there was a statistically significant Reading Span \times Epistemic Knowledge interaction, $F(1, 44) = 4.49$, $MSE = 97.53$, $p < .04$, in addition to the significant effects of reading span, $F(1, 44) = 27.37$, $MSE = 97.53$, $p < .001$, and epistemic knowledge, $F(1, 44) = 9.41$, $MSE = 97.53$, $p < .004$. This finding suggests that students who had mature beliefs about knowledge and integration may have recognized the need to initiate strategies such as backtracking and rereading when they encountered comprehension difficulties or information that might be relevant to the questions they had previewed. These monitoring and remedial strategies would have had the effect of compensating to some degree for their limited capacities to hold information in working memory during reading.

Figure 1(b) also suggests that there was less evidence that mature epistemic beliefs could compensate for a small working memory capacity when no advance reading of questions and passages was allowed. Notice that again it was the students with both small working memory spans and naive epistemic beliefs who performed most poorly; they were correct on only 51% of the questions, which means that they answered 19% fewer questions correctly than did their counterparts with large working memory spans and naive epistemic beliefs. However, having mature epistemic beliefs did not compensate much because students who had mature epistemic beliefs and small working memory spans answered a good 13% fewer questions correctly than did their counterparts with large working memory spans and mature epistemic beliefs. An ANOVA confirmed this description of the results insofar as there was no hint of a Reading Span \times Epistemic Knowledge interaction, $F < 1$. There was a significant effect of reading span, $F(1, 44) = 12.58$, $MSE = 203.26$, $p < .01$, but only a marginal effect of epistemic knowledge based on this mean-split, $F(1, 44) = 3.28$, $MSE = 203.26$, $p < .08$. As predicted, epistemic awareness is less useful when the situation does not allow for the strategic use of advance knowledge about the questions and text to guide the process of reading and searching for correct answers.

Table 5 provides the results of regression analyses on both the advance reading and the no advance reading SAT data as a function of reading span and epistemic knowledge. As the table

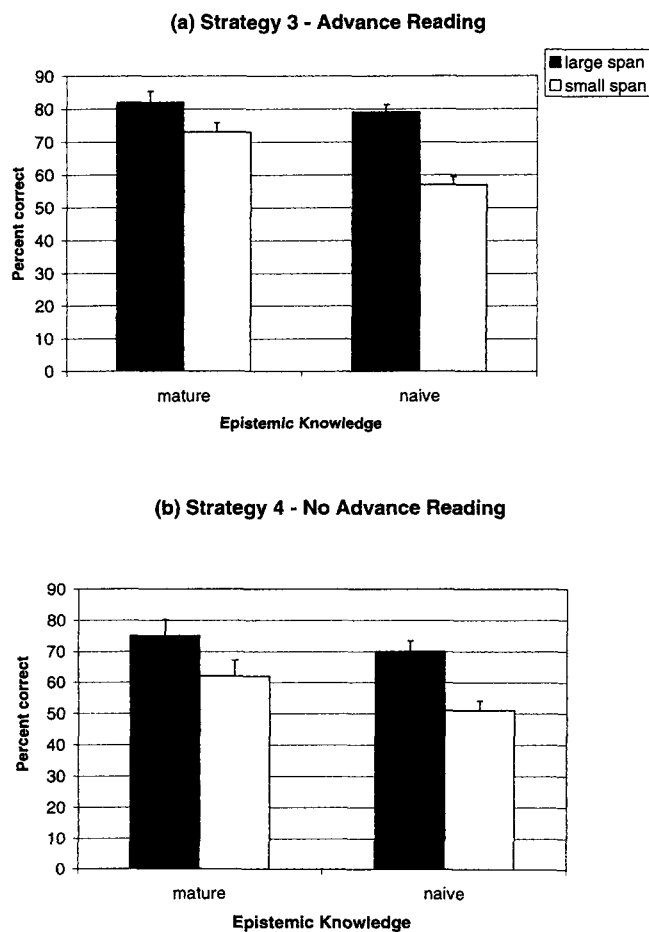


Figure 1. Mean Scholastic Assessment Test performance as a function of the test taker's strategy, reading span, and epistemic knowledge in Experiment 2.

