Adults' and children's monitoring of story events in the service of comprehension

Catherine M. Bohn-Gettler · David N. Rapp ·
Paul van den Broek · Panayiota Kendeou ·
Mary Jane White

Published online: 10 May 2011
© Psychonomic Society, Inc. 2011

Abstract When reading narratives, adults monitor shifts in time, space, characters, goals, and causation. Shifts in any of these dimensions affect both moment-by-moment reading and memory organization. The extant developmental literature suggests that middle school children have relatively sophisticated understandings of each of these dimensions but does not indicate whether they spontaneously monitor these

dimensions during reading experiences. In four experiments, we examined the processing of event shifts by adults and children, using both an explicit verb-clustering task and a reading time task. The results indicate that middle school children's and adults' post-reading memory is organized using these dimensions but that children do not monitor dimensions during moment-by-moment reading in the same manner as adults. These differences were not a function of differentially difficult texts for children and adults, or between-group differences. The findings have implications for models of adult and child text processing and for understanding children's developing narrative comprehension.

Keywords Reading · Comprehension · Development · Narratives · Event indexing · Discourse processing · Text processing

This research was supported in part by grants from the National Institutes of Health (T32-HD007151), the Center for the Improvement of Early Reading Achievement (CIERA), and the Center for Cognitive Sciences, University of Minnesota. This work is based on a doctoral dissertation conducted at the University of Minnesota. We thank Jeanne Scheiderer, April Pemeticky, Deborah Shevlin, Mohsina Ahmed, Jennifer Hodgson, Kaitlyn Wahlsten, Jennifer Willcutt, Melinda Mueller, and Sterling Hutchinson for their assistance in conducting this study. We also thank three anonymous reviewers for their comments on earlier versions of the manuscript.

C. M. Bohn-Gettler (✉)
Department of Counseling, Educational, and School Psychology,
Wichita State University,
1845 Fairmount, Campus Box 123, Wichita, KS 67260, USA
e-mail: kate.bohn@wichita.edu

D. N. Rapp
Northwestern University,
Evanston, IL, USA

P. van den Broek
Leiden University,
Leiden, Netherlands

P. Kendeou
Neapolis University Pafos,
Pafos, Cyprus

M. J. White
University of Minnesota,
Minneapolis, MN, USA

When beginning a work of fiction, readers have the opportunity to become familiar with characters, settings, and events as carried forward by story plot. Sometimes those plots set up situations that are fairly consistent and coherent: a small number of characters interact in a restricted set of locations, and events occur in a clear, causal sequence. But stories are not usually so straightforward; they often include time shifts, surprising character behaviors, and plot twists. Rather than dissuading readers, books with such complexities (including the well-read adventures of Tom Sawyer, Harry Potter, and Encyclopedia Brown) draw readers to them.

Comprehending these stories requires that readers carefully monitor the ways in which events play out, indexing what has and has not changed over the course of the narrative (Zwaan & Radvansky, 1998). The goal of the present study was to examine these monitoring processes and to assess similarities and differences in these processes

between adults and children. We also examined whether any developmental trends in monitoring might be a function of skill-based or resource-driven influences or, perhaps, simply age-appropriate text difficulty. We begin with a brief review of the adult literature on readers' monitoring of text dimensions, followed by the relevant children's literature. These literatures frame the theoretical approaches and questions that guided our investigations of readers' monitoring propensities and the resulting consequences for comprehension. Comparing the literatures also highlights the need for identifying developmental patterns of monitoring during comprehension.

Dimension monitoring among adults

Models of adult reading have focused on the types of mental representations that are encoded and applied in the service of comprehension. One classic account, the *tripartite* theory of comprehension (Kintsch & van Dijk, 1978), identifies three levels of representations readers might build. The most superficial is the surface level, which encodes the exact words in a text. At the intermediary level is the textbase, representing the basic idea units conveyed in the text. Finally, at the deepest level is the situation model, which contains information described by the text but not necessarily contained within it, including inferences and explanations (Fletcher, 1994; Johnson-Laird, 1983; Kintsch & van Dijk, 1978; Zwaan, 1999). Situation model representations are most associated with successful comprehension, since they require that readers make connections between information across the text (and outside of it) to build a holistic and coherent understanding of the material (Zwaan, 1999; Zwaan & Radvansky, 1998). Although readers can encode a variety of text features into situation models, the most crucial features for comprehension have been outlined in the *event-indexing model*: Readers attempt to index time, space, characters, goals, and the causal sequences of events. These dimensions are monitored during moment-by-moment reading and have consequences for the organization of reader memory (Zwaan, Langston, & Graesser, 1995; Zwaan, Magliano, & Graesser, 1995).

