Structure Building Differences Impact Learning from Educational Text: Effects on Encoding,
Retention, and Metacognitive Control

Nicole D. Martin ^a, Khuyen Nguyen ^a, and Mark A. McDaniel ^a

^a Washington University in St. Louis

One Brookings Drive

St. Louis, MO 63130, USA

Nicole D. Martin (ndmartin@wisc.edu), Khuyen Nguyen, (khu.nguyen.10@gmail.com)

Mark A. McDaniel (markmcdaniel@wustl.edu)

Nicole D. Martin is now at Department of Educational Psychology, University of Wisconsin-Madison.

This research was supported with a Collaborative Activity Award from the James S. McDonnell Foundation. The second author was supported by the National Science Foundation Graduate Research Fellowship Program.

Correspondence concerning this article should be addressed to Nicole D. Martin, Department of Educational Psychology, University of Wisconsin-Madison, □1025 West Johnson Street, □Madison, WI 53706. Phone: +1-512-924-3305 Email: ndmartin@wisc.edu

Structure Building Differences Influence Learning from Educational Text: Effects on Encoding,
Retention, and Metacognitive Control

Submitted: December 20, 2015

Abstract

Structure building describes the process by which people mentally organize information while reading in order to comprehend and later recall text. We investigated how individual differences in structure building ability affect students' learning of complex, educational texts. College students studied a complex text with a mechanical theme, engaged in retrieval practice, made metamemory judgments, and then were given another study opportunity. Following a short delay, memory and comprehension of the text were assessed with a range of dependent measures, including performance on free recall, multiple-choice questions, and problem-solving questions. Participants with high structure building ability outperformed those with low structure building ability on information recalled, factual and inference multiple-choice questions, and problem solving questions. Although high and low structure builders demonstrated equivalent metamemory accuracy, high structure builders appear to better regulate their restudy time according to a discrepancy reduction strategy, which allowed them to acquire more new information that they were unable to retrieve initially during recitation. Our findings suggest that low structure builders may suffer from deficiencies at many levels of text representations as well as deficiencies in metacognitive control during restudy. Our study highlights structure building ability as an important individual difference for learning educational texts and furthers our understanding of exactly what aspects of learning are related to these differences.

Keywords: Structure building; Learning; Individual differences; Educational text; Metacomprehension

Structure Building Differences Impact Learning from Educational Text: Effects on
Encoding, Retention, and Metacognitive Control

1. Introduction

Understanding and remembering information from texts is an essential component of learning in educational contexts. Unfortunately, many students struggle with this (Mayer, 2008). Although it is important to recognize that students have difficulties comprehending and remembering texts, it is even more crucial to pinpoint exactly where deficiencies exist in order to target those areas for improvement. Learning and comprehending information from a text is a complicated process that involves encoding, organizing information, forming connections, and creating mental models of the text. Accordingly, identifying the specific nature of the difficulties faced by students who struggle to learn and comprehend texts is a challenging task. In the present study, we take the approach of anchoring our individual-difference analysis within a general dimension of text comprehension, that of *structure building*. We attempt to pinpoint components of text representation, metacomprehension, and metacognitive control of studying text that may vary across individual differences in structure building and which may impact differences in learning and comprehension.

According to Gernsbacher's (1997) structure building model, the first step in text comprehension is laying a foundation, which involves setting the stage for further reading by providing a representational structure upon which to build. Evidence of people spending more time reading the first words of a sentence and the first sentences of a paragraph supports the idea that people use the early components of texts to form a foundation for the larger segments of text to follow (Aaronson & Scarborough, 1976; Cirilo & Foss, 1980). Next, readers map new information onto previous information, enhancing their mental structure of the text. If incoming

information coheres to the segments already organized in their mental model, then mapping proceeds smoothly, building one word, sentence, or idea on another. When incoming information diverges from the structure being built, readers engage in the next step of structure building: shifting. Shifting is a cognitive process whereby readers add a new branch onto their mental structure to begin a new idea. Although shifting is necessary to keep mental structures of information organized, shifting too often can create disorganized structures that have little continuity. A final component of structure building involves the suppression of information that is irrelevant to the developing mental structure. In order to build the most functional mental structure, readers must suppress information that does not meaningfully contribute to comprehending the text.

Gernsbacher and Varner (1988) designed the Multi-Media Comprehension Battery (MMCB) to assess people's general text comprehension from the perspective of structure building framework. The MMCB has been the measure of comprehension used in many studies that strengthen our understanding of the structure building framework and investigate comprehension from a structure building perspective (Gernsbacher, Varner, & Faust, 1990; Gernsbacher, 1993; Gernsbacher, 1997; Gernsbacher, Robertson, Palladino, & Werner, 2004). High and low scores on the MMCB further correspond to the individually targeted processes of shifting and suppression (see Gernsbacher, 1997, for a comprehensive discussion of how scores on the MMCB reflect structure building ability). The MMCB, thus, provides a comprehensive measure of people's ability to lay a foundation, shift appropriately, and suppress irrelevant information to build an overall well-organized mental structure of information. In the present study we do not directly examine these underlying theoretical processes, but instead focus on their consequences for learning from educational materials; further, as developed in the next

section we attempt to illuminate the particular representational levels (extracted by the learner) that are influenced by differences in structure building (as assessed by the MMCB) and examine possible outcomes of individual differences in structure building on metacomprehension processes.

