In press, *Perspectives on Psychological Science*

**Individual Differences in Structure Building: Impacts on Comprehension and Learning, Theoretical Underpinnings, and Supports for Less-Able Structure Builders**

Mark A. McDaniel

Washington University in St. Louis

Elizabeth J. Marsh

Duke University

Reshma Gouravajhala

Washington University in St. Louis

Send correspondence to:

Mark McDaniel

Department of Psychological and Brain Sciences

Washington University

1 Brookings Drive

Saint Louis, MO 63130

Email: markmcdaniel@wustl.edu

Abstract

In this article, we highlight an under-appreciated individual difference: *structure building* (Gernsbacher, 1990). Structure-building is integral to many everyday activities and involves creating coherent mental representations of conversations, texts, pictorial stories, and other events. People vary in this ability, in a way not generally captured by other better-known concepts and individual difference measures. Individuals with lower structure-building ability consistently perform worse on a range of comprehension and learning measures relative to those with higher structure-building ability, both in the laboratory and in the classroom. Their problems include a range of comprehension processes, including encoding factual content, inhibiting irrelevant information, and constructing a cohesive situation model of a text or conversation. Despite these problems, recent research is encouraging in that techniques to improve the learning outcomes for low-ability structure-builders have been identified. We argue that the accumulated research warrants the recognition of structure building as an important individual difference in cognitive functioning and that additional theoretical work is needed to understand the underpinnings of structure building deficits.

**Individual Differences in Structure Building: Impacts on Comprehension and Learning, Theoretical Underpinnings, and Supports for Less-Able Structure Builders**

One of the biggest mysteries of cognition involves understanding individual differences in performance. That is, though the field has a good understanding of the basics of processes like perceiving, remembering, and comprehending, among others, we have much less of a sense of how such abilities vary across people and how such variations affect performance on a wide range of tasks. Working memory is a major exception; it can be measured reliably and predicts individual differences in a wide range of skills and abilities (Conway et al., 2005). However, from an educational standpoint, working memory may not be an ideal target, given that arguments persist as to whether or not this is a trainable individual ability (Shipstead, Redick, & Engle, 2012) and if so whether the benefits last over time and/or transfer across tasks (Melby-Lervåg & Hulme, 2013; Melby-Lervåg, Redick, & Hulme, 2016). In this article, we highlight an individual difference that is measurable, consistently related to learning in the laboratory and authentic educational settings, and potentially trainable. We argue for the value of this under-appreciated individual difference, *structure building* (Gernsbacher, 1990; an individual difference that is not well known outside of the text-processing literature), explaining what it is, how to measure it, and its relationship to learning and comprehension (both in the lab and the classroom). We briefly explain how structure-building differs from other better-known concepts and measures, then review studies that illuminate the component processes that challenge low-ability structure-builders, followed by a discussion of the implications for interventions. We close by identifying open theoretical issues and discussing the potential of this construct for fruitful educational application.

 Structure-building is the creation of coherent mental representations of concepts, conversations, texts, and other events. Structure-building is integral to many everyday activities, such as when one reads a novel or story (Gernsbacher, Varner, & Faust, 1990; Graesser, 1975; Stine-Morrow et al., 2004); watches a picture story (e.g., Baggot, 1975), listens to a lecture (Bui & McDaniel, 2015), attends a work meeting, or engages in conversation. In these situations, the goal is typically not to memorize what one reads, sees or hears, but rather to extract the gist and understand the “who/what/when/where/why” of a particular narrative or event. Critically, Gernsbacher (1990) proposed that there are fundamental individual differences in people’s structure-building ability. She argued that some people—high-ability structure builders—routinely extract coherent mental structures of the narrative events they encounter (texts, conversations, pictorial stories). In contrast, other people—low-ability structure builders—routinely extract fragmented and less cohesive mental structures of the narrative events they encounter.

 Despite the potential theoretical and practical importance of this proposed individual difference in structure building, it has received relatively little empirical attention outside of the text-processing literature since it appeared nearly three decades ago (e.g., Gernsbacher et al., 1990). In the past 15 years, however, we have embarked on a series of studies to (a) establish the range of materials and learning environments in which individual differences in structure building have impact, (b) illuminate the representational and comprehension shortcomings of low-ability structure builders, and (c) develop interventions to improve comprehension and learning for less able structure builders. In this article, we draw on this research to highlight and integrate the progress on these significant issues, and argue that structure-building is a critically important individual difference.

A**ssessing Structure Building**

Structure-building ability is assessed with the Multi-Media Comprehension Battery (MMCB; developed by Gernsbacher & Varner, 1988), and this assessment is highly reliable, alpha = .987 (Gernsbacher et al., 1990). Capturing the theoretical premise that structure building ability is general across linguistic and nonlinguistic media, the Multi-Media Comprehension Battery presents simple stories in print, aurally and in pictures. Participants proceed through each of four stories at their own pace; like fables, each story conveys a moral. In the original version, participants are presented six stories (two self-paced written stories, two experimenter-paced aural stories, and two self-paced pictorial stories; see Gernsbacher & Faust, 1991, Appendix for more details about the battery and its administration1). Participants then answer short-answer questions (scored on a three-point scale) that focus on both specific details and broader concepts from the stories. Test scores across the different media are highly correlated (ranging from .72 to .92), and a principle-component analysis revealed that one factor underlies performance in all modalities (Gernsbacher et al., 1990, who also noted correspondences between listening and reading using other assessments). Accordingly, subsequent studies typically assessed structure building with one modality (either printed text or pictorial narratives; e.g., Maki & Maki, 2002; McDaniel, Hines, & Guynn, 2002).

In particular, most of these studies assessed structure building using a modified version of the Multi-Media Comprehension Battery, wherein participants are presented with four stories, shown line-by-line with one paragraph per slide, in a self-paced fashion. Once a participant has finished each story, the story is removed from view and participants answer 12 multiple- choice questions about it before proceeding to the next story. The resultant scores can range from 0 to 48. Some researchers have treated structure building ability continuously, whereas others have pre-selected subjects who score below 32 or above 36 to indicate low-ability and high-ability structure builders, respectively (Callendar & McDaniel, 2007; Martin et al., 2016; McDaniel et al., 2002). (A natural question is to what extent structure-building ability, as assessed by the Multi-Media Comprehension Battery, is concordant with reading ability as assessed by standard indices such as Nelson-Denny. We address that question in a later section; to preview, the overlap is modest.)

 Figure 1 shows how several samples of individuals vary in their scores on the Multi-Media Comprehension Battery (the text version). This variability is evident in a sample from a mid-size four-year public university (Washburn University) and even within relatively homogenous samples from highly selective private universities (Washington University in St. Louis, Duke University). The pattern is similar in a sample of high school students from excellent school districts in North Carolina as well in the general population (as measured in an MTurk sample). Thus, structure-building is an ability that varies widely.

**Structure Building Matters**

 To highlight the strong relationship between structure building and learning we begin with an illustrative example (conducted by Lin, McDaniel, & Miyatsu, 2018). In one experiment (*n* = 134), students read and studied a passage on either biological anthropology or geology, completed the written version of the Multi-Media Comprehension Battery , and then answered multiple choice and short answer questions about the studied passage. The multiple choice questions focused on details from the passage. For instance for the geology passage, a multiple choice question was:

*What can a paleoanthropologist assess using carbon-dating techniques?*

* 1. *Shell, charcoal, and tuff*
	2. *Charcoal, shell, and peat*
	3. *Ash, peat, and bone*
	4. *Peat, wood, and ash*
	5. *Bone, tuff, and charcoal*

The short answer questions focused on students’ understanding and integration of concepts. An example of this type of question (for the geology passage) is:

*A scientist used fluorine absorption dating to discover a certain material’s relative age compared to all the other local fossils. The scientist later used another dating technique to find the absolute age of this material. The material was determined to be 1,500 years old. a) What absolute dating technique did the scientist use? Explain. b) What could the material be? Explain.*

 Students’ structure building ability (as captured in their Multi-Media Comprehension Battery scores) predicted their performance on both the multiple choice (η2 = .076; medium sized effect) and the short answer (η2 = .099; medium sized effect) questions. This relationship between learners’ structure building ability and test performance held regardless of how students studied the text: without specific instruction (a free-study condition), with experimenter-provided flashcards, or with student-generated flashcards. These findings underscore the importance of structure building ability when learning authentic classroom materials and with a variety of study strategies.

 Low-ability structure building is associated with deficits in learning and retention across a wide range of topics (see Table 1 for a comprehensive summary): astronomy (Arnold, Drew, McDaniel, & Marsh, 2019; Arnold et al., 2017), biology (Gouravajhala, 2018), biological anthropology and geology (Lin et al., 2018; McDaniel et al., 2002, Experiment 2b), biological and abnormal psychology (Callender & McDaniel, 2009, Experiment 4), and mechanics (brakes and pumps; Bui & McDaniel, 2015; Martin, Nguyen, & McDaniel, 2016; St. Hilaire, Carpenter, & Jennings, 2019). The effects are not limited to short texts, but also occur when the goal is learning entire textbook chapters (Psychology: Callender & McDaniel, 2007; Callender & McDaniel, 2009, Experiment 4; and Biology: Gouravajhala, 2018). A strong narrative structure is clearly not a pre-requisite for revealing learning and retention differences as a function of structure-building ability. Multi-Media Comprehension Battery scores predict learning of textbook selections that Coh-Metrix analyses (Graesser, McNamara, Louwerse, & Cai, 2004) confirm were extremely low in narrativity (10th and 6th percentile, respectively for biological anthropology and geology; Lin et al.).

 Critically, structure-building also predicts learning and comprehension across a range of procedural variations. Though most experiments measure comprehension of print materials, structure-building also predicts comprehension when students listen to a “book-on-tape” passage while taking notes (Bui & McDaniel, 2015) and when texts are degraded (scrambled or missing letters; McDaniel et al., 2002, Experiment 2b). Structure-building scores predict long-term learning, as measured days later in laboratory experiments (Arnold et al., 2019; Arnold et al., 2017; Callender & McDaniel, 2009) or weeks later in a classroom setting (Maki & Maki, 2002). It does not matter whether structure-building is measured with an extreme-groups design (Arnold et al., 2019, Experiment 1; Callender & McDaniel; Martin et al., 2016) or treated continuously (Arnold et al., 2019, Experiment 2; Arnold et al., 2017; Bui & McDaniel; Lin et al., 2018); it still predicts comprehension. In sum, up to a dozen laboratory studies (or experiments) with a variety of materials and study (or reading) procedures have examined the extent to which individual differences in structure building are associated with learning and memory. The results are clear cut and consistent in study after study: Lower-ability structure builders learn and remember less than do more able structure builders (see Table 1). These deficits are robust and often yield large effect sizes.