A processing model of these activities has emerged through the incorporation of the event-indexing model with Gernsbacher's (1990) *structure-building framework* (Zwaan & Radvansky, 1998). According to the framework, readers encode mental representations of events as they are described in texts. When new information is consistent with a reader's current representation, the information is integrated into the unfolding situation model. In contrast, inconsistent events necessitate the encoding of new model substructures, and these new substructures make previous

events less accessible in memory (Gernsbacher, 1997). The general process of building text structures can accommodate monitoring of the dimensions described in the event-indexing model. For example, as a reader encodes information about the setting of a text, changes in setting lead to new memory representations, with a concomitant change in the ease with which readers can retrieve earlier setting information from memory (Radvansky, Zwaan, Federico, & Franklin, 1998; Rapp & Taylor, 2004; Wilson, Rinck, McNamara, Bower, & Morrow, 1993; Zwaan, 1996). Any discontinuity then, as exemplified by shifts in the five dimensions, leads to a decrease in accessibility.

Evidence for the processing effects of discontinuities has been obtained with analyses of memory organization and moment-by-moment reading. For the former, post-reading clustering of events is a direct function of dimension-based continuities (Zwaan, Langston, & Graesser, 1995). When asked to group verbs from previously read narratives into related pairings, readers are more likely to pair verbs that are continuous on a dimension (e.g., verbs representing events occurring in the same spatial location) than verbs that are discontinuous. These effects emerge even when surface-level text factors (e.g., number of intervening words between verbs) are taken into account.

With respect to moment-by-moment reading, adult readers exhibit slowdowns when they encounter event shifts (Zwaan, Magliano, & Graesser, 1995; Zwaan, Radvansky, Hilliard, & Curiel, 1998). Memory for words from preceding events also decreases as a function of dimension shifts. For example, participants have more difficulty recognizing words from sentences describing earlier story events when they follow long shifts (e.g., an hour later), as compared with short shifts (e.g., a minute later). These processing decrements increase when multiple shifts occur simultaneously, as exemplified by longer reading times and retrieval latencies (Radvansky et al., 1998; Rapp & Taylor, 2004; Zwaan, 1996). This has highlighted the importance of examining readers' monitoring of multiple dimensions simultaneously, since dimensions are often causally linked in narratives (see also Jahn, 2004; Sundermeier & van den Broek, 2005; Taylor & Tversky, 1997).

As mentioned previously, work on the event-indexing model has focused almost exclusively on adults' propensities for, and the consequences of, monitoring narrative dimensions. The descriptive and explanatory power of the model, though, has not yet been extended to the investigation of children's comprehension.

Comparisons with dimension monitoring among children

One way to contextualize developmental accounts is to evaluate the similarities between adult and child findings.

Recall that adult findings describe how information from a text is encoded, the ease or difficulty of integrating that information, and how new events lead to new representational substructures. Analogously, developmental theories have examined how children encode and integrate new information into their current representations (Oakhill, 1994; van den Broek, 1997). Children can modify new information to fit into their current representation (assimilation) or revise existing representations and create new representations (accommodation) (Oakhill, 1994; Piaget & Inhelder, 2000; van den Broek, 1997) by activating related understandings to form connections between new information and prior knowledge (MacWhinney, Leinbach, Taraban, & McDonald, 1989; Pascual-Leone, 1970, 2000).

Another critical similarity is that the dimensions of interest in adult comprehension are identical to those in accounts of children's comprehension. For example, some neo-Piagetian accounts describe how children develop in their understandings of fundamental conceptual domains, including number (which includes time), space (which includes object relations and spatial layouts), and narrative understanding (which includes distinguishing characters, understanding causal sequences of events, and inferring how character goals may have consequences for the plot) (Case, Demetriou, Platsidou, & Kazi, 2001; Case, Okamoto, Griffin, McKeough, Bleiker, Henderson, & Keating, 1996). This literature documents that, by early middle school, children have sophisticated understandings of each of these dimensions and how they interact.