Table 5
Regression Analyses on SAT Performance in the Advance Reading and No Advance Reading Conditions of Experiment 2

Measure and condition	<i>R</i>	<i>R</i> ²	ΔR^2	<i>F</i>
Advance reading				
Reading span	.60	.36	.36	26.25**
Epistemic knowledge	.69	.48	.12	9.75**
Nelson-Denny	.81	.66	.18	23.51**
No advance reading				
Reading span	.64	.41	.41	32.37**
Epistemic knowledge	.67	.45	.04	3.27*
Nelson-Denny	.81	.65	.20	24.57**

Note. ΔR^2 = change in *R*²; SAT = Scholastic Assessment Test.
 * *p* < .05. ** *p* < .01.

shows, reading span accounted for 36% of the variance on SAT scores with advance reading, and epistemic knowledge contributed an additional 12% of unique variance. In contrast, reading span accounted for 41% of the variance on SAT scores with no advance reading, and epistemic knowledge contributed only an additional 4% of unique variance. The regression analyses in Table 5 also show the amount of variance accounted for by the Nelson-Denny test when it was entered into the regression equation after the effects of reading span and epistemic knowledge were statistically removed. Notice that for both the advance reading and no advance reading strategies, the Nelson-Denny test of reading comprehension accounted for a substantial proportion of residual variance (18% and 20%, respectively); this finding becomes important when we compare these results with those for the no-passage condition in Experiment 3.

Experiment 3

Although the primary goal of the present research was to investigate strategies that students might plausibly be using in the conventional SAT setting, our final experiment included a condition in which students were required to answer the SAT questions without consulting the passages at all (see also Katz et al., 1990; Powers & Leung, 1995). By comparing the relative power of reading span and the Nelson-Denny at predicting SAT performance when test takers do or do not see the passages, we were able to provide support for our view that the SAT assesses passage comprehension as long as the test taker is allowed to do some passage reading.

Our first prediction was that SAT performance would be lower when test takers have no access to the passages as compared with when they are allowed to consult the passages at the time of answering specific questions. On the basis of previous findings (e.g., Katz et al., 1990; Powers & Leung, 1995), we predicted that mean performance would be above chance (20%) even when test takers could not read or consult the passages on which the questions were based. However, if we were correct in concluding that test takers were still doing a significant amount of passage reading even though they were not allowed to do any advance passage reading in the question-driven strategy in Experiment 2, then we would expect performance without any passage reading to be significantly lower than 63%, which was the mean performance level for the no advance reading strategy in Experiment 2.

We also made several predictions concerning the relative power of reading span and the Nelson-Denny to predict SAT question-answering performance when passages were unavailable. Remember that in Experiment 2, reading span was a good predictor of SAT performance whether test takers engaged in advance reading of the passages or only question-driven reading of the passages; however, the Nelson-Denny test of reading comprehension was an even better predictor of SAT performance and still accounted for a substantial proportion of the variance in SAT performance after the effects of reading span were partialled out. For the no-passage reading condition in Experiment 3, we predicted that reading span would again be a good predictor of SAT performance; however, now it should be an even better predictor than the other reading comprehension measure, the Nelson-Denny. Let us explain the logic behind our predictions.