 A critical test of structure-building is whether it predicts performance in the classroom, given that much learning requires the integration of material from multiple experiences, across modalities. That is, the same topic is often covered in a lecture, a textbook assignment and in a recitation section. Learning should not involve creating separate representations for each of these experiences. Rather, coherent and complete understanding would entail extracting an integrated mental representation (structure) across sources (and modalities) that also captures the somewhat non-overlapping sub-content separately gleaned from each experience. Moreover, the successful learner must build his or her mental structures over time (on the order of days or weeks), not just across different sources. Accordingly, one might expect that skill in structure building would be a crucial component of successful classroom learning.

Two studies found the predicted positive relationship between structure-building ability and classroom success: Multi-Media Comprehension Battery scores predicted exam performance in two sections of an introductory psychology course (Maki & Maki, 2002) and final course grades in undergraduate introductory psychology and biology classes (Arnold, Daniel, Jensen, McDaniel, & Marsh, 2016). In the latter study, researchers collected high school GPAs and SAT scores (from a subset of students in the psychology course), as well as scores on the Biology Concept Inventory and Lawson’s Classroom Test for Scientific Reasoning (from a subset of students in the biology course). Critically, differences in structure building ability predicted final grades in both introductory psychology and biology even after taking the other standard predictors into account – an impressive feat given that high school GPA and SAT scores are “gold standard” predictors of success in college. Therefore, these studies show that individual differences in structure building predict learning outcomes in university courses, at least courses at the introductory level. The implication is that structure building is a general skill that holds importance for learning and understanding of complex material in authentic learning environments.

**What Structure-Building is Not**

Many factors predict learning and comprehension, and accordingly, in this section we briefly reassure the reader that structure building is not a familiar ability with a new name. Our goal here is not to comprehensively cover all of the research on these other measures and concepts, but rather to underscore that structure-building captures something different than these more familiar measures.

First,it is important to note that structure-building is *not the same as story comprehension*, a concern that might arise given that both the structure-building assessment (Multi-Media Comprehension Battery; MMCB) and much of the initial elegant work on structure-building measured understanding of simple stories (e.g., Gernsbacher et al., 1990). To be clear, the concern is not that the *general* goals and processes of comprehension are unique to stories. Rather, on many views, stories (and more generally event-based texts) afford specific and potent processes2 that expository texts do not (e.g., two influential comprehension models are based on theoretical components that apply only to event-based texts: the Causal Network Model [Trabasso, van den Broek, & Suh, 1989] and the Event-Indexing model [Zwaan, Langston, & Graesser, 1995; Zwaan & Radvansky, 1998]; see McNamara & Magliano, 2009, for an excellent overview). In particular, an extensive theoretical and empirical literature has identified unique aspects of stories’ contents and structure that facilitate the construction of well-organized representations (e.g., Fletcher & Bloom, 1988; Graesser, Woll, Kowalski, & Smith, 1980; Mandler & Johnson, 1977; Schank & Abelson, 1977; Thorndyke, 1977; Trabasso & van den Broek, 1985; van Dijk, 1980; van Dijk & Kintsch, 1983). Thus, one concern is that structure-building is primarily capturing participants’ ability to utilize and build upon story-specific features (Footnote 2 provides amplification).

 Were it the case that individual differences in structure building primarily reflected unique aspects of story comprehension (e.g., noticing and exploiting causal relations; acquiring or activating story grammars) this individual difference would not be less interesting, but its theoretical reach would be more limited. The evidence reviewed in the previous section, however, makes clear that structure-building ability predicts comprehension and learning of expository text that is not in story format. For example, after reading technical passages, less-able structure builders recall less content, correctly answer fewer multiple–choice questions (regardless of whether they test retention or require inferences), and solve fewer short-answer application questions that require fixing or diagnosing problems, relative to more-able structure builders (Martin et al., 2016; St. Hilaire et al., 2019; see Table 1 for many more examples).

Second, structure building ability, as assessed by the Multi-Media Comprehension Battery, is *not synonymous with general reading ability* as indexed by existing standardized indices such as the Nelson-Denny Reading Test (Brown, Fischo, & Hanna, 1993). Certainly, reading comprehension must involve structure-building, but the Multi-Media Comprehension Battery and Nelson-Denny Reading Test scores should not be highly correlated if these tests emphasize somewhat different skills. In line with this expectation, the two measures are moderately correlated (*r* = .46, Maki et al., 1994; *r* = .28, unpublished; personal communication, Callender, April 5, 2018); that is, a standard reading ability measure captured no more than 21% of the variance in the Multi-Media Comprehension Battery (Maki et al.).

 Further evidence for the difference between the two measures comes from findings that extra time benefits struggling readers on standard reading tests such as the Nelson-Denny Reading Test but not on the structure-building measure (Multi-Media Comprehension Battery). That is, most reading measures, in contrast to the Multi-Media Comprehension Battery, are timed tests and accordingly emphasize the efficiency of low-level reading processes such as word-decoding (for Nelson-Denny Reading Test see Mason, 1978; Petros, Bentz, Hammes, & Zehr, 1990; see Callendar & McDaniel, 2007, for a broader discussion of this and other differences across reading measures). Struggling readers improve and comprehend more when they are given additional time (Baldwin, Murfin, Ross, & Seidel, 1989; Jensen, 1998), and consequently these readers later recall as much of both narrative and expository passages as do readers who score high on the Nelson-Denny Reading Test (McDaniel et al., 2002, Experiments 1a and 1b). In contrast, unlimited reading time does not alleviate lowability structure builders’ problems, as they still recall about 40% less of the same narratives and expository passages than that observed for high-ability structure builders (McDaniel et al., Experiments 2a and 2b). Because the way a reader recalls a text is theorized to reflect how cohesive their representations are (Kintsch & van Dijk, 1978), these patterns suggest that readers who score low on the Nelson-Denny Reading Test (but not low-ability structure builders) are capable of extracting coherent and complete representations, albeit requiring more time to do so.

**Relating Structure-Building to Other Known Predictors of Learning**

***ACT and SAT.*** It is not surprising that Multi-Media Comprehension Battery scores correlate with selected standardized test scores given that reading comprehension is required for success in school (Arnold et al., 2016). However, these correlations are relatively low: Multi-Media Comprehension Battery scores correlate weakly with **SAT verbal scores** (*r* = .16) and not at all with **ACT English scores** (*r* = .01; Gouravajhala, 2018) and with **SAT-writing scores** (r = .028; Arnold et al., 2017). More critically, Multi-Media Comprehension Battery scores explain additional variance in course grades beyond that predicted by high school GPA and SAT scores (two very significant predictors; Arnold et al., 2016). Further, Multi-Media Comprehension Battery explains variance in performance on multiple-choice and problem solving tests on an astronomy text, but SAT writing scores do not (Arnold et al., 2017).

 ***Working Memory.*** One might assume that effective structure building is highly dependent on working memory capacity. The idea is that with greater working memory capacity, more concepts (propositions) can be co-activated (Kintsch & van Dijk, 1978) and more resources are available to construct connections among those concepts **(**cf. Wiley, Jarosz, Cushen, & Colflesh, 2011), or both. Existing evidence suggests, however, that working memory and structure building are moderately related at best. Consider data from Arnold et al. (2017); this paper reported that when Multi-Media Comprehension Battery, working memory, and SAT scores were simultaneously entered into a regression, only Multi-Media Comprehension Battery scores emerged as a significant predictor of learning. Panel A of Figure 2 drills down into those data3, showing the relationship between standardized Multi-Media Comprehension Battery scores and an averaged working memory score (collapsing over standardized Rspan and Ospan scores4). There was essentially no correlation between the standardized Multi-Media Comprehension Battery scores and the working memory composite scores (*r* = .14) In another study with over one hundred participants (Gouravajhala, 20185), we again found only a small association between working memory (as measured by an average score of standardized OSpan and RSpan scores) and standardized Multi-Media Comprehension Battery scores (*r* = .15; see panel B of Figure 2). Thus, Multi-Media Comprehension Battery scores are weakly related to performance on working memory tests -- but should not be considered a stand in for working memory.

**Processes Underlying Structure-Building Deficits**

 Given that structure-building ability predicts comprehension and learning across a variety of materials and in educational settings it is important to understand *why* low-ability structure builders struggle. To do so, we draw on the above findings as well as additional studies to develop a more complete and nuanced description of the component processes impaired in low-ability structure builders. In these sections, our discussions will primarily focus on text processing, reflecting the available literature. In doing so we will occasionally use text-specific terms, but it is important to note that the findings and theoretical implications discussed should extend to the comprehension of non-linguistic inputs. Such an extension would be in line with the central tenet of the structure-building framework, but direct confirmation of that assumption awaits future empirical work. Here, we examine the possibility of deficits in (1) encoding and remembering incoming information, (2) identifying the main points of a text, (3) activating and applying prior knowledge, (4) inhibiting irrelevant knowledge, and (5) connecting ideas across the text.

**Memory Deficits for Factual Content**

At least some information from the to-be-comprehended source (be it a text, video, or other type of event) must be stored in memory to provide the foundation for one’s mental model. In the original structure-building framework, this basic information was represented as memory nodes (Gernsbacher, 1990). However, because a prominent text processing model (in a following section, we motivate and amplify on this model) more frequently formalizes the basic elements of meaning as *propositions* (Kintsch, 1974; Kintsch & van Dijk, 1978; Kintsch, 1988; Schmalhofer, McDaniel, & Keefe, 2002), we adopt that term here. To determine whether a text’s propositions are stored in memory, the standard approach is to test recognition of explicitly stated propositions from the text (e.g., Sachs, 1967; Masson & Sala, 1978; Mayer & Gallini, 1990; Perrig & Kintsch, 1985). The idea is that propositions will be recognized if they are stored in memory, given that recognition tests minimize retrieval demands (as opposed to free recall; Einstein & Hunt, 1980; McDaniel & Masson, 1977).