An important concern for developmental accounts involves understanding influences on any developmental changes and trends. Information-processing theorists argue that working memory is a primary constraint on development (Gathercole, Pickering, Ambridge, & Wearing, 2004; Swanson, 2008; Zelazo, 2004). Working memory encompasses the resources recruited in the service of comprehension and that are utilized, for example, to combine information in the text with prior knowledge. Working memory has a limited capacity that increases throughout childhood and adolescence (Conklin, Luciana, Hooper, & Yarger, 2007; Luciana, Conklin, Hooper, & Yarger, 2005; Miller, 1956; Swanson, 2008), which could play a role in children's ability to monitor multiple narrative dimensions.

A key process that is important for the growth of working memory as it relates to success in comprehension is the development of automaticity in monitoring text features (Samuels, 1987; Tronsky, 2005; Vadasy & Sanders, 2008). Tasks require less strategic attention (and fewer cognitive resources) as they become more practiced (Siegler & Alibali, 2005; Tronsky, 2005). With development, children become more efficient with certain tasks (such as basic reading skills, including decoding), which frees up cognitive resources for tasks such as dimension

monitoring (Duke, Pressley, & Hilden, 2004; Kuhn & Stahl, 2003). Therefore, children's automaticity in reading comprehension could play a critical role in "releasing" mental resources for dimension monitoring.

The present study

The adult literature indicates that readers monitor narrative dimensions during reading, meaning that (1) they process shifts in each dimension and (2) these shifts influence the organization and accessibility of text information in memory (Zwaan, Langston, & Graesser, 1995; Zwaan, Magliano, & Graesser, 1995). In contrast, the child literature has rather generally focused on the notion that, by late elementary school, children have sophisticated understandings of narrative dimensions (Case et al., 1996). Although these results indicate that event dimensions are relevant to both adults and children, they do not specify the degree to which children monitor dimensions during reading.

Because we were precisely interested in the degree to which children monitor dimensions (i.e., whether they utilize the information that they have been shown to understand), we examined early middle school children who, presumably, had relatively sophisticated understandings of each dimension (and might engage in spontaneous monitoring behavior). On the basis of research examining the development of children's understandings of dimensions, one hypothesis is that middle school children encode event shifts and monitor dimensions in a manner similar to that of adults. However, limitations such as those associated with developing working memory suggest that middle school children might not monitor dimensions in the same way as adults.

Middle school students represent a crucial starting point for investigating children's monitoring behaviors because they are in the midst of an important transition: They are now regularly being asked to apply their knowledge of how to read (phonemic awareness, alphabetic principles, decoding) to learn from texts as provided in their content area classes (Chall, 1983; Hagaman & Reid, 2008). As a practical matter, this age is also important to study on the basis of recent calls for research on how adolescent readers differ from elementary school readers (Jacobs, 2008) and how comprehension difficulties (despite successful decoding) pose problems for middle school children (Underwood & Pearson, 2004; Williams, 2005). Indeed, some middle school children struggle with comprehension skills in addition to decoding and vocabulary (Cromley & Azevedo, 2007; Rasinski, Padak, McKeon, Krug-Wilfong, Friedauer, & Heim, 2005), which could affect whether students monitor dimensions. In addition, because reading is a

resource-intensive activity and because younger readers are still building expertise with respect to resource allocation (both strategic and automatic), early middle school children might not monitor narrative dimensions to the same degree as adults, despite possessing an understanding of the properties of these dimensions.

To obtain a better picture of monitoring behaviors, the present project examined children's and adults' post-reading comprehension products and their moment-by-moment reading. This multi-method analysis is important, since monitoring behaviors might be exerted both during and after reading. Because any propensity to monitor dimensions might be affected by skill-based and resource-driven factors, we also evaluated the influence of reading ability and working memory on comprehension behaviors. In [Experiment 1](#), we examined adults' and children's dimension monitoring as measured through a post-reading verb-clustering task. In [Experiment 2](#), we examined their reading latencies for texts containing dimension continuities and discontinuities. In [Experiment 3](#), we examined whether the results might be due to the relative difficulty of texts as a function of age. In [Experiment 4](#), we utilized a within-subjects design to further confirm the findings obtained in [Experiments 1 and 2](#).