Why would we expect reading span to be a good predictor of performance in a condition that entails no passage reading when all along we have been interpreting the working memory span/SAT correlations as evidence that the SAT task is tapping passage reading comprehension processes? Of course, we would not be making this prediction if we believed, as Katz et al. (1990) and others do, that test takers achieve better-than-chance performance as a result of invoking strategies that have no relevance to the reading comprehension construct, strategies that depend on personal opinion, prior knowledge, or general testwiseness. If this were the case, then we would have no reason to expect reading span to predict SAT performance in the no-passage condition. However, we believe that these kinds of strategies play a minor role when test takers try hard to answer questions without the help of the passages on which they are based. Self-reports provided by students taking the SAT test without the passages (Powers & Leung, 1995) strongly suggest that the better-than-chance performance is a result of strategies that "entail legitimate verbal reasoning abilities" (Powers & Leung, 1995, p. 106), strategies such as "choosing answers on the basis of consistency with other questions and reconstructing the main theme of a missing passage from all the questions and answers in the set" (Powers & Leung, 1995, p. 105). These kinds of sophisticated verbal reasoning strategies are precisely the kinds of strategies that have been shown to draw on the processing and storage resources of working memory (Jurden, 1995; Kyllonen & Christal, 1990). Consequently, we would expect reading span to predict SAT performance in the no-passage condition as well. On the other hand, because specific reading comprehension processes would be minimally engaged in the no-passage condition (they would be engaged just to read the question stems and answer choices), we would expect the Nelson-Denny reading comprehension test to be much more weakly correlated with SAT performance in this condition than in the passage conditions of Experiment 2. In other words, we would expect reading span to be better than the Nelson-Denny at predicting SAT performance when no passages are available, whereas the reverse was true for the two strategies that made passages available.

Method

Participants. The participants were 28 University of Toronto students who had not participated in Experiments 1 and 2. All participants were fluent speakers of English. Participants were administered the reading span, epistemic knowledge, and Nelson-Denny measures. In addition, they completed one form of the SAT without seeing the accompanying passages.

SAT with no passages. Participants were told that the purpose of the study was to determine how well people can answer multiple-choice questions without reading the passages on which they were based. They were informed that it has been shown that people can do quite well even with the passage missing (see also Katz et al., 1990), and they were encouraged to try their best to determine what the correct answer should be.

To make the no-passage condition as comparable as possible to the advance reading and no advance reading conditions in Experiment 2, each participant completed 40 SAT questions, with half the participants completing the 40 questions from Form 1 (Passages 1 to 4) and the other half completing the 40 questions from Form 2 (Passages 5 to 8). Participants were given the questions in the order in which they appeared in the original SAT forms; however, they were given the questions for only one passage at a time. In other words, participants were given the 12 questions corresponding to Passage 1 first, and only when they had completed these were they given the 5 questions corresponding to Passage 2, and so on.

Results and Discussion

Table 6 provides the means and standard deviations for the tasks. As Table 6 shows, mean performance on the SAT with no passages was 35% ($SD = 12.06$), a level that is better than chance (20%), $t(27) = 6.70$, $p < .01$, and comparable to the performance levels found in previous studies (e.g., Katz et al., 1990; Powers & Leung, 1995). However, the 35% level of performance is also considerably less than the 63% level achieved by Experiment 2 test takers who were allowed to consult the passages only to answer specific questions, $t(76) = 19.99$, $p < .0001$, a finding that suggests that the passages were helpful in answering many questions and that test takers were doing a significant amount of passage reading even when they were not allowed to do any advance passage reading.

The correlational results are consistent with our position that test takers invoke qualitatively different strategies when the passages are unavailable than when they are available. Table 7 shows that reading span, epistemic knowledge, and Nelson-Denny performance all correlated significantly with SAT performance when passages were unavailable, just as they did in Experiment 2 when passages were available. However, the striking difference is in the degree to which the Nelson-Denny correlated with SAT performance when passages were unavailable as compared with when they were available. As Table 7 shows, the Nelson-Denny correlated only .38 with SAT performance in the no-passage condition, as compared with .74 and .76 in the two passage-available conditions in Experiment 2; this means that the Nelson-Denny was accounting for little more than 14% of the variance in no-passage SAT performance, a large drop from the 55+% variance it ac-

Table 6
Means and Standard Deviations for All Tasks
Used in Experiment 3

Test	<i>M</i>	<i>SD</i>	Range
Reading span (maximum = 100)	60.89	11.24	43–96
Epistemic knowledge (maximum = 60)	33.46	4.35	23–41
Nelson-Denny reading comprehension (maximum = 36)	22.71	6.29	11–33
SAT—no passages (% correct)	35.27	12.06	15.00–60.00

Note. SAT = Scholastic Assessment Test.