 Briefly, the data suggest that low-ability structure-builders are less likely to recognize verbatim propositions from texts. For example, in one study participants were tested on their memories for the contents of relatively long texts (e.g., on the order of 2,600-3,200 words – that is, 16 pages or longer) that were adapted from psychology textbook chapters. For example, readers were asked to answer the question “Hairstyles, dress, and social customs are examples of: (a) culture, (b) groupthink, (c) attractiveness, (d) social conformity” to measure whether this idea unit about culture was stored in memory. Regardless of whether the chapter was read once or twice, performance on the multiple-choice questions was positively associated with higher structure-building ability (Callender & McDaniel, 2009, Experiment 4, continuous variable analysis). A study using an extreme groups design provides a sense of how robust this encoding deficit is: Low-ability structure builders answered only 52% of questions correctly as compared to 80% correct for high-ability structure builders (Callender & McDaniel, 2007). Low-ability structure builders show similar deficits in recognizing verbatim text content even when the tested passages are much shorter. For example, when asked to recognize a concept from each of eight paragraphs from a mechanics text (“Which of the following 4 options is not a component of a disk brake”), low-ability structure builders were significantly impaired (M = .68) relative to high-ability structure builders (M = .81), a large effect size (Martin et al., 2016). Converging results come from studies that treat structure-building as a continuous variable; structure-building is negatively related to recognition of the content of relatively short texts (800-900 words) on geology and biological anthropology (Lin et al., 2018) and astronomy (Arnold et al., 2017)

 In sum, less-able structure builders consistently recognize less factual information from texts relative to more-able structure builders, regardless of text length or topic. This finding highlights the need to provide learning supports to less-able structure builders (a point we return to later) and also signals a critical theoretical implication: Less-able structure builders do not sufficiently encode the propositional information stated in the text. That is, less-able structure builders struggle to store the propositions that are the foundation for building mental structures. A memory deficit for basic to-be-learned information could underlie many of the learning deficits documented for less-able structure builders, instead of or in addition to problems with the coherence of the structures built (cf. Gernsbacher, 1990; Gernsbacher et al., 1991). At the very least, it appears that less-able structure builders are plagued not only by deficits in constructing coherent mental structures but also in encoding incoming information.

**Deficits in Identifying Main Points**

Encoding propositions is a necessary but insufficient step towards comprehending and remembering a text. Construction of coherent mental structures presumably depends, at least in part, on being able to identify the main points of a text and to relate sub-ordinate material to these main points (Kintsch & van Dijk, 1978; Lorch, Lorch, & Inman, 1993). One possibility is that lower-ability structure builders struggle to distinguish main ideas from less important ones, as compared to higher-ability structure builders. This hypothesis is not highlighted in the existing structure-building model (e.g., Gernsbacher et al., 1990, Gernsbacher & Faust, 1991), and as a consequence has received little attention – but is an intriguing idea given the larger literature about summarization and learning (see Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013, for a brief review). Summarization entails many processes, including but not limited to identifying the main points of a text – and summaries that are successful in capturing important ideas are associated with more learning (Dyer, Riley, & Yekovich, 1979) and better mental models (Garner, 1982).

To our knowledge, there is only one study investigating whether there is a relationship between a reader’s ability to identify important ideas and his/her structure-building performance (Gouravajhala, 2018). Participants read a biology textbook chapter, “Evolution and Natural Selection,” (Phelan, 2009); two days later readers received a list of 39 ideas from the chapter and were instructed to mark the ones that they considered to be the most important ideas (no limits were placed on the number of ideas that could be marked). The list contained both important ideas (e.g., “Fitness depends on an organism’s reproductive success compared with other organisms in the population.”) and unimportant ideas (“The name of Darwin’s ship was the *HMS Beagle.*”). Biology instructors independently identified ten ideas as the most important ones; intriguingly, these ideas were more likely to be endorsed by participants higher in structure-building ability, as assessed using a discriminability index (d’; standardized important points minus standardized unimportant points). At the same time, structure-building did not significantly predict the total number of ideas endorsed. That is, low-ability structure builders were not endorsing more ideas as important; they were less likely to identify the actual main ideas as being important. Importantly, participants were neither enrolled in nor had completed a biology or evolution course, and further, these patterns remained after factoring in participants’ self-reported ratings of prior familiarity with the topics, indicating that the results were not simply a by-product of some individuals having more knowledge of the material. Of course, a single study is not conclusive (e.g., the problem identifying main ideas could indicate a problem with one’s mental model rather than reflecting on-line comprehension processes), but the study does suggest that it is worth investigating the hypothesis that structure-building problems may in part derive from a difficulty in identifying main ideas.

**Deficits in Activating and Applying Prior Knowledge**

Knowledge is required for structure-building (as described in Kintsch's, 1988, construction-integration model of text comprehension); the reader must activate prior knowledge and inhibit irrelevant information as she creates a representation of a text. Without adequate prior knowledge, irrelevant information is not adequately deactivated (McNamara & McDaniel, 2004; see McNamara & Magliano, 2009, for further discussion), thereby potentially stimulating formation of many fragmented structures (Gernsbacher, 1990). Accordingly, different degrees of prior knowledge may contribute to individual differences in structure building (McNamara, 1997; McNamara & Magliano, 2009).

Also, it may not be enough to have knowledge *stored* in memory; knowledge must be *activated* to support the construction of a coherent mental representation. Gouravajhala, Nebel, Umanath, and McDaniel (in preparation)6 examined these ideas in a sample of 105 students from two universities (one more selective than the other, with the idea that the two groups would differ in how much they knew). Prior knowledge was quantified with two measures: a vocabulary test and a general knowledge assessment that probed knowledge of geographical, historical, literary, and scientific facts. Knowledge activation was assessed via an error detection task in which students reading stories were asked to press a key when they read anything that contradicted the true state of the world (Marsh & Fazio, 2006; Umanath & Marsh, 2012). Critically, the stories made both neutral references to the world (e.g., “paddling around the largest ocean”) as well as incorrect ones (e.g., “paddling around the largest ocean, the Atlantic”). Only the general knowledge measures significantly correlated with structure building ability (Multi-Media Comprehension Battery scores); *r* = .28 for general knowledge and *r =* .38 for vocabulary. The knowledge-activation index (performance on error detection) correlated slightly but not significantly with structure building ability (.18, *p <* .08). These patterns are in line with the idea that structure-building ability may rely somewhat on available prior knowledge.

**Inhibitory Deficits**

As noted in the last section, learners must also deactivate irrelevant information, and crucial for present purposes is that several studies show that less-able structure builders are slower to inhibit the irrelevant meanings of words. For instance, in one task participants read sentences like “He dug with a spade” and decided whether a probe word fit in the sentence. Critically, the probe word tapped a different meaning that the one implied in the sentence (e.g., “ACE”). Low-ability structure builders were slower to reject the related but irrelevant probe words, as compared to more able structure builders (Gernsbacher et al., 1990, Experiment 4; see also Gernsbacher & Faust, 1991, for a similar task and findings). In a converging cross-modality task, participants saw pictures with words superimposed on them (e.g., the word “RAIN” might be superimposed on a line drawing of a hand) and were instructed to ignore the words; each was followed by a picture slide and participants judged whether the two pictures were related. Critically, on some trials the to-be-judged picture was related to the to-be-ignored word (e.g., a picture of an umbrella). Low-ability structure builders took longer to reject this picture and label it as unrelated to the prior picture (the hand; Gernsbacher & Faust, 1991, Experiment 3). A similar effect was found when participants were instructed to ignore the pictures on the context slide and instead judge whether a word probe was related to the word super-imposed on the picture. These and similar results suggest that low-ability structure builders are deficient in inhibiting irrelevant meanings of incoming information, which is problematic because successful comprehension requires that “irrelevant or inappropriate information must not affect ongoing processes; it must be efficiently suppressed” (Gernsbacher & Faust, 1991, p. 247).

 To our knowledge, published research has generally not built on this rich set of initial findings. However, a recent study pursued the idea that compromised structure building ability is related to inhibitory deficits. In a continuous-measure design, Gouravajhala et al. (in prep) assessed three indices of inhibitory control: the irrelevant-word meaning task from Gernsbacher’s paradigm (Gernsbacher et al., 1990; Gernsbacher & Faust, 1991), Stroop interference, and flanker interference. Regression analyses indicated that performance on the flanker task, but not the other tasks7, was associated with structure-building ability (increased flanker interference was associated with lower structure-building scores), and this relation persisted when scores on measures of prior knowledge and indices of connection making (described below) were taken into account.

**Making Connections**

 Finally, one might posit that structure building is basically the skill of connecting ideas and building relations across patterns of information. To test this hypothesis, Gouravajhala et al. (in prep) administered the Multi-Media Comprehension Battery and assessed three measures of subjects’ abilities to make connections across concepts. In the first, an analogies task (ETS, 2015), subjects discerned the critical relation in each of a series of base pairs (e.g., *rare : common*) and then selected the pair (out of four multiple choice options) reflecting the same kind of relation (e.g., *old : new*). Prior research has suggested that completing analogies can benefit comprehension of biological and statistical concepts, as well as serve as a building block for cognition in development (Carey & Gelman, 1991; Goswami, 1996; Newby, Ertmer, & Stepich, 1995; Quilici & Mayer, 1996). The second measure was the Raven’s Progressive Matrices Test (Arthur & Day, 1994), which requires subjects to decide which of several choices (patterns) complete a matrix conveying a relational pattern. A well-validated measure of fluid intelligence and cognitive functioning, the Raven’s Progressive Matrices Test has also been shown to be highly correlated with verbal learning tasks (Tamez, Myerson, & Hale, 2008; Williams & Pearlberg, 2006). Finally, the third measure was the Remote Associates task (Mednick, 1962); participants read three-word triads and generated a linking word for each triad (e.g., “sweet” is the best response to “tooth”, “potato”, and “heart”).

Performance on the Raven’s (*r* = .38) and the Remote Associates (*r* =.43) tasks (but not the analogies task) significantly correlated with Multi-Media Comprehension Battery scores (written version). Multiple regression models showed that prior knowledge predicted Multi-Media Comprehension Battery scores (as indexed by vocabulary but not the general knowledge index), but this relationship disappeared once scores on the Ravens and the Remote Associates tasks were entered into the model (both of which were significant predictors). These patterns show that the ability to draw connections and discover relations is an important component of structure-building. Of course, both of these tasks (Raven’s and remote associates) are also correlated with prior knowledge measures, and thus prior knowledge may contribute to several levels of structure building (i.e., deactivating irrelevant information, drawing connections) (cf. McNamara & Magliano, 2009). These interpretations clearly await more in-depth investigations.

**Theoretical Implications**

Identifying structure-building as a predictive individual difference raises the question of *how* structure-building ability relates to the construction of representational structures. Addressing this issue is complicated by the long-standing and vigorous theoretical debate regarding the nature of the representations that are constructed during comprehension. At present, there is a lack of consensus and continuing debate (McNamara & Magliano, 2009), and it is beyond the scope of this article to consider all of the various competing theoretical positions.