Experiment 1

The goal of this experiment was to assess whether children's memory for narratives, like adults' memory, is structured around the continuity of narrative dimensions. Adult and child participants read narratives and, afterward, grouped verbs from each story into related pairings. Each possible pairing was scored for how often the pairings were made by participants. These scores, which have been used in previous research, provide an index of how related the two verbs are in the text and, thus, the strength of a reader's long-term memory connection between the two verbs (Britton & Gulgoz, 1991; McNamara, Kintsch, Songer, & Kintsch, 1996; Zwaan & Brown, 1996; Zwaan, Langston, & Graesser, 1995). These pairings can also reflect the level of representation at which the verbs are related. For example, if verbs are paired because they are in the same sentence, this could reflect a surface-level representation. If verbs are paired because they represent events that occurred in the same spatial setting, this could indicate that the spatial dimension at the situation model level was involved in linking the two events in memory.

We hypothesized that both adults and children would pair verbs that represent continuities across dimensions, consistent with previous work with adults (Zwaan, Langston, & Graesser, 1995) and consistent with the notion that middle school children's sophisticated understandings of narrative dimensions might extend to their

reading experiences (Case et al., 2001, 1996; van den Broek, 1997). An alternative hypothesis might suggest that children must expend more cognitive resources on basic reading tasks, such as decoding, because these skills require strategic attention (e.g., Duke et al., 2004; Kuhn & Stahl, 2003). This would reduce the resources available for children to monitor dimension changes, thus influencing the likelihood that they would pair verbs representing dimension continuity.

Because the latter hypothesis relies on considerations of working memory and reading fluency, we measured those constructs in the children to examine their impact on verb clustering. Even if children perform similarly to adults, it is worth considering whether individual differences in working memory or fluency predict the proficiency with which children monitor dimensions in the service of comprehension (Siegler & Alibali, 2005; Swanson, 2008).

Method

Participants

Eighty-two adults and 47 twelve-year-olds (mean age = 12.67 years, $SD = 0.39$) participated in this experiment. All were native English speakers. The children's data were part of a larger longitudinal study examining the development of comprehension skills across different media.

Materials

Texts Each participant read four stories from Zwaan, Langston and Graesser (1995); these stories have also been studied in other projects (Graesser, 1981; Graesser & Clark, 1985). Each story was approximately 100 words long, with an average Flesch–Kincaid grade level of 4.12. The stories were “The Czar and His Daughters,” which was about heroes saving a Czar's daughters from a dragon; “The Boy and His Dog,” which was about a dog returning to his owner after chasing a rabbit and fox; “John at Leone's,” which was about a man paying a restaurant bill; and “The Ant and the Dove,” which was about an ant saving a dove from a bird catcher.

Ten unique verbs were selected from each story. Two trained coders rated every possible pairing of the 10 verbs (a total of 45 pairs per story) for continuity in each event-indexing dimension (1 = *shift*, 0 = *no shift*), the number of words intervening between the verbs (surface distance), whether the words appeared in the same sentence (surface connections; coded as 1 = *yes*, 0 = *no*), and whether they shared an argument (a noun, pronoun, or noun phrase; argument overlap, coded as 1 = *yes*, 0 = *no*). Interrater agreement was high: $.84 < ks < 1.00$.

To determine whether the dimension shifts were related in the stories, categorical pairwise correlations were calculated between all dimensions. Shifts in time were correlated with shifts in space ($\phi = .44, p < .001$), characters ($\phi = .22, p < .01$), causation ($\phi = .53, p < .001$), and goals ($\phi = .47, p < .001$). Shifts in space were correlated with shifts in characters ($\phi = .26, p < .01$), causation ($\phi = .55, p < .001$), and goals ($\phi = .50, p < .001$). Shifts in characters were correlated with shifts in causation ($\phi = .18, p < .05$) and goals ($\phi = .20, p < .01$). Finally, shifts in causation were correlated with shifts in goals ($\phi = .57, p < .001$).

In addition, we asked 20 adults to pair the verbs without reading the narratives. These pairings provided an index of the likelihood that verbs would be paired on the basis of general lexical knowledge, rather than story comprehension (Zwaan, Langston, & Graesser, 1995), and were used as a control in our examination of participant performance.

Fluency The children completed a curriculum-based measurement task to assess oral reading fluency (Deno, 1985). Children were asked to read aloud three texts for 1 min per text. The sum of the total number of words read correctly was calculated. Words that the children did not read correctly, including omissions, insertions, hesitations of more than 3 s, mispronunciations, and substitutions, were subtracted from the sum. The reliability and validity of this measure are well supported (Marston, 1989). The task took 5 min to complete.