Table 7
Correlations for No-Passage Condition in Experiment 3

Measure	Epistemic knowledge	Nelson-Denny comprehension	SAT (no passage)
Reading span	-.16	.53**	.54**
Epistemic knowledge	—	-.31	-.40**
Nelson-Denny comprehension	—	—	.38*

Note. SAT = Scholastic Assessment Test.

* $p < .05$. ** $p < .01$.

counted for when passages were available. On the other hand, as we had predicted, reading span still correlated well with SAT performance in this no-passage condition (.54) and consequently was a much better predictor of SAT performance than was the Nelson-Denny.

This reversal of the relative predictive powers of reading span and the Nelson-Denny is highlighted in Table 8, which juxtaposes the results of a regression analysis on the no-passage SAT data with regression analyses on the SAT data for the two passage-available strategies in Experiment 2. In each regression analysis, only reading span and the Nelson-Denny were included as predictor variables, with reading span entered into the regression equation before the Nelson-Denny. As Table 8 shows, the Nelson-Denny accounted for a substantial 23% of the residual variance when test takers were required to read the entire passage in advance and a substantial 22% of the residual variance when test takers were not allowed to read the passage in advance but could still consult it to answer specific questions. However, the Nelson-Denny accounted for a mere 1% of residual variance when test takers could not consult the passages at all.¹³ We take these findings to reflect the fact that test takers invoke a qualitatively different strategy when the passages are not available than when they are available (see also Powers & Leung, 1995). When the passages are available, test takers make use of the passages by reading them and answering the multiple-choice questions on the basis of what is stated or implied in them. Working memory capacity plays a role in how well individuals can execute the reading comprehension processes, and so reading span, as a measure of working memory capacity, is a good predictor of performance on the SAT when passages are involved. However, the Nelson-Denny, a task that has great surface similarity to the SAT passage-available conditions, is an even better predictor of performance, capturing variance not accounted for by differences in working memory capacity. On the other hand, when passages are not available, test takers do not do much reading at all, and so the Nelson-Denny is no longer a good predictor. Test takers probably do invoke sophisticated verbal reasoning and inferential skills as they attempt to reconstruct the theme and contents of the passage from what the questions and answer choices reveal. Working memory capacity plays a role in how well individuals can engage

¹³ Reading span was the better predictor of SAT performance in the no-passage condition even when Nelson-Denny scores were entered into the regression equation first. Reading span accounted for 16% unique variance as compared with 14% accounted for by the Nelson-Denny.

Table 8
Regression Analyses on Three Different SAT Strategies With Reading Span and Nelson-Denny as Predictors

Strategy and predictor	<i>R</i>	<i>R</i> ²	ΔR^2	<i>F</i>
Experiment 2				
Advance reading				
Reading span	.60	.36	.36	26.25**
Nelson-Denny	.77	.59	.23	25.11**
No advance reading				
Reading span	.64	.41	.41	32.37**
Nelson-Denny	.80	.63	.22	27.17**
Experiment 3				
No passages				
Reading span	.54	.29	.29	10.73**
Nelson-Denny	.55	.30	.01	0.44

Note. ΔR^2 = change in *R*²; SAT = Scholastic Assessment Test.
 ** *p* < .01.

in complex verbal reasoning, and so reading span, as a measure of working memory capacity, is a good predictor of SAT performance in the no-passage condition. (Notice, it was not because reading span is a good predictor of everything; it correlated only $-.16$ with epistemic knowledge in this study; see Table 7). This pattern of results suggests that even if test takers were to adopt the extreme approach of entirely ignoring the passages when answering all or some of the SAT questions, the strategy would not compromise the construct validity of the test entirely. The no-passage strategy is qualitatively different from the passage-reading strategies; nevertheless, it implicates the kinds of sophisticated verbal reasoning skills that are important for and predictive of academic success.