It is first important to note that structure building is theoretically different from reading abilities indexed by standard reading tests (see Callender, 2008, for discussion). In line with this observation, the MMCB measure of structure building ability only modestly correlates with the Nelson-Denny reading test (Maki, Jonas, & Kallod, 1994). Furthermore, McDaniel, Hines, and Guynn (2002) showed that low structure builders demonstrated different deficiencies than low Nelson-Denny readers. Specifically, low structure builders showed improved recall when asked to engage in a sentence unscrambling task (sentences from a folk-tale were presented in random order and had to be rearranged to produce a coherent tale) relative to reading an intact folk tale. Having to reorder the sentences presumably stimulated the low structure builders to engage in organizational processing that for them appears not to spontaneously occur otherwise. In contrast, low Nelson-Denny readers did not benefit from this sentence unscrambling task, relative to the intact folk-tale condition. This may be because low Nelson-Denny readers do not have impaired abilities to develop an organized representation of the text but instead have deficient word decoding processes. (This latter possibility dovetails with the additional findings that low Nelson-Denny readers but not low structure builders showed lower recall of a text with letters deleted from words relative to the intact text.) These results suggest that although both structure building and Nelson-Denny measures are predictive of reading comprehension, they focus on two different processes: structure building is focused on higher level processes that contribute to building a coherent mental model, whereas Nelson-Denny is more focused on word

decoding processes (e.g., Mason, 1978; Petros, Bentz, Hammes, & Zehr, 1990). Consequently, there appears to be value in investigating structure building as a separate construct, distinct from more traditional reading comprehension tests, as the processes captured in this construct may uniquely contribute to individuals' abilities to learn and to comprehend texts.

Though much research has been conducted to understand reading comprehension (Graesser, Singer, & Trabasso, 1994; Zwaan & Radvansky, 1998; McNamara & Magliano, 2009), the consequences of individual differences in structure building for learning *educationally* relevant materials has received virtually no investigation. The previous research on structure building has focused on individual differences to distinguish skilled and less-skilled structure builders (McNamara & Magliano, 2009), but structure building has been predominantly assessed with narrative materials—materials in which causal structures, familiar scripts, and narrative superstructures (narrative grammars) are prominent in contributing to readers' efforts to construct effective mental representations (e.g., see van Dijk & Kintsch, 1983; McDaniel, Hines, Waddill, & Einstein, 1994). Expository and didactic passages common in educational settings do not afford these kinds of organizational advantages (van Dijk & Kintsch, 1983; McDaniel & Einstein, 1989). Accordingly, it is possible that the structure building construct (as currently assessed) might not be especially informative with regard to the real world context of classroom learning and studying. Indeed, only three extant studies (of which we are aware) have investigated the relationship between structure building and learning for educationally-relevant materials (Bui & McDaniel, 2015; Callender & McDaniel, 2007; Maki & Maki, 2002).

Bui and McDaniel (2015) and Callender and McDaniel (2007) examined the relationship between individual differences in structure building and learning with expository materials. Low structure builders showed impaired performance compared to high structure builders on a

classroom-type test (e.g., multiple-choice and short answer questions) after reading a social-psychology chapter (Callender & McDaniel) and after listening to a short text (being read) on how mechanical devices work (Bui & McDaniel). Maki and Maki (2002) found that structure building ability was a significant predictor of performance in online Introductory Psychology courses. These initial findings suggest that structure building ability could be a critical individual difference factor in determining successful learning in educational contexts. Therefore, one purpose of the present study was to extend the investigation of the consequences of individual differences in structure building for learning and retention of technical information communicated in expository text. Our first aim was to further document possible deficiencies in low structure builders' understanding and remembering of expository discourse.

1.1 Different Levels of Text Representation

The initial evidence of differential learning from educationally authentic texts between high and low structure builders raises a central question of precisely how and why low structure builders' text representations are deficient. The structure building framework predicts that learners who have low structure building abilities, as measured by performance on the MMCB, would have poor access to recently comprehended information due to an inefficiently organized mental structure of the information (Gernsbacher, Varner, & Faust, 1990); however, it is not well documented how this framework predicts performance at a variety of different levels of text representation. To gain insights into this issue, a second aim of the current study was to explore whether low structure builders have difficulties encoding information, creating organized networks of information, using such networks to later remember information, applying that information to solve problems, or a combination of these processes (see Kintsch, 1988, and McNamara & Magliano, 2009, for further discussion of different levels of text representation).

The issue of how mental representations are impaired can be informed by probing learners on a range of tests over a previously studied text.

Following Mayer and Gallini (1990; see also Bui & McDaniel, 2015), we adopted a battery of comprehension and retention measures that included multiple-choice items on facts, free recall, inference questions (multiple-choice), and problem solving. In line with Mayer and Gallini, we assumed that encoding of propositions conveyed by the text would be revealed by the factual multiple-choice and free recall performances. Free recall would in addition reflect potential differences between low and high structure builders' abilities to build an organized, interconnected representation of the information in the text. We reasoned that performance on the inference multiple-choice and the problem solving questions would index possible differences in structure building ability in terms of being able to form a coherent mental model of what the text was about (how brakes worked, how pumps worked; see Mayer & Gallini, 1990). Given the posited deficiencies in suppression and shifting for low structure builders, we anticipated that structure building deficits would emerge for free recall, inference multiple-choice, and problem solving performance. We were less certain about performance on factual multiple-choice questions; it may be that low structure builders adequately encode individual propositions in a text and consequently would perform as well as high structure builders on factual multiple-choice items.

1.2 Metacomprehension

Finally, though current theorizing is mute on this point, structure building ability may also impact metacomprehension, which refers to the ability to monitor ongoing comprehension and use effective study policies (Dunlosky & Lipko, 2007). The first important metacognitive component to consider in learning is that of metacomprehension accuracy. Much research has

examined metacomprehension accuracy, because being aware of what you do or do not know is crucial for deciding how to allocate study time. When students read a text with the goal of learning the material for a future test, they must be able to monitor their learning and comprehension in order to successfully fill the gaps in their mental models when given another opportunity to study.