Working memory The children completed the sentence span task of working memory, adapted by Swanson, Cochran and Ewers (1989) from an original task by Daneman and Carpenter (1980). In the task, the experimenter reads a set of sentences aloud, and each participant responds to a comprehension question about the sentences. The participant is then asked to recall the last word of each sentence in the order originally presented. Participants begin with smaller sets of sentences (two sets of two sentences each) and proceed sequentially to larger sets (two sets of five sentences each). A participant's final score is calculated as the total number of words correctly recalled in the sets for which the comprehension question was answered correctly (i.e., Friedman & Miyake, 2005). This task took 10–15 min to complete.

Procedure

All participants were tested individually. Children completed the sentence span task, followed by the fluency task. (Neither of these tasks were completed by the adult participants.) Following the tasks or, in the case of the

adult participants at the beginning of the experiment, participants received a booklet with the four stories to read at their own pace. After reading each story, each participant received a list of 10 verbs from the story and a column of seven boxes. In each box, participants were asked to write two verbs that they thought belonged together, with no further explanation to guide their pairings. The pairing task took about 20 min to complete.

Results

A challenge for previous research with verb pairing methodologies is that the units of analysis are often aggregated verb pairs, which ignores participant variation. Here, hierarchical generalized linear modeling (HGLM) was used to account for the multi-level structure of the verb pairings nested within participants, which reduces both aggregation bias and misestimation of the standard errors (Richter, 2006). The dependent variable was whether each participant matched each possible pair of verbs. Because this is a dichotomous variable (e.g., participants did or did not make each possible pairing), the dependent variable was transformed to fit a Bernoulli distribution with a logit-link function (Raudenbush & Bryk, 2002), making the dependent variable the log-odds likelihood of providing each pairing. For example, a log-odds coefficient of .67 for surface connections would mean that the verbs are .67 (log) times more likely to be paired if they are in the same sentence than if they are not. A log-odds coefficient of -.67 for surface connections would mean that the verbs are .67 (log) times less likely to be paired if they are in the same sentence than if they are not.

The predictor variables were the control variables (surface distance, surface connections, argument overlap, and lexical knowledge) and the text dimensions (shifts in each of the five dimensions). For the children, fluency and working memory were added to the models. Because of the large number of parameters, an alpha level of $\alpha = .01$ was used for all tests. The penalized quasi-likelihood method was used (Raudenbush & Bryk, 2002). Finally, the adult and child data were modeled separately to assess qualitative differences in dimension monitoring.

The steps in the models involved first creating an unconditional model to check whether there was enough variation in the data to warrant further modeling. The next step was to examine the predictive influence of the control variables on verb pairings. Next, the dimensions were added to the model to assess their influence over and above the control variables. Finally, for the children only, the working memory and fluency variables were added to the model to assess their unique contributions to verb pairings over and above the other variables.

Verb-clustering scores To assess the variation, unconditional models were estimated. Among the adults, the average log-odds of making a verb pairing was $\hat{\gamma}_{00} = -1.82$, $SE = 0.01$, $p < .001$. Among the children, the average log-odds of making a pairing was $\hat{\gamma}_{00} = -1.73$, $SE = 0.01$, $p < .001$. This result indicated that there was variability, thus justifying further modeling.

To assess whether the text controls predicted pairings, we added the control variables to the level-1 models. Among the adults, all of the controls predicted pairings ($ps < .001$). Adults were more likely to pair verbs that had fewer intervening words (surface distance), $\hat{\gamma}_{10} = -0.02$, $SE = 0.002$, $p < .001$, were in the same sentence (surface connections), $\hat{\gamma}_{20} = 1.29$, $SE = 0.11$, $p < .001$, shared argument overlap, $\hat{\gamma}_{30} = 0.46$, $SE = 0.05$, $p < .001$, and were consistent with general lexical knowledge, $\hat{\gamma}_{40} = 0.92$, $SE = 0.13$, $p < .001$. Children were also more likely to pair verbs that had fewer intervening words, $\hat{\gamma}_{10} = -0.02$, $SE = 0.003$, $p < .001$, occurred in the same sentence $\hat{\gamma}_{20} = 0.95$, $SE = 0.14$, $p < .001$, shared argument overlap, $\hat{\gamma}_{30} = 0.39$, $SE = 0.05$, $p < .001$, and were consistent with general lexical knowledge, $\hat{\gamma}_{40} = 1.12$, $SE = 0.16$, $p < .001$.