Do the present findings reveal anything about what working memory span tasks like reading span really measure? Of course, the present study was not designed to investigate this issue. However, we think that the reversal of the predictive powers of reading span and the Nelson-Denny test (captured in Table 8) is more consistent with our theory of what reading span measures than it is with at least one rival theory. Specifically, we think that the reversal is more consistent with our view of reading span as a measure of the combined capacity individuals have for simultaneously processing and storing verbal information in working memory than it is with Kintsch's (1998) view of reading span as a measure of "the efficiency with which readers can comprehend sentences and hence store them in long-term memory" (p. 239). It is not that we disagree with Kintsch about our reading span measure being sensitive to individual differences in people's abilities to comprehend sentences and store them in long-term memory. Indeed, we now have some direct evidence that reading span is correlated with the ability to retrieve newly acquired text information from long-term memory (see Hannon & Daneman, 2001). It is just that we do not wish to take the dynamic working memory component out of reading span and simply equate reading span with comprehension skills or with long-term memory storage and retrieval as Kintsch preferred to do (see also Ericsson & Kintsch, 1995). If reading span were simply another measure of comprehension skills, then its predictive power should mirror that of the

direct measure of comprehension skills used in this study, namely the Nelson-Denny. However, we have just shown that this was not the case. The Nelson-Denny was the better predictor of SAT performance when passages were available (Experiments 1 and 2), whereas reading span was the better predictor of SAT performance when passages were not available (Experiment 3). We think the account we gave earlier is the most parsimonious account. As the more direct test of comprehension skills of the two, the Nelson-Denny was better than reading span at predicting SAT performance when passage comprehension processes were significantly engaged. However, when passages were not available, test takers did not do much reading at all, and so the simple test of comprehension skills was no longer the better (or even a particularly good) predictor of SAT performance. However, test takers still had to invoke working-memory-demanding verbal reasoning and inferential skills to perform above chance on the no-passage task, and so reading span as a measure of an individual's resources for simultaneously manipulating and storing verbal information in working memory was still a good predictor in the no-passage condition.

Finally, even if Kintsch's (1998) view of reading span as a measure of "the efficiency with which readers can comprehend sentences and hence store them in long-term memory" (p. 239) has some surface plausibility with respect to the reading span task, it has less plausibility with respect to the operation span task, given that the process component of the operation span task taps numerical computations rather than sentence comprehension processes. In the current study, reading span and operation span were highly correlated with one another, and both predicted SAT performance when passages were available (Experiments 1 and 2). These findings are consistent with our view that both tasks measure the combined resources an individual has for simultaneously processing and temporarily storing symbolic information in working memory. Reading span is the better predictor of SAT comprehension performance of the two span measures because it measures an individual's resources for executing language-specific processes, whereas operation span measures an individual's resources for executing predominantly numerical manipulations and computations (see Daneman & Merikle, 1996).

Summary and Conclusions

Educational Issues

Does the particular strategy adopted by a test taker affect test success on the reading comprehension portion of the revised SAT? Does the particular strategy adopted by a test taker affect the construct validity of the SAT? These were the two main questions addressed by our research. These questions were motivated by a recent resurgence of attacks on the construct validity of multiple-choice reading comprehension tests such as the SAT, the most damaging being the claim that test takers do not or need not read and comprehend the passages on which the questions are based. The conclusion that "the passages are not necessary to the task" (Katz et al., 1990, p. 126) is based on the finding that test takers can achieve better-than-chance performance on the SAT multiple-choice items even if they are not given the passages to read. Although we also investigated the consequences of completing the SAT when the passages are unavailable (Experiment 3), our main

tactic was to investigate strategies that test takers might be more likely to invoke under typical testing circumstances where the passages are always fully available. Experiment 1 was an exploratory study in which we investigated four test-taking strategies that college students spontaneously use when completing a multiple-choice reading comprehension test (see Farr et al., 1990). Although the effects of the different strategies did not appear to be large, we were concerned that our design lacked power and reliability because test takers used a given strategy for only two passages, the equivalent of half of one SAT form. Consequently, in Experiment 2, we investigated only two of the four strategies, the strategy that involved the greatest amount of advance reading of the passages and questions and the strategy that involved the least (i.e., no advance reading, just question-directed search).