The second important metacognitive component is to implement study policies that depend, at least in part, on metacomprehension judgments. A student's general knowledge of metacognitive strategy has been shown to be an important predictor of text comprehension and learning (Schaffner & Schiefele, 2013). Schaffner and Schiefele (2013) found that students who demonstrate superior knowledge of study strategies learn more than students without such knowledge. Schaffner and Schiefele investigated the participants' general metacognitive strategies, but did not examine which strategies participants actually used while restudying. Thus, further research is necessary to better understand what strategies students use to guide their study time. More critically for present purposes, we were interested in whether differences in structure building ability might be related to metacomprehension. Because poor structure builders have difficulties with learning from texts (Gernsbacher, Varner, & Faust, 1990) and students with poor metacognitive strategy also have difficulties learning from texts (Schaffner & Schiefele, 2013), there may be an important link between these two abilities that has yet to be uncovered.

One metacognitive study policy that is thought to be effective, and thus motivates our current investigation, comes from the Discrepancy-Reduction framework (Dunlosky & Thiede, 1998). According to this framework, a student has a desired level of learning and a perceived level of learning. Because there is often a discrepancy between the desired level and the

perceived level of learning, the goal of studying is to reduce this discrepancy. Reducing this discrepancy typically requires learners to constantly monitor what they do and do not know in order to accurately target the information yet to be learned (Thiede & Dunlosky, 1999). Thus, this framework predicts that students will allocate less study time to information judged as well-learned and more study time to information judged as less well-learned. Early research using word lists and paired-associates has largely confirmed these predictions. For instance, Nelson, Dunlosky, Graf, and Narens (1994) found that students are more inclined to study items judged more difficult than items judged less difficult. Findings from Mazzoni, Cornoldi, and Marchitelli (1990) also support the Discrepancy-Reduction framework by showing that students do in fact allocate more study time to items they judge as more difficult to learn. More recently, Thiede, Anderson, and Theriault (2003) extended these findings using more educationally relevant materials (i.e., text passages). Specifically, Thiede et al. (2003) found that learners were more likely to restudy passages judged as less well-learned than passages judged as better learned.

1.3 The Present Study

In the present experiment, we used the Read-Recite-Review (3R) study strategy developed by McDaniel, Howard, and Einstein (2009) as a vehicle to evaluate where low structure builders suffer in learning from educational texts. By engaging participants in the 3R study strategy, we can gain insight into different phases of the studying process that might be impaired for low structure builders. In the 3R study strategy, students read through a passage once, recite as much information as they can remember, and then review (restudy) the passage for as long as they want. Although this strategy has been shown to be significantly more effective than traditional study strategies such as rereading or note-taking (McDaniel et al., 2009), we utilized this experimental paradigm to investigate different facets of the studying and

learning process across high and low structure builders. The recite phase of this paradigm gauges how much information people learn on an initial reading. Further, recitation performance can be compared to the performance on a final test to reveal how much information people learn during the review phase. The review (or restudy) phase of the paradigm provides an opportunity to investigate learners' metacognitive processes. Specifically, by requiring participants to make judgments of learning (JOLs) before this restudy phase, we can determine their metacomprehension accuracy, and we can analyze their study policies, particularly with regard to their fidelity to a discrepancy reduction strategy. Finally, the 3R paradigm allows us to examine whether participants' study policies promoted learning new information during the review phase.

We predicted that high structure builders would not only learn more information overall than low structure builders, but they would additionally learn more during their restudy of the passage. If in fact high structure builders perform better on the outcome measures and learn more than low structure builders, as predicted, the remaining question is *why* this difference occurs. We intended to answer this question by analyzing how high and low structure builders differ in their metacognitive study strategies. It is possible that high structure builders can more accurately judge what information they do not know and are more likely to adopt a discrepancy-reduction study policy to effectively regulate their study, thus resulting in high performance on the final test. Another possibility is that both high and low structure builders are equally as likely to utilize the discrepancy reduction strategy but high structure builders are more accurate about what they do not know and, therefore, better allocate their study time to the information they have not learned. A third possibility is that high and low structure builders are equally as accurate about what information they do not know, but high structure builders are more likely to utilize their metacomprehension judgments to engage a discrepancy-reduction strategy.

2. Methods

2.1 Participants and Design

The participants in this study were 72 students from a mid-sized university located in a large city in the mid-western United States. Participants consisted of 44 females and 28 males ranging from ages 18 to 24. Additionally, the participants came from a highly capable student population with average verbal SAT scores of 700 to 760. All participants received partial course credit or payment. We dropped one participant for failing to follow instructions, leaving 71 participants tested for the experiment. This was an extreme-groups (between-subjects) design in which participants were divided into high and low structure-builders based on their performance during the experiment on the written portion of the Multi-Media Comprehension Battery (Gernsbacher & Varner, 1988). Using Callender and McDaniel's (2007) cutoff (low SB = below 32, high SB = 36 and above), we were left with a total of 20 low structure builders ($M = 26.50$, $SD = 4.10$) and 32 high structure builders ($M = 38.13$, $SD = 2.21$) for further analysis. Participants who scored in between the low and high cutoffs were considered neither low nor high structure builders and were not included in analyses (all participants studied the passage using the aforementioned 3R strategy with a subset [$n = 48$] also making metacomprehension judgments). In the subset of participants who also made metacomprehension judgments, there were 16 low and 16 high structure builders¹.

2.2 Materials

We selected two scientific passages on brakes and pumps from Mayer and Gallini (1990) and slightly modified each passage so they both contained eight distinct paragraphs. A subset of the participants ($n = 16$ and 11 for high and low SB respectively) read the brakes passage (1029 words) and a subset ($n = 16$ and 9 for high and low SB respectively) read the pumps passage

(996 words). Following McDaniel et al. (2009), we used three types of criterial tests to assess performance and learning from the passages. First, participants received a free recall test in which they were asked to write down as much as they could remember from the passage. Recall performance was scored as the proportion of idea units recalled. The brakes passage consisted of 101 idea units and the pumps passage included 105 idea units, with each idea unit roughly corresponded to one phrase of a sentence. For example, the sentence “a brake is a device that slows down or stops a moving object” contains two idea units, “a brake slows down a moving object”, “a brake stops a moving object”. The first and second authors scored approximately 20% of participants’ free recall protocols. Inter-rater reliability was high ($r = .95$) so the first author scored the remaining tests.