For an initial theoretical treatment, we situated our analysis of individual differences within Kintsch’s influential model of text processing (Kintsch, 1988; Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983). A number of reasons compelled our choice of this particular model, the most prominent being: (1) The model specifies several parallel representational structures (as is generally embraced by the field; McNamara & Magliano, 2009; see also Zwaan & Radvansky, 1998), which allows the possibility for a more nuanced specification of the representational outcomes of individual differences in structure building (than models embracing one representational level). (2) The model has stimulated a literature that aligns particular behavioral measures with each representational structure. (3) The model is an influential benchmark theory that has provided the foundation for several competing models, thereby providing some commonality across a range of specific theoretical perspectives (though not all). (4) The Kintsch model can be instantiated in a computational model (see e.g., Schmalhofer, McDaniel, & Keefe, 2002), providing a formal basis for (future) work attempting to characterize how individual differences in structure-building affect processing and representations (McNamara, 1997). Though the Kintsch model focuses on text comprehension, this choice of model does not limit the reach of structure-building to reading. That is, the seminal structure-building framework was explicitly intended to encompass processes general to many modalities; we use the Kintsch framework because it is the most developed, not because it involves reading.

In the Kintsch model, understanding involves constructing two parallel mental structures. The first is a coherent, interrelated network of the propositions stated in the text as well as those inferred from the text (the *text-based* representation). The second is a mental representation of the larger event or action described in the text. This representation might capture the overarching activity described (e.g., laundry rather than the individual steps of the process; Bransford & Johnson, 1972), a character’s path through a physical setting (e.g., Bower & Morrow, 1990; Morrow, Greenspan, & Bower, 1987), or how interacting mechanical parts support a device’s function (e.g., Mayer & Gallini, 1990). This representational structure has been termed a *mental model,* Johnson-Laird, 1981; or *situation model* (Kintsch, 1988; Perrig & Kintsch, 1985; Schmalhofer et al., 2002). Theoretically, individual differences in structure building could affect the coherence of the text-based representation or the quality of the situation model, or both. We examine these possibilities next.

**Text-Based Level Structures**

Free recall is typically used to gauge the coherence of text-based representations (see Kintsch & van Dijk, 1978, for an example of this approach). The idea is that a tightly interconnected network of propositions provides an organizational structure for recalling information; the more coherent the text-based representation, the more that will be recalled. Accordingly, if structure building ability is related to the construction of a coherent text-based level representation, it should also be associated with greater recall. Below we review the evidence supporting this prediction.

Across studies and a number of procedural variations, higher structure-building ability is associated with greater recall of a text. Figure 3, for example, shows that this pattern holds when participants initially recall a technical mechanics text (with high-ability structure builders recalling almost 40% more) and even after receiving a chance to restudy the passage before recalling again (Martin et al., 2016). As would be expected, final recall was higher than initial free recall, but if anything, the low-ability structure builders were at even more of a disadvantage when recalling for the second time. Arnold et al. (2017) found similar results when participants were instructed to either recall or write an essay about an astronomy text. Participants’ written protocols (free recall or essays) were scored for the number of content words (nouns, action verbs, and adjectives from the text) included. Structure building ability was significantly related to number of content words retrieved (semipartial r = .51), whereas neither working memory nor writing ability were. Clearly, these two studies must be considered preliminary; firm conclusions await additional research, ideally using a range of texts and converging measures to gauge the coherence of the text-based representations constructed as a function of structure-building ability.

Several of the processes described earlier may explain why lower-ability structure builders have less coherent text-based representations. For example, insufficient encoding is likely involved, as it stands to reason that connecting ideas across an event requires one to remember the preceding parts of the event (be it a text, lecture, film, picture-story). Similarly, if a reader fails to identify the main points, the coherence of the text-based representation arguably could be compromised. A key process in constructing a coherent text-based representation is maintaining activation of important points (propositions) in working memory, because the important points have a higher likelihood of argument overlap with incoming propositions than do subordinate points. (Failure to establish argument overlap between newly input propositions and those held in working memory penalizes free recall of texts, Britton & Gulgoz, 1991, presumably because the coherence of the text-based representation suffers.) If the main points are not identified, it is less likely they will be kept active in working memory to anchor the construction of a coherent text-based representation. Inhibitory deficits might also impair the coherence of a reader’s text-base. Less facile inhibition of irrelevant meanings of words (reviewed in the previous section) could produce inappropriate or failed argument overlap. Beyond text comprehension, inhibiting irrelevant pictorial material, which is associated with less-able structure builders (Gernsbacher & Faust, 1991), could be important in structuring pictorial events.

 Finally, it is less clear how differences in prior knowledge and making connections would contribute to problems with the text model; knowledge is typically considered more essential to construction of situation-model representations (Kintsch, 1988; McNamara & Magliano, 2009). Making connections is arguably a strategic, skilled process, and constructing causal connections in stories is likely one such strategic process. However, for expository texts (as used in nearly all studies of individual differences in structure building and text recall; see Table 1) making connections involves noticing argument overlap, which is presumably a straightforward and non-complex process (essentially a matching process) that most, if not all, comprehenders presumably accomplish.

**Situation Model Representations**

Coherent mental/situation models theoretically represent an overarching understanding of the referent of the text or situation— for example, the layout of a town or how the components of a pump work together. These representations—the constructed mental models—allow the comprehender to reason and make inferences. For instance, a mental model of a town allows one to infer where two buildings are relative to one another, even if this information was never explicitly learned (Perrig & Kintsch, 1985). Similarly, a coherent mental model of a brake allows one to trouble-shoot the problem when a brake fails (Mayer & Gallini, 1990). We note that a rich and active body of work concerns the representational nature of mental models. Theoretical positions range from representations with static and dynamic dimensions (e.g., Copeland, Magliano, & Radvansky, 2006), multilayered structures (e.g., temporal, causal, and spatial layers; Zwaan et al., 1995), and representations embodied in the motor and perceptual system (e.g., Barsalou, 1999, 2009; Glenberg, 1997). Even the assumption that the representations are “stored” in memory has been extended; a mental model may be simulated in real-time and thus be influenced by the current context (Barsalou, 2009). However, a common theme is that accurate mental models represent a coherent understanding of the event (text, film, picture-story) from which accurate inferencing, reasoning, prediction, and problem solving can be supported. Thus, our analysis below is agnostic regarding the specifics of the various theoretical positions. We believe that it would be premature to take a stance in light of the absence of agreement concerning this construct and the surprisingly limited work on this issue with regard to structure building. Our approach is to focus on outcome measures that are used to generally characterize the quality and coherence of the mental model.

The ability to draw inferences is a common way to quantify the quality of a mental model (e.g., see Mayer & Gallini, 1990; Perrig & Kintsch, 1985). For example, one’s mental model of brakes could be probed with the multiple-choice question, “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?” (Mayer & Gallini, 1990). In the same vein, short-answer questions that require explanation or problem-solving also probe whether a reader has a coherent mental model (e.g., Mayer & Gallini, 1990). For both types of questions, structure-building ability is related to performance, suggesting that the quality of one’s mental model varies as a function of individual differences in structure-building. For example, less-able structure builders were substantially less accurate (M = .45) when making inferences about the mechanics of brakes and pumps than were more-able structure builders (M = .73), a difference that represented a large effect size (d = 1.16) (Martin et al., 2016). The same pattern held with problem solving questions (d = .80) and when the passages were presented aurally (as opposed to read) (Bui & McDaniel, 2015). Similar results were obtained with astronomy passages as well (Arnold et al., 2019). Combined, these findings provide much needed support for the theoretical premise that low-ability structure builders construct fragmented mental representations of connected discourse.

 Whereas past work treats inferences as an indicator of mental-model quality, a formal modeling technique (Pathfinder modeling) provides an externalization of learners’ mental models (Schvaneveldt, 1990). Pathfinder modeling calculates the semantic proximity of concepts in a text, based on relatedness ratings provided by experts and participants. (See Britton & Gulgoz, 1991, for an application that successfully illuminated the structure of the mental representations derived from less-well written and better-written versions of the same didactic text.) Gouravajhala (2018) used this technique to provide converging evidence that structure-building scores predict differences in the structures of the mental representations. Her participants read part of a biology textbook chapter (Phelan, 2009), and then provided relatedness ratings for all pair-wise combinations of ten important terms from the reading (e.g. “heritability” and “allele frequency”). Each pair was rated on a 5-point Likert scale anchored with 1 “*highly unrelated”* and 5 “*highly related*”. Two PhD. Biology instructors completed the same task. Pathfinder was then used to determine the proximity of concepts based on relatedness ratings, and create a proximity matrix for each participant and expert based on their ratings (following common practice, the experts’ matrices were combined into an average expert matrix). Each matrix was then reflected in a network structure in which each node represents a main term and the lines reflect connections between those terms. Physically closer nodes are more similar in the semantic space, such that, for example, concepts given a “highly related” rating (5) appear closer together in the network structure than those given a rating of 3 (Goldsmith, Johnson, & Acton, 1991). See Figure 4 for example networks.

 The accuracy of the students’ representational structures was quantified by comparing them to the averaged expert network; crucially, accuracy was significantly related to structure-building ability. In other words, students with lower structure-building ability created mental models that aligned less well with the expert representations. For instance, the experts’ network relates the key terms “mutation”, “individual”, and “mutagens” in a cluster, with that cluster in turn linked to “heritability” and “population” through the concept “natural selection” (see right panel in Figure 4). In contrast, the sample model shown in the middle panel of Figure 4 fails Figure 4 to directly relate “mutation” to “individual” and “mutagen” and also fails to cluster them as a structure. By sharp contrast, the sample model from a high-ability structure builder (left panel, Figure 4) overlaps tightly with that the expert model: “mutation”, “individual”, and “mutagens” are a related cluster that links to “natural selection,” which itself is related to “heritability” and “population.”

 These findings are striking for two reasons. First, it is the first report directly revealing differences in the quality of mental models as a function of structure-building ability. Second, the implications are exciting, because the particular index used (measuring the similarity of a representation to that of experts) is one that predicts an extremely important learning outcome in other studies: the degree to which students master content in academic courses (statistics, Goldsmith et al., 1991; chemistry, Neiles, Todd, & Bunce, 2016; physics, Stoen, McDaniel, Fry, Hynes, & Cahill, 2020). Thus, the results provide initial evidence that low-ability structure builders’ deficiencies in constructing coherent, accurate representations may significantly undermine their ability to adequately grasp academic content conveyed in course lectures (cf. Bui & McDaniel, 2015) and textbooks.

 There are two clear candidate processes for why less able structure builders might struggle to construct coherent, rich mental models. First, as mentioned in the preceding section, prior knowledge and its activation are essential to construction of mental (situation) models (McNamara & Magliano, 2009). Accordingly, if less able structure builders know less, or are less likely to activate their knowledge, their mental models would necessarily be less rich and accurate. Second, connecting ideas and relating across incoming information are fundamental to constructing cohesive and accurate mental models (e.g., Graesser, Singer, & Trabasso, 1994). Theories differ in whether they emphasize explanative connections (e.g., Graesser et al., 1994; Magliano, Trabasso, & Graesser, 1999; Singer & Richards, 2005) or causal connections (e.g., Trabasso et al., 1989), but regardless of the particular nature of the connections, difficulty connecting events, ideas, characters and their goals, pictorial events and so on would likely result in less coherent and accurate mental models.