There are two simple messages we take from the results of this research. The first message comes in the form of advice to would-be test takers. Although our research corroborates Katz et al. (1990) insofar as we found that passage reading is not necessary to achieve better-than-chance performance, our research strongly advises test takers to do as much passage reading as they can if they want to achieve competitive test scores (see also Freedle & Kostin, 1994; Powers & Leung, 1995). Indeed, our results show that test takers achieved the poorest level of performance (35%) if they did not read or consult the passage at all, they achieved an intermediate level of performance (63%) if they did no advance global reading of the passage but used it to search for answers to the specific questions, and they achieved the highest level of performance (71%) if they read the entire passage in advance and then reread portions of the passage to answer specific questions.

The second message is an encouraging one for test makers, test users, and anyone who has a stake in the construct relevance of the SAT. By using an individual-differences approach that compared the relative powers of working memory span and the Nelson-Denny at predicting SAT performance for different test-taking strategies, we showed that the SAT appears to be tapping reading comprehension processes as long as test takers engage in at least some reading of the passages themselves. Moreover, even though the question-directed reading strategy produces poorer levels of performance than the advance reading strategy, it does not appear to undermine the construct validity of the SAT as a measure of reading comprehension. Indeed, even if test takers do no reading of the passage at all, the SAT retains construct validity in the broader sense because test takers are still engaging in complex verbal reasoning skills that are important to and predictive of academic success.

These positive messages notwithstanding, our findings do raise a number of important issues that may be of concern to developers of multiple-choice reading comprehension tests such as the SAT. We have stressed the positive aspects of our findings, namely, that performance is enhanced if test takers do global reading of the passage rather than simply question-directed reading and that even if they do question-directed reading, construct validity is not jeopardized. However, one could give these findings a more negative spin. After all, an 8% improvement in SAT performance for having done all that advance passage reading is not such a large improvement. Some of that 8% improvement may be due to advance reading of the questions and not to advance reading of the passages.¹⁴ Even if the improvement was all due to advance reading of the passages, its relatively modest size suggests that the

SAT does not place a heavy premium on forming a coherent model of the overall text meaning (Johnston, 1984). If anything, our findings of relatively little difference between the advance reading and question-directed reading strategies can be taken to mean that test takers need to use information stated or implied in the passages to answer most of the SAT questions but that they do not necessarily have to form a coherent representation of the passage meaning to answer most of the questions. Taking this perspective, one could conclude that the SAT has construct validity inasmuch as it assesses individual differences in test takers' abilities to extract meaning from the passages; however, one might question whether the SAT has construct validity in the sense of assessing test takers' abilities to form coherent representations of passage meaning.

Theoretical Issues

Although the primary goals of this research were applied rather than theoretical, the research does have some theoretical implications. The main theoretical insight from the present research has to do with our finding that epistemic awareness about the importance of knowledge and integration can compensate to some extent for a limited working memory capacity. Like many previous studies, this study provides strong correlational evidence for working memory's role in reading comprehension processes. The study shows that reading span, a measure of the combined processing and storage capacity of working memory, was an excellent predictor of performance on the Nelson-Denny test of reading comprehension in all three experiments and that it was also an excellent predictor of SAT performance when test takers read the passages in advance or searched the passages in a question-directed manner. So, for example, individual differences in test takers' reading spans accounted for 36% of the variance in SAT performance when test takers were required to read the questions and the passage before beginning to answer the questions. The new contribution of the present study is to show that epistemic knowledge is a good predictor of performance on this advance reading task, with individual differences in test takers' awareness of the complexity of knowledge and the importance of integration to learning accounting for a further 12% of the variance in SAT performance after the effects of reading span were partialled out. Moreover, even though reading span is clearly the more important contributor to test success of the two, mature epistemic awareness can compensate for a limited working memory capacity in an interesting way. If test takers had large working memory spans, they performed well on the advance reading SAT task regardless of whether they had mature epistemic beliefs ($M = 82\%$) or naive epistemic beliefs ($M = 79\%$). On the other hand, if test takers had small working memory spans, they did very poorly if they also had naive epistemic beliefs ($M = 57\%$), but they were able to perform quite well if they had mature epistemic beliefs ($M = 73\%$). These findings suggest that students who had mature beliefs about