Second, participants received a 12-item multiple-choice test that included eight factual questions and four inference questions. We developed one factual question for each paragraph of text that required only detailed knowledge provided in that paragraph (e.g., Which of the following is not a component of a disc brake?). The inferential questions required integration of knowledge from two or more different paragraphs as well as prior knowledge to make inferences (e.g., Your fancy new car includes both an antilock brake system and a traction control system. After owning the car for a few months you realize that your brakes are still working but neither of these systems seems to be functioning. Failure of which of the following components is most likely to cause this problem?). All multiple-choice questions were scored as either correct, receiving one point, or incorrect, receiving zero points. Finally, everyone received four short-answer problem-solving questions (e.g., What are some ways that a car’s brake system might be improved to make the system more efficient, that is, to reduce the distance needed to stop?). A scoring rubric was developed for each question, and participants received points for correct

answers with partial credit possible. The first and second authors both scored approximately 20% of the problem-solving answers. Inter-rater reliability was high ($r = .90$) so the first author scored the remaining tests.

We developed the metacomprehension judgments concerning participants' future memory for the passages read. Those assigned to the metacomprehension groups made either general judgments (i.e., On a scale of 0 to 100...How likely would you be to remember this information on a future test?) or judgments that probed specific knowledge presented in a given paragraph (e.g., How likely would you be to remember the components of hydraulic brakes, how they work, and what they are used for?). Metacomprehension judgments regarding performance on a future memory test on recently read texts have been shown to be stable over time and have high test-retest reliability, with a correlation between judgments made one week apart of $r = .69$ (Kelemen, Frost, & Weaver, 2000). People are equally as confident about their future memory of texts when asked to make a second round of judgments a week later.

The written portion of the Multi-Media Comprehension Battery (Gernsbacher & Varner, 1988) was used to characterize participants' structure building ability and consisted of four short narratives ranging from 500 to 900 words. Participants were instructed to read each short story at their own pace. They were instructed to press the SPACE key to advance from one paragraph to the next. At the conclusion of each passage, participants answered 12 multiple-choice comprehension questions about key details presented in the text. As a result, structure building scores could range from zero to 48. We conducted a split-half reliability test to determine the reliability of the structure building assessment. This method of examining the reliability of a test is typically viewed as an index of internal consistency, and such measures are commonly generalized to reflect how well each participant's performance on the test items can be

generalized to possible questions that could have been asked (Crocker & Algina, 1986). This method also allowed us to purposefully determine half-tests based on odd and even number questions to evenly distribute content and difficulty. We correlated participants' structure building performance on the odd and even numbered questions, resulting in a correlation of .55.

2.3 Procedure

Participants were presented with a passage, about how brakes or pumps work, and were instructed to read through the passage once at their own pace. After they indicated that they were finished reading the passage, participants were instructed to orally recite as much as they could remember from the passage into a tape recorder. Following the recitation phase, participants played Tetris for two and a half minutes and were then given an opportunity to restudy the passage again. The passage during this restudy was presented paragraph by paragraph in the same order as in the original presentation, and participants were allowed to study each paragraph for as long as they wanted. A subset of participants was given instructions to make metacomprehension judgments after playing Tetris and prior to the restudy opportunity. For these participants, each paragraph was presented one by one on the screen for 15 seconds and they were asked to make a metacomprehension judgment. After the study phase, participants completed the Multi-Media Comprehension Battery and then the three criterial tests were administered.

3. Results

3.1 Study Time

We first looked at average time high and low structure builders spent studying the passages to see if any differences we found in learning could simply be attributed to more time spent studying. Means and standard deviations of study time are presented in Table 1. There

were no differences in amount of time spent during initial study, $t(50) = 0.43, p > .10$, or restudy, $t(50) = 0.87, p > .10$. There was, however, a marginally significant difference in the amount of time spent reciting, $t(50) = 1.91, p = .06, d = .56$. High structure builders spent more time engaging in recitation. Nevertheless, overall study time (initial study plus recitation plus review) did not significantly differ between the two groups, $t(50) = 1.21, p > .10$.

3.2 Free Recall Performance

The free recall performance was used to investigate potential differences between low and high structure builders' abilities to encode propositions conveyed by the text and to build an organized, interconnected representation of the information in the text. Building such an organized representation during studying would help participants recall more information later. We scored free recall performance as the proportion of idea units recalled. We calculated this proportion by dividing the number of idea units each participant recalled by the total idea units present in the passage they read, 101 idea units for the brakes passage and 105 idea units for the pumps passage. As seen in Table 1, high structure builders ($M = .33, SD = .11$) recalled substantially more idea units than low structure builders ($M = .19, SD = .10$), $t(50) = 4.60, p < .05, d = 1.33$. This difference between high and low structure builders is a large-size effect.

3.3 Multiple-Choice Performance

We used factual and inference multiple-choice questions to investigate differences between high and low structure builders at different levels of text representation. We used the factual multiple-choice questions as measures of participants' abilities to encode propositions from the text, and we used the inference multiple-choice questions as measures of participants' abilities to build coherent mental models of the text that they could apply to new situations. We conducted analyses for multiple-choice performance separately for factual and inference

questions. As seen in Table 1, high structure builders ($M = .81$, $SD = .13$) outperformed low structure builders ($M = .68$, $SD = .18$) on factual questions, $t(50) = 3.05$, $p < .05$, $d = .83$, and high structure builders ($M = .73$, $SD = .22$) outperformed low structure builders ($M = .45$, $SD = .25$) on inference questions, $t(50) = 4.14$, $p < .05$, $d = 1.16$. Again, these differences reflect large effect sizes.