 Other deficits associated with lower structure-building ability may also have implications for the extraction of a coherent mental model, although these ideas have not been empirically tested. For example, inhibitory deficits could interfere with the construction of a coherent mental model, as posited in the seminal Gernsbacher (1990) view. Inhibitory failures would cause learners to inappropriately create new structures, obviating the construction of a single coherent representational structure. Similarly, a difficulty in identifying the main or important points of an event could reduce the accuracy or cohesiveness of less able structure-builders’ mental models, given that mental models capture the important points. (For instance, Mayer and Gallini,1990, reported that recall of explanative (important) idea units increased 2- to 7-fold for a condition in which accurate mental-model construction was facilitated by providing an illustrative schematic, relative to the control-text condition. Critically, the “mental-model” condition did not improve recall of other factual idea units.) In contrast, encoding deficits are less likely to matter for the construction of mental models, so long as the main points are encoded (a failure to encode less central material would be unlikely to impact the quality of the mental model constructed). This idea awaits further research.

 **Aids that Improve Learning Outcomes for Less Able Structure Builders**

 Low-ability structure builders exhibit ubiquitous comprehension and learning deficits, underscoring the need for techniques to improve their learning outcomes. Though work on this front is in its infancy, the available findings provide insights into the kinds of supports that could improve their learning outcomes. Most of the research to date focuses on testing existing interventions (e.g., embedded questions) which past work suggests would alleviate one or more of the processing deficits outlined in the earlier section (namely, encoding and remembering incoming information; identifying the main points of a text; activating and applying prior knowledge; inhibiting irrelevant knowledge; connecting ideas across the text). An alternate approach would be to target and train particular processes rather than leveraging existing structural supports, but training would likely require a time commitment (as opposed to the relatively immediate support provided by interventions such as embedded questions). Any process-specific training would also need to address whether the processes could be trained independently or if they needed to be trained in a particular order. Although there are no direct data on this issue, across literatures there are successes in training individual processes (e.g., main idea identification, Ritchie, 1985; activation and application of prior knowledge; McNamara, 2004), suggesting that the order may be less important than the total skill set. Process-specific interventions await further research; currently, the initial evidence suggests that embedded questions, exercises that require re-organization, and diagrams illustrating relations all have the potential to help low-ability structure builders, likely supporting multiple key processes.

**Embedded Questions**

Much research has examined whether embedding questions into texts improves learning (Hamaker, 1986, for a review). Perhaps because of mixed results, interest in embedded questions has waned over time. However, embedded questions have great potential when low-ability structure builders are considered, given that answering them stimulates readers to review key ideas and concepts. In other words, embedded questions are a promising tool to support low-ability structure builders’ encoding of the targeted concepts.

 Callender and McDaniel (2007) reported strong support for this possibility. Learners read a 16-page social psychology chapter with eight questions embedded. For example, readers answered the question “What is a scapegoat?” after reading about the concept; they could refer back to the earlier text when answering. As shown in Table 2, answering embedded questions improved low-ability structure builders’ performance almost to the level of high-ability structure builders. This benefit extended to test questions that were related to the targeted concepts but not to completely unrelated test questions (Table 2). The same patterns were observed on short answer test questions that required learners to apply their knowledge (“There are many presidential candidates campaigning to be elected. What is a correspondent inference, and can one be drawn about someone when looking at their behavior on the campaign trail?”). By contrast, high-ability structure builders, who were already performing at a high level, did not profit from the embedded questions. Thus, embedded questions may be best viewed as a potential support for less able-structure builders as opposed to a general learning adjunct.

 Embedded questions are most likely improve encoding of targeted concepts for the lower-ability structure builders (i.e., to address their insufficient encoding), but they may also provide anchors for constructing more unified representations (Callender & McDaniel, 2007). That is, embedded questions may alert less-able structure builders to the main points of the passage and foster inhibition of irrelevant information. Evidence for these possibilities awaits further work.

 The type of embedded questions likely has implications for how and whether they support less-able structure builders. For example, Callendar and McDaniel (2007) included embedded questions designed to activate prior knowledge (explanative questions); however, these questions improved the performance of *high-ability* structure builders rather than *low-ability* structure builders – perhaps because the high-ability structure builders had more knowledge stored in memory to activate (as suggested earlier in this article). Callendar and McDaniel used questions that targeted single concepts (“what is a scapegoat”), which were unlikely to encourage connections across concepts – although another type of question might be able to do so, as suggested by a study on integrative pre-questions (St Hilaire et al., 2019, Experiment 2). Pre-questions differ from embedded questions in that they are provided before reading, with the instruction to look for the answers while reading. Most studies on pre-questions use items that tap individual concepts, but all participants benefited in a study that used integrative pre-questions (e.g., “What is the primary difference between mechanical brakes and hydraulic brakes?”), including those at the lower end of the structure-building spectrum.

**Organizational Supports**

Because low-ability structure builders construct less well-organized representations, they should benefit from techniques that encourage them to organize the concepts, facts, or events in a text, conversation, lecture, or other event. In an initial experiment supporting this prediction, some readers read traditional texts whereas others were given the same texts but with the sentences in the wrong order. These readers had to reorder the sentences to create a text that made sense (Einstein et al., 1990; McDaniel, Einstein, Dunay, & Cobb, 1986). Critically, both low- and high-ability structure builders successfully re-ordered the scrambled sentences (the technique would not be expected to help learning otherwise). Low-ability structure builders who unscrambled the text recalled significantly more of it than low-ability structure builders who read the intact text. This effect held for narrative passages (Experiment 2a) and didactic texts (McDaniel et al., 2002, Experiments 2a and 2b). Given that greater recall is considered to be a marker of a more coherent representation, the results imply that low-ability structure builders can improve the coherence of their mental structures with appropriate intervention.

Of course, educators are unlikely to embrace the use of scrambled texts, even if they do promote organization. More practical and typical would be to encourage learners to create outlines for organizing the concepts covered in lectures or assigned readings (Miyatsu, Nguyen, & McDaniel, 2018). In general, outlining produces similar effects to reordering scrambled sentences, in terms of benefits in recalling a text (Einstein et al, 1990). Accordingly, Bui and McDaniel (2015) examined whether lower structure-builders benefitted from using an outline to guide note-taking during a lecture on mechanics (brakes and pumps). In one condition, students received skeletal outlines to support their note-taking; the outline provided the main ideas and the important subtopics subsumed beneath each. In the control condition, learners took notes without any external support.

Across all levels of structure-building, participants remembered more of the lecture if they had received the outlines to guide their note-taking. Though the outlines did not selectively benefit those at the lower structure building levels, receiving an outline did nevertheless benefit them. In contrast, lower-ability structure builders did not show any benefits from the outlines when taking a problem-solving test that required making inferences. Apparently, the outline promoted effective construction of the discourse (text) level representation and did not extend to the coherence of the lower-ability structure builders’ situational model (Bui & McDaniel, 2015, provide additional discussion). Consideration of participants’ notes support this interpretation. Lower-ability structure builders in the outline condition no longer had to identify the main ideas and recorded significantly more ideas from the lecture in their notes (the idea units in the outline were not included in this total). Presumably recording an idea in one’s notes enhanced its memorability, thus attenuating less-able structure-builders’ deficit in encoding.

The content of the notes also implied that outlining stimulated connections across ideas, creating a cause-and-effect chain of how brakes and pumps worked. The subset of the ideas (from the lecture) that captured this content was significantly more prominent in notes when the outline was provided than when it was not. However, the cause-and-effect content in the outline condition was still significantly less prominent than in a condition that fostered construction of a better mental model for less-able structure builders (described in the next section). This finding is in line with the just-mentioned conclusion regarding limited benefits of outlining for mental-model construction for lower-ability structure builders. Whether the outline alleviated other processing deficits for lower-ability structure builders is less certain: The outline theoretically might have assisted less-able structure builders in inhibiting irrelevant information. The outline would not necessarily have promoted increased activation of prior knowledge, because the outline was tied directly to the structure of the text.

Another promising technique to assist low-ability structure builders to encode relationships among ideas is to provide learners with flashcards that focus on integrating to-be-learned information. For instance, the flashcard, *How are isotopes and half-life related?,* requires learners to integrate and relate two concepts described in a biological anthropology chapter (Lin et al., 2018), as compared to more traditional flashcards that test the definition of a single term or a specific detail about a concept. Lin et al. (2018, Experiment 2) found that the conceptual flashcards produced better performance on short-answer questions that required relating information across the passage (compare-contrast, describe a process, or application) than did “detail” flashcards. Of interest for present purposes, this effect was observed primarily at the lower levels of structure-building ability. Studying with “integrative” flashcards helped lower-ability structure-builders to connect ideas in the event. It likely also enhanced encoding, given the known benefits of retrieval practice (McDaniel, Roediger, & McDermott, 2007) – but that benefit was not documented in the Lin et al. study given that there was no control group. Finally, we note that it is unknown whether the conceptual flashcards would assist with activating prior knowledge and that it is unlikely that they would help with inhibitory deficits.

Importantly, these initial studies show much promise in identifying and developing techniques that can assist lower-ability structure builders in constructing more coherent interconnections and organization of didactic discourse, which in turn supports better learning outcomes.

**Schematic Diagrams and Mental Models**

Understanding technical material is difficult, but can be made easier with the provision of schematic diagrams. For example, a text on the mechanics of pumps is easier to read and understand if accompanied by a diagram that illustrates the pump’s parts and their movements while also showing how function emerges (Mayer & Gallini, 1990). Does this pattern hold across structure-building ability? The study on outlining described earlier (Bui & McDaniel, 2015) included a condition in which learners who were listening to (and taking notes on) a “lecture” on brakes and pumps received a schematic diagram. These participants later answered short-answer questions requiring them to explain or solve problems about brakes and pumps. The results were dramatic. As Figure 5 shows, providing the schematic diagram during note-taking substantially improved later problem-solving performance (short-answer questions) for the lower structure-ability end of the spectrum relative to note-taking with or without a provided outline. Moreover, receiving the schematic diagram completely eliminated differences in performance as a function of structure building ability (a similar pattern was obtained for free recall performance). This initial finding shows that with appropriate assistance, lower-ability structure builders can perform as well on challenging problem-solving questions as high-ability structure builders, implying that they are not incapable of constructing coherent mental models from challenging didactic text. This result clearly opens up important educational possibilities.