¹⁴ Of course, we cannot determine how much of the 8% advantage of advance reading over no advance reading was a result of advance reading of the set of questions or advance reading of the whole passage. However, we suspect it was largely from reading the passage because there was absolutely no benefit from advance reading of the passage and the questions (Strategy 3) over advance reading of the passage alone (Strategy 1) in our exploratory Experiment 1.

knowledge and integration may have recognized the need to initiate strategies such as backtracking and rereading when they encountered comprehension difficulties or information that might be relevant to the questions they had previewed. These strategies would have had the effect of compensating to some degree for the students' limited capacities to hold information in working memory during reading. Such compensatory strategies should be evident in the eye movement patterns of test takers who have small working memory spans but mature epistemic beliefs, and we have studies underway to explore them.

So far, we have been conceptualizing epistemic knowledge (or knowledge about knowledge) as if it were a kind of long-term knowledge that is theoretically unrelated to the working memory construct. On the other hand, we have also been conceptualizing epistemic knowledge as a form of metacognitive knowledge or awareness, and we have been arguing that individuals with mature epistemic awareness seem to be monitoring the reading situation for ways to facilitate knowledge acquisition and integration. The monitoring and controlling functions of cognition are functions that are typically attributed to the central executive component of working memory (see Baddeley, 1986; Baddeley & Logie, 1999; Engle, Kane, & Tuholski, 1999), and so, to that extent, the skills exhibited by test takers with mature epistemic beliefs would be under the control of the central executive. Our findings have shown one way in which the control and monitoring functions of working memory can serve to compensate for limitations in the active maintenance functions of the working memory system.

References

- Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. (1985). *Becoming a nation of readers: The report of the Commission on Reading*. Washington, DC: National Academy of Education.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.
- Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.
- Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). New York: Cambridge University Press.
- Baddeley, A. D., Logie, R., Nimmo-Smith, I., & Brereton, N. (1985). Components of fluent reading. *Journal of Memory and Language*, *24*, 119–131.
- Benton, S. L., Kraft, R. G., Glover, J. A., & Plake, B. S. (1984). Cognitive capacity differences among writers. *Journal of Educational Psychology*, *76*, 820–834.
- Brown, J. I., Bennett, J. M., & Hanna, G. (1981). *Nelson-Denny Reading Test*. Chicago: Riverside.
- Cohen, A. (1984). On taking language tests: What the students report. *Language Learning*, *1*, 70–81.
- 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., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 561–583.
- Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*, 1–18.
- Daneman, M., & Merikle, P. M. (1996). Working memory and comprehension: A meta-analysis. *Psychonomic Bulletin and Review*, *3*, 422–433.
- Donlon, T. F. (1984). *The College Board technical handbook for the SAT and Achievement Tests*. New York: College Entrance Examination Board.
- Drum, P. A., Calfee, R. C., & Cook, L. K. (1981). The effects of surface structure variables on performance in reading comprehension tests. *Reading Research Quarterly*, *16*, 486–514.
- Engle, R. W. (1996). Working memory and retrieval: An inhibition-resource approach. In J. T. E. Richardson, R. W. Engle, L. Hasher, R. H. Logie, E. R. Stoltzfus, & R. T. Zacks (Eds.), *Working memory and human cognition* (pp. 89–119). New York: Oxford University Press.
- Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). New York: Cambridge University Press.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.
- Farr, R. (1972). *Iowa Silent Reading Test, Level 3, Form E*. New York: Harcourt Brace Jovanovich.
- Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement*, *27*, 209–226.
- Freedle, R., & Kostin, I. (1994). Can multiple-choice reading tests be construct valid? *Psychological Science*, *5*, 107–110.
- Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Hillsdale, NJ: Erlbaum.
- Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology*, *93*, 103–128.
- Johnston, P. (1984). Prior knowledge and reading comprehension test bias. *Reading Research Quarterly*, *19*, 219–239.
- Jurden, F. H. (1995). Individual differences in working memory and complex cognition. *Journal of Educational Psychology*, *87*, 93–102.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.
- Katz, S., Blackburn, A. B., & Lautenschlager, G. J. (1991). Answering reading comprehension questions without passages on the SAT when items are quasi-randomized. *Educational and Psychological Measurement*, *51*, 747–754.
- Katz, S., Lautenschlager, G. J., Blackburn, A. B., & Harris, F. H. (1990). Answering reading comprehension items without passages on the SAT. *Psychological Science*, *1*, 122–127.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Cambridge University Press.
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity?! *Intelligence*, *14*, 389–433.
- Kyllonen, P. C., & Stephens, D. L. (1990). Cognitive abilities as determinants of success in acquiring logic skills. *Learning and Individual Differences*, *2*, 129–150.
- Masson, M., & Miller, J. A. (1983). Working memory and individual differences in comprehension and memory of text. *Journal of Educational Psychology*, *75*, 314–318.
- Posner, M. I., & Rossman, E. (1965). Effects of size and location of informational transforms upon short-term retention. *Journal of Experimental Psychology*, *70*, 496–505.