3.4 Problem Solving Performance

We used problem-solving performance as another measure of potential differences between high and low structure builders' abilities to create coherent mental models of the text that they could use to solve new problems. We scored problem-solving performance as the proportion correct. There were 10 possible points for the brakes passage and 12 possible points for the pumps passage. As seen in Table 1, high structure builders ($M = .35$, $SD = .13$) outperformed low structure builders ($M = .25$, $SD = .12$), $t(50) = 2.85$, $p < .05$, $d = .80$, with the difference reflecting a large effect size.

3.5 Metacomprehension: Judgments of Learning

We investigated whether there were differences in high and low structure builders' abilities to monitor their understanding and use effective study policies to learn. Means and standard deviations of all metacomprehension measures are presented in Table 1. For the subset of participants who engaged in metacomprehension monitoring, we calculated a mean JOL by averaging across the eight paragraph-by-paragraph JOLs. High ($M = 50.95$, $SD = 16.64$) and low ($M = 42.59$, $SD = 24.74$) structure builders did not differ in the magnitude of their metacomprehension judgments, $t(30) = 1.12$, $p > .10$.

We measured metacomprehension accuracy both in terms of *calibration* (absolute accuracy) and *resolution* (accuracy relative across different portions of the text, in this case

across paragraphs). We assessed calibration by computing bias scores, which were calculated by subtracting a participant's recall proportion from their mean JOLs. The range of bias scores is from -1 to 1, with zero reflecting no bias. Scores below zero reveal underconfidence and scores above zero reflect overconfidence. Both high ($M = .32$, $SD = .18$) and low ($M = .30$, $SD = .27$) structure builders were overconfident, ($ps < .05$), and there was no difference between the two groups, $t(30) = 0.26$, $p > .10$.

We assessed resolution by computing correlations between each participant's JOLs and their recall performance by paragraphs. For each participant, we computed a Goodman-Kruskal gamma correlation, which is recommended for analyzing metacognitive measures because it appears to be the most accurate measure of feeling-of-knowing in comparison to other commonly used measures of feeling-of-knowing (Nelson, 1984). All analyses were also conducted using Pearson correlations; because the pattern of results did not change, we will only report the results for the Goodman-Kruskal gamma correlations. The mean gammas for both high ($M = .25$, $SD = .39$) and low ($M = .45$, $SD = .38$) structure builders were greater than zero, ($ps < .05$), and the gammas between the two groups did not differ, $t(30) = 1.47$, $p > .10$.

We also computed gamma correlations between each participant's JOLs and their recitation performance by paragraphs. The mean gammas for both high ($M = .37$, $SD = .42$) and low ($M = .39$, $SD = .37$) structure builders were greater than zero ($ps < .05$), and the gammas between the two groups did not differ, $t(30) = .09$, $p > .10$.

3.6 Metacomprehension: Study Time Allocation

We looked at how participants allocated their study time during the restudy phase to investigate the study strategies that high and low structure builders used. For each participant, we computed a Goodman-Kruskal gamma correlation between each participant's JOLs and the

amount of time they spent studying each paragraph during the restudy phase. Because the paragraphs differed in length, we adjusted restudy times as a function of idea units in each paragraph². According to the Discrepancy-Reduction (DR) model of self-regulated learning, participants should allocate more study time for paragraphs given low JOLs and less study time for paragraphs given high JOLs (Thiede & Dunlosky, 1999). Both of these situations (high JOLs corresponding to low study time or low JOLs corresponding to high study time) would produce a negative correlation between study time and JOLs. Thus, a negative correlation would mean that people were strategically regulating their restudy according to the DR model, allocating more study time to paragraphs they judged as less well-learned and less study time to paragraphs they judged as more well-learned. As shown in Table 1, the mean gamma for the high structure builders was reliably less than zero ($M = -.22$, $SD = .36$), $p < .05$ whereas the mean gamma for the low structure builders was not significantly different from zero ($M = -.03$, $SD = .44$), $p > .10$. High and low structure builders did not significantly differ in gamma correlations, $t(30) = 1.33$, $p > .10$, however.

3.7 3R Recitation Performance

We assessed performance during the recitation phase of the experiment in order to see how much information participants learned from their initial reading of the passage. We scored recitation performance as the proportion of idea units recited. High structure builders ($M = .18$, $SD = .07$) recited more idea units than low structure builders ($M = .13$, $SD = .06$), $t(50) = 2.84$, $p < .05$, $d = .82$.

We also used recitation performance as a base line to compare to the information participants recalled on the final test in order to assess differences in how much new information participants learned during the restudy phase. In light of the finding that high structure builders

but not low structure builders appeared to adopt a DR strategy during restudy, one might expect that high structure builders gained more from restudy than did low structure builders. For completeness we computed the probabilities that people recalled previously recited (old) or non-recited (new) idea units on the final free recall test. The old idea units were recalled both during recitation and final recall, whereas the new idea units were recalled during final recall but not during recitation. Thus, these new idea units represented the amount of information that was presumably learned during restudy. High structure builders recalled more old information, $t(50) = 2.05, p < .05, d = .57$. Most critically, high structure builders recalled more new information than did low structure builders, $t(50) = 3.94, p < .05, d = 1.17$, (see Table 1 for means).