The content of participants’ notes suggested that the schematic diagram focused learners on the causal connections key to understanding how the mechanical devices worked: The steps of the cause-and-effect chain were significantly more prominent in the notes taken when the schematic diagram was present than when it was absent (i.e., relative to the outline and control conditions). The implication is that the schematic diagram helped less-able structure builders to make the critical causal connections necessary for an accurate mental model. The schematic diagram also identified the main points with regard to how the mechanical devices worked, removing that burden from the lower-ability structure-builders and potentially assisting with the inhibition of irrelevant information (but as discussed above this possibility awaits more direct evidence). Finally, for completeness we note that it is not clear whether the schematic diagrams helped less-able structure builders activate prior-knowledge.

**Theoretical Post-Script about Improving Learning of Low-Ability Structure Builders**

A key theoretical assumption in the seminal structure-building model is that a central impairment of less-able structure builders is a deficiency in inhibiting irrelevant information (Gernsbacher, 1990; Gernsbacher & Faust, 1991). As a consequence, low-ability structure builders theoretically construct more sub-structures at the expense of creating an overarching coherent representational structure. Though this account did not preclude other deficiencies (others were not explicitly identified or examined), theoretically an inhibitory deficiency could be foundational to less-able structure builders’ poor comprehension and memory of complex events. Some of the findings reviewed in this section do not converge, however, with the possibility that a deficient inhibitory mechanism solely underlies poor structure building. One finding is that low-ability structure builders but not high-ability structure builders benefitted (in free recall of the text) from processing a narrative in which they had to construct the narrative from randomly ordered sentences, relative to reading the intact narrative (McDaniel et al., 2002, Experiment 2a). Indeed the random-text condition showed a dramatic effect: Low-ability structure builders approached the relatively high free recall levels for the narrative displayed by high-ability structure builders. It seems reasonable to expect that many possible inferences would be generated when trying to piece together a narrative from randomly ordered sentences, a number of which would certainly be irrelevant. If low-ability structure builders are challenged by inhibiting irrelevant information, then unscrambling a scrambled-text would likely be overwhelming for low-ability structure builders. Instead, it seems that low-ability structure builders can create coherent structures if the processing task directly demands it.

 More generally, if low-ability structure building resulted solely from an inhibitory deficit, it would be difficult to remediate. Contrary to this expectation, as just reviewed, multiple techniques improve low-ability structure builders’ learning and retention: embedded questions (Callender & McDaniel, 2007), outlines and schematic illustrations (Bui & McDaniel, 2015), and conceptual flashcards (Liu et al., 2018, Experiment 2). It is unclear how these techniques would moderate or overcome inhibitory deficiencies. We hasten to acknowledge that further research is needed to guide a more definitive understanding of the key components underlying individual differences in structure building, and by so doing potentially illuminate a broader range of techniques to overcome the learning deficits exhibited by low-ability structure builders.

**Summary and Conclusions**

 Differences in comprehension skill are well established (e.g., McNamara & Magliano, 2009). The present article highlights a particular individual difference—structure-building (Gernsbacher, 1990) —that seems to capture unique aspects not reflected in standard comprehension tests, and does not seem to represent general verbal or cognitive ability. Over the past 15 years, our and others’ (Maki & Maki, 2002; St. Hilaire et al., 2019) research has consistently revealed the striking importance of this individual difference in learning and retention of texts, lectures, and classroom materials. Indeed, in every study examining individual differences in structure building of which we are aware, there is a relation between structure-building ability and comprehension, retention, and transfer performances (see Table 1). Structure-building likely involves multiple cognitive processes, including basic comprehension processes such as encoding the key ideas (propositions) comprising the event, connecting and relating those ideas, and inhibiting irrelevant information. And importantly, we now have initial evidence that structure-building predicts the alignment of a learner’s representational structure with an expert ideal. Accordingly, this individual difference has the potential to be a powerful tool for identifying and characterizing individuals who, though they may display high verbal scores on standardized tests, may be at risk when digesting, synthesizing, and integrating information.

 The work considered in this article significantly extends the initial investigations of individual differences in structure building (e.g., Gernsbacher & Faust, 1991). This earlier research primarily examined single sentences and pictures, using reaction time as focal measure, or used relatively short stories constructed for experimental purposes. As just noted, the newer work has considerably expanded the scope of findings to a range of naturalistic texts (and lectures), to learning in the classroom, and to a variety of learning outcome measures. From a theoretical perspective, the initial focus was that low-ability structure builders struggled to inhibit irrelevant information, leading them to construct more sub-structures instead of an overarching coherent representational structure. There is evidence to support an inhibitory deficit, however, the results reviewed here suggest that lower-ability structure builders struggle in other ways as well. This is good news, from a practical stance; it might be difficult to improve performance if a basic inhibitory deficit were the sole mediator for low-ability structure building. In contrast, recent work suggests successes in designing practical interventions that scaffold and stimulate improvements in comprehension and learning for less-able structure builders (e.g., Bui & McDaniel, 2015; Callender & McDaniel, 2007; Liu et al., 2018).

In conclusion, we believe that the accumulated research warrants the recognition of structure building as an important individual difference in cognitive functioning. Doing so will be fruitful for stimulating additional theoretical work to understand the underpinnings of low-ability structure building and for guiding applied work to identify less-able students and design interventions to assist them.

Author Notes

 Mark McDaniel’s and Elizabeth Marsh’s contributions on this article were supported in part by grant R305A130535 from the Institute of Education Sciences, US Department of Education. We thank Kathleen Arnold for providing the data on structure building in Figure 1 from Duke University students, North Carolina high school students, and Amazon Mechanical Turk workers. Correspondence concerning this article should be addressed to Mark A. McDaniel, Washington University in St Louis, Psychological and Brain Sciences, 1 Brookings Dr. Campus Box 1125, St Louis, MO 63130. Contact: markmcdaniel@wustl.edu

Footnotes

1 The Multi-Media Comprehension Battery is posted on the Open Science Framework at <https://osf.io/qjwvs>. Both the original and modified versions are posted. At the site, access to the battery can be requested from the administrators (Gernsbacher, Marsh, or McDaniel).

2 For the interested reader and to make this idea more concrete, we briefly mention several unique aspects of stories that could relate to Multi-Media Comprehension Battery scores. One is that stories explicate action sequences that are typically connected through causal links (Trabasso & Sperry, 1985; Trabasso et al., 1989), and these links can be further organized into coherent causal chains relating initial events to final outcomes—provided that the comprehender extracts these causal links. Some have argued that doing so is at least in part a strategic self-regulated process, one that relies on the comprehender’s skill (McNamara & Magliano, 2009). Thus one possibility is that less-able structure builders are less likely to recognize and connect causal links, relying instead on referential connections signaled through argument overlap (e.g., Kintsch, 1995; Kintsch & van Dijk, 1978). Though a good general strategy for a range of text genres, relying on referential connections would provide a less tightly organized and memorable representation than constructing a causal chain from a story (Fletcher & Bloom, 1988; see also Zwaan, Magliano, & Graesser, 1995).

 Stories are also unique in that they reflect a particular grammar (Mandler, 1978; Mandler & Johnson). Story grammars theoretically allow readers to appreciate the functional significance of information in the text and organize that information according to the categories of the story grammar. Perhaps less-able structure builders are those who fail to appreciate and exploit a story grammar during comprehension (in text or in picture presentation; e.g., Baggot, 1975). Low-ability structure builders do show significant reductions in recall of folk tales relative to high-ability structure builders (McDaniel, Hines, & Guynn, 2002), a finding that could be aligned with the just-mentioned possibilities.

3 The correlation between the working memory and the Multi-Media Comprehension battery scores was not reported in Arnold et al. (2017); it was computed from their data for the present article.

4 OSpan refers to the Operation Span task (Unsworth et al., 2005). In this version of the task, participants are shown a series of letters (ranging in set size from three to seven letters each) with a simple arithmetic problem preceding each letter. A potential solution to each arithmetic problem is displayed after the problem, and the participant indicates whether it is the answer; after answering a letter is displayed. Following each set of arithmetic problems and letters, participants are then asked to report the presented letters in the exact order of presentation. The participant’s score is the sum of the number of letters from the perfectly reported sets of letters.

RSpan refers to the Reading Span task (Unsworth et al., 2005). Participants are shown a series of letters (ranging in set size from three to seven letters each) with a sentence preceding each letter. For each sentence, participants determine whether each sentence makes sense (by pressing keys corresponding to Yes or No responses). At the end of each set, participants are asked to report the presented letters in the exact order of presentation.  The score is the same as for OSpan.

5 This Master’s thesis can be accessed at <https://openscholarship.wustl.edu/art_sci_etds/1179/>. For the correlation presented, to parallel the computation from the Arnold et al. (2017) data, we excluded 9 participants with an Operation Span or Reading Span score of *0* or for whom a score was not available for one of the tasks.

6 These unpublished data (manuscript in preparation) are deposited on Open Science Framework. The data set, a code book, and the full correlation table can be found at <https://osf.io/5mz8u/>

7 The generally poor reliability of the reaction times (which are used to compute the difference scores to quantify inhibitory control in these measures) possibly limited the predictive utility of the two other inhibition tasks. With relatively good split-half reliability (.76 in this study), the flanker task provided more stable response times than did Stroop (.31 in this study) and the irrelevant word-meaning task (reaction times ranging from .27 to .48 for sentences with unambiguous and ambiguous target words, respectively). (But see Hedge et al., 2018, indicating relatively low test—retest reliability for Flanker and Stroop interference difference scores, the scores used to gauge inhibition.) Note in the extreme groups approach adopted in the Gernsbacher and Faust (1991) experiment using the irrelevant word-meaning tasks, difference scores were not required because interference was indexed as a within-subjects variable.

References

Arnold, K. M., Daniel, D. B., Jensen, J. L., McDaniel, M. A., & Marsh, E. J. (2016). Structure

building predicts grades in college psychology and biology. *Applied Cognitive Psychology, 30*, 454-459. doi: 10.1002/acp.3226

Arnold, K. M., Drew, E., McDaniel, M. A., & Marsh, E. J. (2019). Two routes to the same

place: Learning from quick closed-book essays versus open-book essays. Manuscript under review.

Arnold, K. M., Umanath, S., Thio, K., Reilly, W. B., McDaniel, M. A., & Marsh, E. J. (2017).