- Powers, D. F., & Leung, S. W. (1995). Answering the new SAT reading comprehension questions without the passages. *Journal of Educational Measurement*, 32, 105-129.
- Reitman, W. (1970). What does it take to remember? In D. A. Norman (Ed.), *Models of human memory* (pp. 469-509). New York: Academic Press.
- Royer, J. (1990). The sentence verification technique: A new direction in the assessment of reading comprehension. In S. Legg & J. Algina (Eds.), *Cognitive assessment of language and math outcomes* (pp. 144-191). Norwood, NJ: Ablex.
- Rukavina, I., & Daneman, M. (1996). Integration and its effect on acquiring knowledge about competing scientific theories from text. *Journal of Educational Psychology*, 88, 272-287.
- Schneider, W. (1988). Micro experimental laboratory: An integrated system for IBM-PC compatibles. *Behavior Research Methods, Instruction and Computers*, 20, 207-217.
- Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology*, 82, 498-504.
- Schommer, M., Crouse, A., & Rhodes, N. (1992). Epistemological beliefs and mathematical text comprehension: Believing it is simple does not make it so. *Journal of Educational Psychology*, 84, 435-443.
- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, 125, 4-27.
- Shute, V. J. (1991). Who is likely to acquire programming skills? *Journal of Educational Computing Research*, 7, 1-24.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127-154.
- WAIS-III technical manual. (1997). San Antonio, TX: Harcourt Brace.
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third Edition*. San Antonio, TX: Psychological Corporation.

Received August 13, 1999

Revision received November 15, 1999

Accepted November 30, 1999 ■



AMERICAN PSYCHOLOGICAL ASSOCIATION SUBSCRIPTION CLAIMS INFORMATION

Today's Date: _____

We provide this form to assist members, institutions, and nonmember individuals with any subscription problems. With the appropriate information we can begin a resolution. If you use the services of an agent, please do NOT duplicate claims through them and directly to us. **PLEASE PRINT CLEARLY AND IN INK IF POSSIBLE.**

PRINT FULL NAME OR KEY NAME OF INSTITUTION _____

MEMBER OR CUSTOMER NUMBER (MAY BE FOUND ON ANY PAST ISSUE LABEL) _____

ADDRESS _____

DATE YOUR ORDER WAS MAILED (OR PHONED) _____

CITY _____

STATE/COUNTRY _____

ZIP _____

 PREPAID CHECK CHARGE
CHECK/CARD CLEARED DATE: _____

YOUR NAME AND PHONE NUMBER _____

(If possible, send a copy, front and back, of your cancelled check to help us in our research of your claim.)

ISSUES: MISSING DAMAGED

TITLE _____

VOLUME OR YEAR _____

NUMBER OR MONTH _____

Thank you. Once a claim is received and resolved, delivery of replacement issues routinely takes 4-6 weeks.

(TO BE FILLED OUT BY APA STAFF)

DATE RECEIVED: _____

DATE OF ACTION: _____

ACTION TAKEN: _____

INV. NO. & DATE: _____

STAFF NAME: _____

LABEL NO. & DATE: _____

Send this form to APA Subscription Claims, 750 First Street, NE, Washington, DC 20002-4242

PLEASE DO NOT REMOVE. A PHOTOCOPY MAY BE USED.