4. Discussion

In the present study, we sought to: 1) investigate whether individual differences in structure building would be associated with comprehension and learning for educationally-relevant scientific texts; of interest was whether the structure building construct would capture individual differences that are important in learning educationally relevant material. 2) Attempt to illuminate the specific levels of representation deficiency associated with poor structure builders. 3) Explore whether low structure builders' deficiencies include reduced metacognitive ability (relative to high structure builders) in terms of both metacomprehension accuracy and effectiveness of guiding restudy of to-be-learned material. Our findings suggest that low structure builders exhibit deficits across a wide range of processes, including deficiencies at different levels of text representation and in metacognition.

4.1 Low Structure Builder Deficiencies at Different Levels of Representation

The current study is the first to explore the nuanced differences in the different levels of representation with which low structure builders might have difficulties. A range of tests was

administered to investigate how structure building ability influences participants' representation of information at a propositional, text-based level (see Kintsch, 1974, and Kintsch & van Dijk, 1978, for detailed characterizations of this representational level) and at a situational model level (see Kintsch, 1988; Schmalhofer, McDaniel, & Keefe, 2002). Consider first, the structure building effects at the propositional and text-based level of representation. To provide a gauge of this level of representation, we appeal to performance on recitation, free recall, and factual multiple-choice tests. We found that low structure builders performed significantly worse than high structure builders on all three tests. Poor performance on recitation and free recall suggests that low structure builders struggle with encoding and retrieving propositional information presented in the text. In Gernsbacher's (1997) account, low structure builders have trouble forming cohesive, unified representations of the information in the text. The present recitation results (first oral recall) and final free recall results may suggest more specifically that low structure builders have impaired representations for at least the text-base (i.e., forming a coherent, interrelated network of propositions; Kintsch, 1978).

Further, low structure builders' relatively poor performance on the factual multiple-choice questions suggests that low structure builders' deficiencies include not just forming a coherent structure, but also include challenges in sufficiently encoding the propositional information. This conclusion is based on the assumption that a recognition test reduces demands on retrieval (relative to recall), and thus is not as dependent on well-organized mental structures as is free recall (Einstein & Hunt, 1980). Accordingly, it may be that low structure builders are plagued not just by deficits in their representational structures but also in the amount of information that they extract from the text.

Consider next performance on the inference multiple-choice and problem-solving tests, indices of mental models of the referent of the text (in this case how brakes work, how pumps work; Mayer & Gallini, 1990). Low structure builders' significantly lower inference multiple-choice and problem-solving performances, relative to high structure builders, suggest that their representations of the mental or situation model(s) that can be derived from the text are also not as rich or coherent as for high structure builders. This may be because low structure builders are forming a number of fractionated mental models (cf. Gernsbacher, 1997), or because they are challenged by the processes of constructing a rich single mental model (integrating prior knowledge with the propositional text base, constructing necessary inferences, and making higher-level connections; Kintsch, 1988; Schmalhofer et al., 2002). In line with this latter possibility, low ability readers are less likely to generate inferences while reading (Cain, Oakhill, Barnes, & Bryant, 2001).

It is important to note that we did not observe any differences in the amount of time spent reading the text, which suggests that low structure builders do not have problems decoding the text, which is assumed to be the underlying deficiency for low-ability Nelson-Denny readers (McDaniel, Hines, & Guynn, 2002). Thus, our results are consistent with previous research showing that structure building ability is only modestly associated with performance on the Nelson-Denny reading test (Maki, Jonas, & Kallod, 1994). Furthermore, unpublished studies in our lab indicate that structure building ability is also not highly correlated with working memory (i.e., Automated Operation Span). Participants ($n = 264$) showed a small, reliable correlation between structure building ability and working memory (.13) so we cannot definitively rule out that working memory plays a role in deficiencies associated with low structure building. However, given that the amount of variance in structure building accounted for by working

memory is very low (.017), it is implausible that the effects of structure building are accounted for by working memory per se. This finding suggests that low structure builders' deficiencies do not simply reflect a basic working memory problem but are more specifically tied to aspects of comprehension. Our findings taken together suggest that structure building ability may reflect proficiency at encoding and constructing several representational or structural levels that contribute to complete comprehension and learning from textual material.

4.2 Deficiencies in Metacomprehension Control

According to the Structure Building framework, low structure builders shift too often and thus store information in an unorganized manner that can adversely affect their memory and comprehension of the text (Gernsbacher, 1997). Accordingly, we might expect low structure builders to have a poor understanding of what they know and do not know. Low structure builders would have stored information, thus causing them to believe they know the information, but their disorganized structure would prevent actual understanding and later retrieval – reflecting a mismatch between their metacomprehension and actual performances on indices of learning and memory. However, our results suggest that low structure builders do not show any deficits in their ability to monitor their learning, as evidenced by comparable metacomprehension accuracy relative to high structure builders.

One potential explanation for the lack of difference in metacomprehension accuracy is that the use of contiguous texts, in which all eight paragraphs to be learned were related in content, prevented participants from using contextual cues to discriminate which paragraphs were less well learned. Thiede, Griffin, Wiley, and Anderson (2010) showed that the use of contextual cues drives differences in metacomprehension accuracy. If this is the case, then a lack of contextual cues might explain the lack of metacomprehension differences observed in our

study. It may have been difficult for participants to distinguish between less and better learned paragraphs due to their related nature and similar contextual cues. This potential explanation also has important implications for students studying educational texts because educational texts are most often contiguous with many related paragraphs and sections. We will return to these educational implications later in this paper.