To assess whether shifts in each dimension predicted pairings over and above the control variables, the five dimensions were added to the level-1 models (see Table 1 for the results). Among the adults, verbs that took place within a continuous event were more likely to be paired than were verbs that were separated by a shift, for each of the five dimensions ($ps < .001$). Similarly, among the children, verbs that took place within a continuous event were more likely to be paired than were verbs that were separated by shifts in time, characters, causation, and goals ($ps < .01$), but not space (although space was in the same direction, $p = .09$). In order to directly address potential differences between the adults and children in dimension monitoring, *t*-tests were run comparing the coefficients of the adults versus children for each dimension. The only significant difference was that adults were more likely than children to monitor changes in space, $t(125) = 2.63$, $p < .01$, $d = 0.48$. None of the other effects were significant, ($ts < 1.66$).

Finally, fluency and working memory were added to the level-2 model for the children only. The average score for the fluency task was $M = 184.28$ ($SD = 39.66$) words. The average score for the working memory task was $M = 20.16$ ($SD = 3.83$) words. Fluency and working memory did not predict verb pairings as main effects. However, because they were added as level-2 variables, we tested whether they had cross-level interactions with any of the other level-1 predictors. Such interactions would indicate that a level-2 variable either strengthens or weakens the effect of a level-1 variable on the verb pairing scores. There was one

statistically significant interaction (at an alpha level set to .01): The higher a child's reading fluency, the less of an effect argument overlap had on verb pairings, $\hat{\gamma}_{31} = -.01$, $p < .001$ (see Table 1 for the results).

Discussion

The adult data replicated Zwaan, Langston and Graesser's (1995) findings that adults are more likely to pair verbs associated with continuous, as compared with discontinuous, events along the dimensions of time, space, characters, goals, and causation. The results also supported the hypothesis that children's pairings show a similar pattern: Continuity in all of the dimensions except space predicted verb pairings. We note, though, that the pattern of data for space was in the same direction as the other dimensions.

These data support the view that middle school children have relatively sophisticated understandings of each dimension (Case et al., 2001, 1996; van den Broek, 1997). Narrative dimensions influence readers' memory organization for narratives, even in the previously untested case of younger readers. Additionally, fluency and working memory did not show any main effects on verb pairings. However, lower fluency resulted in children relying more on argument overlap. Since children with lower fluency skills tend to be less automatic and proficient with basic reading skills, such as decoding (Wood, 2006), it makes sense that they would be more likely to rely on surface-level characteristics when making verb pairings and that such reliance could lead to a decreased influence of dimensions.

Experiment 1 focused on post-reading memory organization. However, examining both moment-by-moment comprehension processes and post-reading memory is crucial for obtaining a complete picture of how dimension monitoring develops (Magliano & Graesser, 1991; Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007; Zwaan & Rapp, 2006). Although children's memory organization showed patterns similar to adults', there is no a priori reason to expect that the processes by which such organizations are shaped are necessarily identical.

Experiment 2

The goal of Experiment 2 was to assess whether children monitor narrative dimensions in a manner similar to that of adults during moment-by-moment reading. Adults and children read narratives, and their reading times were recorded (similar to the work of Lorch & Myers, 1990). Previous studies have demonstrated that adult reading times increase following dimension shifts (Zwaan, Magliano, &

Table 1 Experiment 1: Summary of HGLM analyses for variables predicting verb pairings

	Controls + Dimensions				Controls, Dimensions, WM, Fluency	
	Adults		Children		Children Only	
	Coef.	SE	Coef.	SE	Coef.	SE
Fixed effects						
Intercept, γ_{00}	−0.62***	.09	−.64***	.13	−1.53*	.75
Fluency, γ_{01}					.003	.003
WM, γ_{02}					.02	.04
Surface dist., γ_{10}	−.002	.001	−.003	.002	.001	.01
Fluency, γ_{11}					−.00004	.00004
WM, γ_{12}					.0001	.0004
Surf. conn., γ_{20}	.67***	.11	.40**	.14	−.86	1.01
Fluency, γ_{21}					−.001	.01
WM, γ_{22}					.08	.04
Arg. overlap, γ_{30}	−.04	.07	.03	.07	.88**	.29
Fluency, γ_{31}					−.01***	.001
WM, γ_{32}					.01	.01
Lexical, γ_{40}	.51**	.14	.77***	.18	.17	.94
Fluency, γ_{41}					−.003	.004
WM, γ_{42}					.06	.05
Time, γ_{50}	−.59***	.07	−.37***	.11	.93	.53
Fluency, γ_{51}					−.005*	.002
WM, γ_{52}					−.02	.02
Space, γ_{60}	−.42***	.07	−.14	.08	.34	.42
Fluency, γ_{61}					.001	.002
WM, γ_{62}					−.03	.02
Characters, γ_{70}	−.43***	.07	−.44***	.08	−.13	.43
Fluency, γ_{71}					.002	.002
WM, γ_{72}					−.04	.02
Causation, γ_{80}	−.52***	.06	−.73***	.12	−.72	.50
Fluency, γ_{81}					.005*	.002
WM, γ_{82}					−.05*	.02
Goals, γ_{90}	−.31***	.06	−.32***	.07	−.85	.51
Fluency, γ_{91}					−.003	.001
WM, γ_{92}					.05	.03