Understanding the cognitive processes involved in writing to learn. *Journal of*

*Experimental Psychology: Applied*, *23*, 115-127. doi: 10.1037/xap0000119

Arthur Jr., W., & Day, D. V. (1994). Development of a short form for the Raven Advanced

Progressive Matrices Test. *Educational and Psychological Measurement, 54*, 394-403. doi: 10.1177/0013164494054002013

Baldwin, R. S., Murfin, P., Ross, G., Seidel, J., Clements, N., & Morris, C. (1989). Effects of

extending administration time on standardized reading achievement tests. *Literacy*

*Research and Instruction*, *29*, 33-38. doi: 10.1080/19388078909557994

Baggot, P. (1975). Memory for explicit and implicit information in picture stories. *Journal of*

*Verbal Learning and Verbal Behavior*, *14*, 535-548.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*,

577–609. doi: 10.1017/S0140525X99002149

Barsalou, L. W. (2009). Simulation, situational conceptualization, and predictions. *Philosophical*

*Transactions of the Royal Society*, *364*, 1281-1289. doi: [10.1098/rstb.2008.0319](https://doi.org/10.1098/rstb.2008.0319)

Bower, G. H., & Morrow, D. G. (1980). Mental models in narrative comprehension. *Science, 24,*

44-48. doi: 10.1126/science.2403694

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some

investigations of comprehension and recall. *Journal-of Verbal Learning*

*and Verbal Behavior, 11,* 717-726.

Britton, B. K., & Gülgöz, S. (1991). Using Kintsch's computational model to improve

instructional text: Effects of repairing inference calls on recall and cognitive structures.

*Journal of Educational Psychology*, *83*, 329-325. doi: 10.1037/0022-0663.83.3.329

Brown, J. I., Fishco, V. V., & Hanna, G. (1993). *Nelson-Denny reading test: Manual for scoring*

*and interpretation, Forms G & H*. Riverside Publishing Company.

Bui, D. C., & 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. doi: 10.1016/j.jarmac.2015.03.002

Callender, A. A., & McDaniel, M. A. (2007). The benefits of embedded question adjuncts for

low and high structure builders. *Journal of Educational Psychology*, *99*, 339-348. doi:

10.1037/0022-0663.99.2.339

Callender, A. A., & McDaniel, M. A. (2009). The limited benefits of rereading educational texts.

*Contemporary Educational Psychology*, *34*, 30-41. doi: 10.1016/j.cedpsych.2008.07.001

Carey, S., & Gelman, R. (Eds.) (1991). Epigenesis of Mind: Essays on Biology and Cognition.

Hillsdale, NJ: Lawrence Erlbaum.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W.

(2005). Working memory span tasks: A methodological review and user’s guide.

*Psychonomic Bulletin & Review*, *12*, 769-786. doi: 10.3758/BF03196772

Copeland, D. E., Magliano, J., & Radvansky, G. A. (2006). Situation models in

comprehension, memory, and augmented cognition. In M. Bernard, J. C. Forsythe, &

T. Goldsmith (Eds.), *Human cognitive models in system design* (pp. 37–66). Mahwah, NJ: Lawrence Erlbaum.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013).

Improving students’ learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4-58. doi: 10.1177/1529100612453266

Dyer, J. W., Riley, J., & Yekovich, F. R. (1979). An analysis of three study skills: Notetaking,

summarizing, and rereading. *The Journal of Educational Research, 73,* 3-7.

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*, 588-598. doi: 10.1037/0278-7393.6.5.588

Einstein, G. O., McDaniel, M. A., Owen, P. D., & Cote, N. C. (1990). Encoding and recall of

texts: The importance of material appropriate processing. *Journal of Memory and*

*Language*, *29*, 566-581. doi: 10.1016/0749-596X(90)90052-2

Fletcher, C. R., & Bloom, C. P. (1988). Causal reasoning in the comprehension of simple

narrative texts. *Journal of Memory and* *Language, 27,* 235-244. doi: 10.1016/0749-

596X(88)90052-6

Garner, R. (1982). Efficient text summarization: costs and benefits. *Journal of Educational*

*Research, 75,* 275-279. doi: 10.1080/00220671.1982.10885394

Gernsbacher, M. A. (1990). *Language comprehension as structure building.* Hillsdale, NJ:

Erlbaum.

Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: a component of

general comprehension skill. *Journal of Experimental Psychology: Learning, Memory,*

*and Cognition*, *17*, 245-262. doi: 10.1037/0278-7393.17.2.245

Gernsbacher, M. A., & Varner, K. R. (1988). *The multi-media comprehension battery* (Tech. Re.

No. 88-07). Eugene, OR: Institute of Cognitive and Decision Sciences.

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general

comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and*

*Cognition*, *16*, 430-445. doi: 10.1037/0278-7393.16.3.430

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*, 1–55.

doi: 10.1017/S0140525X97000010

Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge.

Journal of Educational Psychology, 83(1), 88-96. doi: 10.1037/0022-0663.83.1.88

Goswami, U. (1996). Analogical reasoning and cognitive development. In Reese, H. W. (Ed.)

*Advances in Child Development and Behavior.* Academic Press.

Gouravajhala, R. (2018). Exploring the underlying mechanisms of structure building

(Unpublished Master’s thesis). Washington University in St. Louis, St. Louis, Missouri.

Gouravajhala, R., Nebel, C., Umanath, S., & McDaniel, M. A. (in prep).

Identifying cognitive processes related to individual differences in structure building. Manuscript in revision.

Graesser, A. C. (1975). Study times for sentences in stories. In *Annual Meeting of the Western*

*Psychological Association.* Los Angeles.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of

text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*,

*36*, 193-202. doi: 10.3758/BF03195564

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative

text comprehension. *Psychological Review, 101*, 371–395.

doi: [10.1037/0033-295X.101.3.371](https://psycnet.apa.org/doi/10.1037/0033-295X.101.3.371)

Graesser, A. C., Woll, S. B., Kowalshi, D. J., & Smith, D. A. (1980). Memory for typical and

atypical actions in scripted activities. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 503-515. doi: 10.1037/0278-7393.6.5.503

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational*

*Research*, *56*, 212-242. doi: 10.3102/00346543056002212

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks

do not produce reliable individual differences. *Behavior Research Methods*, 50, 1166-1186. doi: 10.3758/s13428-017-0935-1

Jensen, A. R. (1998). *The g factor: The science of mental ability* (Vol. 648). Westport, CT:

Praeger.

Johnson‐Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, *4*, 71-115.

doi: 10.1016/S0364-0213(81)80005-5

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

Erlbaum.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-

integration model. *Psychological Review*, *95*, 163-182. doi: 10.1037/0033-295X.95.2.163

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production.

*Psychological Review*, *85*, 363-394. doi: 10.1037/0033-295X.85.5.363

Lin, C., Miyatsu, T., & McDaniel, M. A. (2018). Effects of flashcards on learning authentic materials: The role of detailed versus conceptual flashcards and individual differences in structure-building ability. *Journal of Applied Research in Memory and Cognition,* 7 529-539. doi: 10.1016/j.jarmac.2018.05.003

Lorch, R. F., Lorch, E. P., & Inman, W. E. (1993). Effects of signaling topic structure on text

recall. *Journal of Educational Psychology, 85,* 281-290.

doi: 10.1037/0022-0663.85.2.281

Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during

comprehension. *Journal of Educational Psychology*, *9*, 615–629.

doi: [10.1037/0022-0663.91.4.615](https://psycnet.apa.org/doi/10.1037/0022-0663.91.4.615)

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. doi: 10.1037/1076-898X.8.2.85

Mandler, J. M. (1978). A code in the node: The use of a story schema in retrieval. *Discourse*

*Processes, 1*, 14-35. doi: 10.1080/01638537809544426

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and

recall. *Cognitive Psychology*, *9*, 111-151. doi: [10.1016/0010-0285(77)90006-8](https://doi.org/10.1016/0010-0285%2877%2990006-8)

Marsh, E. J., & Fazio, L. K. (2006). Learning errors from fiction: Difficulties in reducing

reliance on fictional stories. *Memory & Cognition, 34,* 1140-1149.

doi: 10.3758/BF03193260

Martin, N. D., Nguyen, K., & McDaniel, M. A. (2016). Structure building differences influence

learning from educational text: Effects on encoding, retention, and metacognitive control.

*Contemporary Educational Psychology*, *46*, 52-60. doi: 10.1016/j.cedpsych.2016.03.005

Mason, M. (1978). From print to sound in mature readers as a form of reader ability and two

forms of orthographic regularity. *Memory & Cognition, 61,* 170-173.

doi: 10.3758/BF03198246

Masson, M. E., & Sala, L. S. (1978). Interactive processes in sentence

comprehension and recognition. *Cognitive Psychology*, *10*, 244-270.

doi: 10.1016/0010-0285(78)90015-4

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words?. *Journal*

*of educational psychology*, *82*, 715-726. doi: 10.1037/0022-0663.82.4.715

McDaniel, M. A., Einstein, G. O., Dunay, P. K., & Cobb, R. E. (1986). Encoding difficulty and

memory: Toward a unifying theory. *Journal of Memory and Language*, *25*, 645-656.

doi: 10.1016/0749-596X(86)90041-0

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.

 doi: 10.1006/jmla.2001.2819

McDaniel, M. A., & Masson, M. E. (1977). Long-term retention: When incidental semantic

processing fails. *Journal of Experimental Psychology: Human Learning and Memory*, *3*,

270-281. doi: 10.1037/0278-7393.3.3.270

McDaniel, M. A., Roediger, H. L., III, & McDermott, K. B. (2007). Generalizing test-enhanced

 learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, 14, 200-

 206.

McNamara, D. S. (1997). Comprehension skill: A knowledge-based account. In M. G. Shafto &

P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science* (pp. 508 – 513). Hillsdale, NJ: Lawrence Erlbaum.

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes,* 38, 1-

30.

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension.

In B. Ross (Ed.), *Psychology of Learning and Motivation*, Vol 51 (pp 297-384).

McNamara, D. S., & McDaniel, M. A. (2004). Suppressing irrelevant information: Knowledge

activation or inhibition? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 465-482. doi: 10.1037/0278-7393.30.2.465

Mednick, S. (1962). The associative basis of the creative process. *Psychological Review, 69,*

220-232. doi: 10.1037/h0048850

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic

review. *Developmental Psychology, 49*, 270-291. doi: 10.1037/a0028228

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not

improve performance on measures of intelligence or other measures of “far transfer” evidence from a meta-analytic review. *Perspectives on Psychological Science, 11*(4), 512-534. doi: 10.1177/1745691616635612

Meyer, B. J. F. (1971). Idea units recalled from prose in relation to their position in the logical

structure, importance, stability, and order in the passage (Unpublished Master’s thesis).

Cornell University, Cornell, NY.