Low structure builders do, however, falter in another metacognitive process – regulation of restudy time. This deficit in regulation of restudy time can be seen in the amount of new information learned during restudy. Specifically, on the final recall test high structure builders were able to recall a significantly higher proportion of the non-recited information than low structure builders (24% and 13%, respectively). It is important to note that these findings emerge despite the fact that both groups spent the same amount of time restudying the text passages. A potential factor that may be underlying this deficit in low structure builders is their inability to utilize effective study policies. JOL-restudy time correlations suggest that high structure builders (-.22) followed a discrepancy reduction strategy whereas low structure builders (-.03) did not. Although the correlations are not significantly different, these findings suggest that poor structure builders have a difficult time focusing and directing their study towards information that they judge to be not well learned, when given an opportunity to engage in self-regulated restudy. This finding is consistent with Schaffner and Schiefele (2013), who reported that some people have a good understanding of effective study methods whereas others do not. Furthermore, they found that high metacognitive knowledge of study strategies is correlated with reading comprehension when students do not have access to a text while being tested. With respect to our study, this suggests that perhaps students who do not have a good knowledge of effective metacognitive study strategies cannot control their study time allocation, and thus

develop poor structures, which leads to poor test performance as compared to the high structure builders.

Gernsbacher (1997) attributes impaired structure building to the inability to ignore irrelevant information while reading, thus causing the reader to shift too often when creating a mental representation of the text. The idea is that low structure builders construct mental models that are cumbersome and disorganized, with many substructures that are difficult to navigate when retrieving information. Our data implying that low structure builders do not use the effective study strategy of discrepancy reduction suggests an even richer theoretical characterization of individuals identified as low structure builders. Specifically, our results suggest that low structure builders may also suffer from a deficit in the control of their restudy. Low structure builders learned less new information during restudy than high structure builders, supporting this idea that low structure builders engage in less effective study strategies and monitoring. This deficit in effective metacognitive study strategies may be an underlying factor in the relation between low structure building ability and performance in authentic learning contexts (Maki & Maki, 2002). We discuss these implications in a following section.

4.3 Limitations

Although this study increases our understanding of how differences in structure building ability are related to learning from educational texts, there are several limitations. One limitation is the small sample size of high and low structure builders who were in the subset of participants who made metacomprehension judgments. In the design of this study, we randomly selected a subset of participants to make metacomprehension judgments, with two thirds of this subset falling into low and high structure building categories. The resultant sample size (16 low and 16 high structure builders) limits the robustness of our correlational findings implicating differences

in self-regulated study policies between high and low structure builders. The low sample size also limited the statistical methods we were able to use for our analyses. It could have been informative to conduct mediation analyses to investigate whether differences in metacomprehension might have mediated the relationship between structure building ability and learning performances, but the sample size was too low to warrant such analyses (Fritz & MacKinnon, 2007).

4.4 Educational Implications

The findings from the present study suggest that structure building ability is an important individual difference for learning from educational texts. This is consistent with Maki and Maki's (2002) finding that structure building ability was predictive of class performance in web-based Introductory Psychology courses. Our findings suggest that poor structure builders not only exhibit deficits in their representations of the text, but they also demonstrate poor metacognitive control during restudy, thereby further increasing their learning gaps relative to high structure builders. This individual difference of structure building ability produced large effects on learning and retention from authentic education text (the effect sizes for all of our learning/retention measures are in the "large" category), and thus, structure building may have significant bearing on how students perform in the classroom.

Regarding the role of structure building in metacomprehension, it is important to reiterate that our study purposefully used expository, educational texts similar to those that might be found in a classroom setting. Our use of contiguous text (rather than unrelated paragraphs) was intentionally to create a naturalistic educational setting, in contrast to most metacomprehension research which uses distinct paragraphs, sentences, or even simple word pairs. As discussed earlier, contiguous texts with related content might have contributed to the absence of

metacomprehension differences between low and high structure builders that were observed.

Clearly, further research is needed, as these initial metacomprehension findings relating to structure building are based on relatively small numbers of participants. More generally, however, the inherent lack of discriminating contextual cues in related contiguous texts may be important to consider when thinking about ways to help students better monitor their learning.

Furthermore, it is important to note that the current experiment was conducted with very able students, coming from a student population with average verbal SAT scores of 700 to 760. Even students who appear to be highly able on standard verbal measures, such as the SAT, may still have reduced structure building abilities that impair their academic performance. The fact that a fair proportion of low structure builders exist even among a highly able student population warrants important investigations into how to help these low structure builders. Some interventions have been shown to augment text comprehension in poor structure builders. Callender and McDaniel (2007) found that embedding key questions within the text enhanced performance for low but not high structure builders relative to a rereading condition. More recently, Bui and McDaniel (2015) found that supplying low structure builders with schematic diagrams as they were listening (and taking notes) to aural presentations (using texts adopted from those used here) significantly improved their free recall and problem solving performances. These findings suggest that poor structure builders can successfully learn from texts when armed with the proper interventions to help them build a more coherent representation of the text.

4.5 Future Research

This study highlights structure building as an important component for understanding individual differences in learning from education texts, but there is still much to be learned about how structure building affects learning. Future research could first strengthen the findings of this

current study by conducting a larger scale study of individual learning differences and metacognitive strategies. A larger scale study could also provide opportunities for further investigating whether differences in metacomprehension may mediate the effects of structure building. Future research should also continue investigating the specific mechanisms that underlie structure building ability, specifically to tease apart whether inhibitory deficits or difficulty activating relevant information are the core reasons why poor structure builders display reduced success in learning from educational texts. Additionally, future research might examine whether poor structure builders can be taught how to implement effective metacognitive strategies at restudy. Identifying important individual differences in learning is a necessary first step towards understanding how to help students who struggle to learn.

Acknowledgments

This research was supported with a Collaborative Activity Award from the James S. McDonnell Foundation (No. 220020041). The second author was supported by the National Science Foundation Graduate Research Fellowship Program.