Coef. = slope coefficient; SE = standard error

* $p < .05$, ** $p < .01$, *** $p < .001$

Graesser, 1995; Zwaan et al., 1998). We expected to replicate this work, and hypothesized that shifts in the dimensions of time, space, characters, goals, and causation would predict increases in reading times. On the basis of the previous experiment, we predicted that 12-year-olds should have relatively sophisticated understandings of the various dimensions (Case et al., 1996). If these understandings influence online reading activity, their reading times should increase when a dimension shift occurs.

However, moment-by-moment reading is likely to be influenced by reading skill and reading strategies (Siegler &

Alibali, 2005). In fact, children can continue to struggle with comprehension skills well into adolescence and adulthood (Cromley & Azevedo, 2007; Duke et al., 2004). This may lead to a focus on lower-level reading tasks (e.g., decoding and identification of words), with less attention to the moment-by-moment monitoring of multiple dimensions simultaneously. Likewise, children's knowledge of narrative dimensions may not be as well practiced or refined as adults', which might contribute to a failure to encode dimension shifts.

As in Experiment 1, we included a measure of working memory—but this time, for both adults and children. This

allowed for the testing of resource-based processing constraints on moment-by-moment reading behaviors (Calvo, 2001, 2005; Linderholm, Cong, & Zhao, 2008). Recall that, in Experiment 1, we also included a measure of reading fluency, which failed to interact with children's verb pairing behaviors. Fluency has generally been associated more with basic reading skills than with situation model construction (Collins & Levy, 2008). Therefore, in this experiment, we included a measure of comprehension to assess higher-level processes (such as inferencing) associated with situation model activity (Graesser, Singer, & Trabasso, 1994). This allowed us to test whether comprehension skills influence readers' processing of dimension shifts.

Method

Participants

Seventy-six adults and 71 twelve-year-olds (mean age = 12.68 years, $SD = 0.49$) participated in this experiment. All participants were native English speakers.

Materials

Stories Four narratives from *Aesop's fables* (1975) were adapted by replacing archaic language and making them age appropriate (mean Flesch–Kincaid grade level = 6.5). These texts were chosen to be similar to the materials utilized in Zwaan et al. (1998). Each story was approximately 300 words long. The stories were “The Town Mouse and the Country Mouse,” “The Donkey Carrying Salt,” “The Miller, His Son, and Their Donkey,” and “The Old Woman and the Physician.”

The stories were broken down into clauses, which were analyzed for shifts in each of the five dimensions (1 = *shift*, 0 = *no shift*; Zwaan, Magliano, & Graesser, 1995). The clauses were also coded for control variables known to predict reading times, including the number of syllables per clause, clause position within the text, argument overlap, the number of new argument nouns, and the number of infrequent vocabulary words (low frequency < 250) (Kučera & Francis, 1967). The texts were counterbalanced to be presented in different orders across four text order conditions.

Two trained analysts coded for dimension shifts and control variables, using Zwaan et al.'s (1998) criteria, and agreement was high, $.80 < ks < .92$. Categorical correlations were computed for the dimensions. Shifts in time were correlated with shifts in space ($\phi = .27, p < .01$) and characters ($\phi = .24, p < .01$), and shifts in space were correlated with shifts in characters ($\phi = .19, p < .05$). None of the other dimensions were correlated ($-.11 < \phi < .14, ps > .05$).

For each story, participants answered two true/false comprehension questions. These questions assessed the participants' explicit memory for the main events in each story and were included to ensure that participants read each text.