Miyatsu, T., Nguyen, K., & McDaniel, M. A. (2018). Five popular study strategies: their

pitfalls and optimal implementations. *Perspectives on Psychological Science*, *13*(3), 390-407. doi: 10.1177/1745691617710510

Morrow, D. G., Greenspan, S. L., & Bower, G. H. (1987). Accessibility and situation models in

narrative comprehension. *Journal of Memory and Language, 26,* 165-187.

doi: 10.1016/0749-596X(87)90122-7

Neiles, K. Y., Todd, I., & Bunce, D. M. (2016). Establishing the validity of using network

analysis software for measuring students’ mental storage of chemistry concepts. *Journal of Chemical Education, 93,* 821-831. doi: 10.1021/acs.jchemed.5b00748

Newby, T. J., Ertmer, P. A., & Stepich, D. A. (1995). Instructional analogies and the learning of

concepts. *Educational Technology Research and Development*, *43*, 5-18.

doi: 10.1007/BF02300478

Perrig, W., & Kintsch, W. (1985). Propositional and situational representations of text. *Journal*

*of Memory and Language*. 24, 503-519. doi: 10.1016/0749-596X(85)90042-7

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. doi: 10.1080/01638539009544767

Phelan, J. (2009). *What is life? A guide to biology.* New York, NY: W. H. Freeman.

Preacher, K. J., Rucker, D. D, MacCallum, R. C., & Nicewander, W. A. (2005). Use of the

extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods, 10*, 178-192. doi:  10.1037/1082-989X.10.2.178

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize

statistics word problems. *Journal of Educational Psychology*, *88*, 144-161.

doi: 10.1037/0022-0663.88.1.144

Ritchie, P. (1985). The effects of instruction in main idea and question generation. *Reading-*

*Canada-Lecture*, *3*, 139-146.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected

discourse. *Perception & Psychophysics*, *2*, 437-442. doi: 10.3758/BF03208784

Schank. R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures.* Hillsdale, NJ: Lawrence Erlbaum.

Schmalhofer, F., McDaniel, M. A., & Keefe, D. (2002). A unified model for predictive and

bridging inferences. *Discourse Processes*, *33*, 105-132.

doi: 10.1207/S15326950DP3302\_01

Schvaneveldt, R. W. (1990). *Pathfinder associative networks: Studies in knowledge*

*organization.* Norwood, NJ: Ablex Publishing.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective?.

*Psychological Bulletin, 138*(4), 628 – 654. doi: 10.1037/a0027473

St. Hilaire, K. J., Carpenter, S. K., & Jennings, J. M. (2019). Using prequestions to enhance learning from reading passages: The roles of question type and structure building ability. *Memory*, 27, 1204-2113. doi: 10.1080/09658211.2019.1641209

Stine-Morrow, E. A. L., Gagne, D. D., Morrow, D. G., & Dewall, B. H. (2004). Age differences

in rereading. *Memory & Cognition*, 32, 696-710. doi: 10.3758/BF03195860

Stoen, S., McDaniel, M. A., Frey, R., Hynes, M., Cahill, M. (2020). The force concept

inventory: More than just conceptual understanding. *Physical Review Physics Education Research*. doi: [10.1103/PhysRevPhysEducRes.16.010105](https://www.researchgate.net/deref/http%3A//dx.doi.org/10.1103/PhysRevPhysEducRes.16.010105?_sg%5B0%5D=crjZFjRHHCcNYUWUuLXT7qt1oamLWyUZpz39vNOO336QIHizYBx6VzcXLOIuocQ6RhhlAtex7aVWKW33OYR95xpK5Q.wDD_YSx04kbXfznK_yB8Oi-mosjJ0r9exbDqkCAMGtjhg04wnh1md95GzncKSZwZOUWpLX5C1xjHV9GesmiY5Q)

Tamez, E., Myerson, J., & Hale, S. (2008). Learning, working memory, and intelligence

revisited. *Behavioural Processes, 78,* 240-245. doi: 10.1016/j.beproc.2008.01.008

Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative

discourse. *Cognitive Psychology, 9*, 77-110. doi: 10.1016/0010-0285(77)90005-6

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of

narrative events. *Journal of Memory and Language, 24*, 612–630.

doi: [10.1016/0749-596X(85)90049-X](https://psycnet.apa.org/doi/10.1016/0749-596X%2885%2990049-X)

Trabasso, T., van den Broek, P., & Suh, S. (1989). Logical necessity and transitivity of causal

relations in stories. *Discourse Processes, 12*, 1–25. doi: [10.1080/01638538909544717](https://doi.org/10.1080/01638538909544717)

Umanath, S., & Marsh, E. J. (2012). Aging and the memorial consequences of catching

consequences with prior knowledge. *Psychology and Aging, 27,* 1033-1039.

doi: 10.1037/a0027242

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the

operation span task. *Behavior research methods*, *37*, 498-505. doi: 10.3758/BF03192720

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Orlando, FL:

Academic Press.

Wiley, J., Jarosz, A. F., Cushen, P. J., & Colflesh, G. J. (2011). New rule

use drives the relation between working memory capacity and Raven's Advanced

Progressive Matrices. *Journal of Experimental Psychology: Learning, Memory, and*

*Cognition*, *37*, 256-263. doi: 10.1037/a0021613

Williams, B.A., & Pearlberg, S.L. (2006). Learning of three-term contingencies correlates with

Raven scores, but not with measures of cognitive processing. *Intelligence*, *34*, 177–191.

doi: 10.1016/j.intell.2005.03.007

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation

models in narrative comprehension: An event-indexing model. *Psychological Science, 6*,

292–297. doi: [10.1111/j.1467-9280.1995.tb00513.x](https://doi.org/10.1111/j.1467-9280.1995.tb00513.x)

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension

and memory. *Psychological Bulletin, 123*, 162–185. doi: 10.1037/0033-2909.123.2.162

**Table 1**

*Studies Reporting a Relationship between Structure Building Ability and Learning: Methodological Features and Effect Sizes*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Length | Study | Topic | Design | Test Format(s) | Effect Sizes |
|  |  |  |  |  |  |
| Short descriptive passages (14-20 sentences) | McDaniel et al. (2002, Experiment 2a) | Fairy tales | Extreme groups (Low: <32; High: >36) | Free recall | Large (Hedges’ *g* = .82) |
|  | McDaniel et al. (2002, Experiment 2b) | Avalanches (ice flows) | Extreme groups (Low: <32; High: >36) | Free recall | Large (Hedges’ *g* = .63) |
| Moderate length passages (3 - 4 pages) |  |  |  |  |  |
|  | Martin et al. (2016) | Mechanical devices (brakes and pumps) | Extreme groups (Low: <32; High: >36) | Free recall; multiple-choice; problem-solving | Large (Cohen’s *d* = 1.33) |
|  | Arnold et al. (2017) | Astronomy | Continuous variable | Multiple-choice; problem-solving | Multiple choice: Large (η2 = .39); Problem-solving: Large (η2 = .43) |
|  | Lin et al.(2018, Experiments 1 and 2)St. Hilaire et al. (2019, Experiments 1 and 2) | Biological anthropology; GeologyMechanical devices (brakes and pumps) | Continuous variableContinuous variable | Multiple-choice; short-answerShort-answer; Fill-in the-blank | Expt. 1 Multiple choice: Small (ηp2 = .076): Short-answer: Medium (ηp2 = .099)Expt. 2 Multiple choice: Small (ηp2 = .073); Short-answer: Small (ηp2 = .04)Could not be obtained |
| Chapter-length passages |  |  |  |  |  |
|  | Callender & McDaniel (2007) | Social psychology textbook | Extreme groups (Low: <32; High: >36) | Multiple-choice; short-answer | Multiple choice: Large (Cohen’s *d*  = .92); Short answer: Large (Cohen’s *d* = .78) |
|  | Callender & McDaniel (2009, Experiment 4) | Biopsychology and abnormal psychology textbook  | Continuous variable | Multiple-choice; short-answer | Multiple choice: Biopsychology = Large (Cohen’s *d* = .93); Abnormal psychology = Medium (Cohen’s *d* = .63)Short answer: Biopsychology = Large (Cohen’s *d* = .99); Abnormal psychology = Medium (Cohen’s *d* = .66) |
|  | Gouravajhala (2018) | Biology textbook | Continuous Variable | Main point identification; short-answer; relatedness ratings | Main Point ID: Small (Cohen’s *d* = .35)Short Answer: Small (Cohen’s *d* = .44)Relatedness Ratings Task 1: Small (Cohen’s *d* = .38)Relatedness Ratings Task 3: Medium (Cohen’s *d* = .53) |
| Aurally presented passages |  |  |  |  |  |
|  | Bui & McDaniel (2015) | Mechanical devices (brakes and pumps) | Continuous variable | Free recall; short-answer | Free recall: Small (ηp2 = .04)Short-answer: Small (ηp2 = .05) |

Note: Effect sizes are typically amplified in extreme-groups designs relative to continuous-variable designs (Preacher, Rucker, MacCallum, & Nicewander, 2005).

Table 2

*Mean Performance on Multiple-choice and Application Questions by Ability and Question Type in Callender and McDaniel (2007)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Low-Ability Structure Builders |  | High-Ability Structure Builders |
| Multiple-choice |  |  |  |  |  |  |
|  |  | Control | Embedded |  | Control | Embedded |
|  | Target | .54 (.35) | .75 (.24) |  | .84 (.21) | .83 (.19) |
|  | Related | .46 (.24) | .68 (.24) |  | .68 (.25) | .75 (.25) |
|  | Unrelated | .59 (.29) | .67 (.28) |  | .88 (.16) | .67 (.24) |
| Application  |  |  |  |  |  |  |
|  |  | Control | Embedded |  | Control | Embedded |
|  | Target | .40 (.21) | .59 (.24) |  | .57 (.26) | .54 (.20) |
|  | Unrelated | .40 (.18) | .42 (.25) |  | .60 (.22) | .47 (.29) |

*Note*. Standard deviations in parentheses. Adapted from Table 1 of Callender and McDaniel (2007).

*Figure 1*. Distribution of Multi-Media Comprehension Battery scores across five samples: Duke University (*n* = 70), high schoolers in North Carolina (*n* = 31), Amazon Mechanical Turk (*n* = 191), Washburn University (*n* = 55), and Washington University in St. Louis (*n* = 100). Scores on the y-axis represent raw scores on the Multi-Media Comprehension Battery (total possible range: 0 – 48).

**A B**

*Figure 2.* Standardized Multi-Media Comprehension Battery scores as related to an average Working Memory Composite Score (Standardized R-span averaged with Standardized O-Span). Panel A shows data from Arnold et al. (2017) and Panel B shows data from Gouravajhala (2018).



*Figure 3*. Mean proportion recalled by low- and high-ability structure builders on initial and final recall tests. Data from Martin et al. (2016).

Figure 4. The average median expert network (left panel), an example of a low-similarity network (to the expert) for a low-ability structure builder (middle panel; Multi-Media Comprehension Battery = 26), and an example of a relatively similar network from a high-ability structure builder (right panel; Multi-Media Comprehension Battery = 42) network (from Gouravajhala, 2018)



*Figure 5*. Proportion of correct answers on the short-answer test as a function of learning-aid condition and structure building ability. From Bui & McDaniel (2015; Figure 3).