References

- Aaronson, D., & Scarborough, H. S. (1976). Performance theories for sentence coding: Some quantitative evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 56-70.
- Bui, D., & McDaniel, M. A. (2015). Enhancing learning during lecture note taking using outlines and illustrative diagrams. *Journal of Applied Research in Memory and Cognition*, 4, 129-135.
- Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference making ability, and their relation to knowledge. *Memory & Cognition*, 29, 850-859.
- Callender, A. (2008). *The effects of prior knowledge relevance and organization on text comprehension and memory*. (Unpublished doctoral dissertation). Washington University, St. Louis, Mo.
- Callender, A., & McDaniel, M. A. (2007). The benefits of embedded questions adjuncts for low and high structure builders. *Journal of Educational Psychology*, 99, 339-348.
- Cirilo, R. K. & Foss, D. J. (1980). Text structure and reading time for sentences. *Journal of Verbal Learning and Verbal Behavior*, 19, 96-109.
- Crocker, L., & Algina, J. (1986). Procedures for Estimating Reliability. In *Introduction to modern and classical test theory* (pp. 131–156).
- Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science*, 16, 228-233.
- Dunlosky, J., & Thiede, K. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica*, 98, 37-56.
- Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of

- individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory*, 6(5), 588.
- Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological science*, 18(3), 233-239.
- Gernsbacher, M. A. (1993). Less skilled readers have less efficient suppression mechanisms. *Psychological Science*, 4(5), 294–298.
- Gernsbacher, M. A. (1997). Two decades of structure building. *Discourse Processes*, 23, 265-304.
- Gernsbacher, M. A., & Faust, M. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 245-262.
- Gernsbacher, M. A., Robertson, R. R. W., Palladino, P., & Werner, N. K. (2004). Managing Mental Representations During Narrative Comprehension. *Discourse Processes*, 37(2), 145–164.
- Gernsbacher, M. A., & Varner, K. R. (1988). *The multi-media comprehension battery*. Eugene, OR: University of Oregon, Institute of Cognitive and Decision Sciences.
- Gernsbacher, M. A., Varner, K. R., & Faust, M. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 430-445.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371-395.
- Kelemen, W., Frost, P., & Weaver, C. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition*, 28(1), 92-107.

Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163-182.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological review*, 85, 363-394.

Maki, R. H., Jonas, D., & Kallod, M. (1994). The relationship between comprehension and metacomprehension ability. *Psychonomic Bulletin & Review*, 1, 126-129.

Maki, W. S., & Maki, R. H. (2002). Multimedia comprehension skill predicts differential outcomes of web-based and lecture courses. *Journal of Experimental Psychology: Applied*, 8, 85-98.

Mayer, R. E. (2008). *Learning and instruction* (2nd Edition). Upper Saddle River, NJ: Pearson Education.

Mayer, R., & Gallini, J. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, 82, 715-726.

McDaniel, M. A., & Einstein, G. O. (1989). Material appropriate processing: A contextualist approach to reading and studying strategies. *Educational Psychology Review*, 1, 113-145.

McDaniel, M. A., Hines, R. J., & Guynn, M. J. (2002). When text difficulty benefits less-skilled readers. *Journal of Memory and Language*, 46, 544-561.

McDaniel, M. A., Hines, R. J., Waddill, P. J., & Einstein, G. O. (1994). What makes folk tales unique: Content familiarity, causal structure, scripts, or superstructures? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 169-184.

McDaniel, M. A., Howard, D., & Einstein, G. O. (2009). The read-recite-review study strategy:

- effective and portable. *Psychological Science*, 20, 516-522.
- Mason, M. (1978). From print to sound in mature readers as a function of reader ability and two forms of orthographic regularity. *Memory & Cognition*, 6, 568–581.
- Mazzoni, G., Cornoldi, C., & Marchitelli, G. (1990). Do memorability ratings affect study-time allocation? *Memory & Cognition*, 18, 196– 204.
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, 95, 109-133.
- Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science*, 5, 207-213.
- Petros, T. V., Bentz, B., Hammes, K., & Zehr, H. D. (1990). The components of text that influence reading times and recall in skilled and less skilled college readers. *Discourse Processes*, 13, 387-400.
- Schaffner, E., & Schiefele, U. (2013). The prediction of reading comprehension by cognitive and motivational factors: Does text accessibility during comprehension testing make a difference? *Learning and Individual Differences*, 26, 42-54.
- Schmalhofer, F., McDaniel, M. A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. *Discourse Processes*, 33, 105-132.
- Thiede, K., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 1024-1037.
- Thiede, K. W., Anderson, M. C. M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, 95, 66-73.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162-185.

Table 1. Mean study time, test performance, metacomprehension judgments of learning and study time allocation, and 3R recitation performance (all standard errors are in parentheses) as a function of structure building (SB) ability.

		High SB	Low SB
Study Time (Minutes)	Initial Study	8.68 (.62)	8.22 (.9)
	Recitation	3.4 (.3)	2.53 (.3)
	Review	7.95 (.85)	7.07 (.53)
	Total Study	20.02 (1.22)	17.83 (1.22)
Test Performance (Proportion correct)	Free Recall	.33 (.11) *	.19 (.10)
	Factual Multiple-Choice	.81 (.13) *	.68 (.18)
	Inference Multiple-Choice	.73 (.22) *	.45 (.25)
	Problem Solving	.35 (.13) *	.25 (.12)
Metacomprehension (Correlations)	Mean JOL	50.95 (16.64)	42.59 (24.74)
	JOL Calibration (Absolute accuracy)	.32 (.18) ^a	.30 (.27) ^a
	JOL Resolution (Relative accuracy)	.25 (.39) ^a	.45 (.38) ^a
	JOL and Recitation Correlation	.37 (.42) ^a	.39 (.37) ^a
	Study Time Allocation	- .22 (.36) ^a	- .03 (.44)
3R Recitation Performance (Proportion recited)	Old Idea Units	.75 (.03) *	.64 (.05)
	New Idea Units	.24 (.02) *	.13 (.02)

* High SB significantly outperformed Low SB, $p < .05$

^a Correlation significantly different from zero, $p < .05$