Comprehension Participants completed the comprehension subsection of the Gates MacGinitie Reading Test (adults, level AR; children, level 7/9) (MacGinitie, MacGinitie, Maria, & Dreyer, 2000). This standardized test assesses the ability to understand prose with multiple-choice questions that require an understanding of explicit and implicit text information. Participants were given 35 min to complete the test.

Working memory Participants completed the sentence span task (Swanson et al., 1989), as described in Experiment 1.

Apparatus

The experiment was run on a Dell computer using E-Prime software. Participants were seated in front of a color monitor with their right hand resting on the mouse. The text was centered on the screen in standard upper- and lower case type.

Procedure

Participants were randomly assigned to one of the four text order conditions. They read each story, clause-by-clause and at their own pace, on a computer screen. They proceeded to the next screen by pressing the mouse button, and reading times were collected. They could not go back and reread any section of the story. After each text, the experimenter asked the participants two true/false questions about the text and did not provide feedback. The task took participants between 15 and 30 min. After reading the narratives, participants completed the sentence span task, followed by the Gates MacGinitie test.

Results

A series of multi-level analyses (clauses nested within participants) were performed using HLM (Raudenbush & Bryk, 2002). In these analyses, reading times per clause served as the dependent variable. The predictor variables were the control variables (number of syllables per clause, clause position, argument overlap, number of new argument nouns, number of infrequent vocabulary words, and three dummy variables coded for text), the dimension variables (shifts in time, space, causation, goals, and characters), and the individual difference variables (working memory and the Gates MacGinitie comprehension scores).

Because of the number of parameters, an alpha level of $\alpha = .01$ was set. No serious evidence of non-normality or non-linearity was found. However, using Bartlett's test of level-1 variance ($ps < .001$ for all steps of the models), unequal variances between variables were problematic even after removing outliers and transformation attempts. Because the data were normally distributed and because no other assumptions were violated, we opted to use the original variables for ease of interpretation; this also led us to use the more conservative estimation method that does not assume robust standard errors. Finally, since reading time data tend to have substantial intraclass correlations (Richter, 2006), maximum likelihood estimation methods were used. Full maximum likelihood was used so as to compare the deviance scores (D) in nested models that did not contain the same fixed parts (Raudenbush & Bryk, 2002). These scores allowed us to determine whether adding new variables to the model leads to a better fit. Finally, the adult and child data were modeled separately in order to more clearly assess qualitative differences between the two groups.

Reading times greater than three standard deviation units above the overall mean or shorter than 200 ms were removed. Reading times were also removed if a participant answered fewer than 80% of the comprehension questions correctly. This resulted in a loss of less than 2% of the data.

The average reading time per clause was $M = 2,404.63$ ms ($SD = 1,018.80$) for the adults and $M = 3,666.90$ ms ($SD = 1,810.05$) for the children. The average percentile score on the Gates was 71.13 ($SD = 16.76$) for the adults and 54.77 ($SD = 20.18$) for the children. The average working memory scores were $M = 20.72$ ($SD = 3.62$) words correctly recalled for the adults and $M = 17.71$ ($SD = 3.88$) words for the children.

Reading times To ensure that there was enough variability to justify further modeling, an unconditional model was fit. For the adults, this was significant $\hat{\gamma}_{00} = 2,419.67$, $SE = 48.42$, $p < .001$. There was significant variation among the reading times, $\chi^2 = 1,966.19$, $p < .001$, with $\rho = 16\%$ of the variation attributable to between-person variation. For the children, this was also significant, $\hat{\gamma}_{00} = 3,668.10$, $SE = 112.05$, $p < .001$. There was significant variation among their mean reading times, $\chi^2 = 2,688.81$, $p < .001$, with $\rho = 23\%$ attributable to between-person variation. Further modeling was therefore justified.

First, the text control variables were added to the level-1 model. For the adults, all of the controls except argument overlap predicted reading times ($ps < .001$). Reading times increased when there were more syllables per clause, $\hat{\gamma}_{10} = 104.90$, $SE = 3.24$, $p < .001$, at the beginning of the text, $\hat{\gamma}_{20} = -6.58$, $SE = 1.08$, $p < .001$, when new argument nouns were introduced, $\hat{\gamma}_{30} = 88.57$, $SE = 7.92$, $p < .001$, and when infrequent vocabulary

words were included, $\hat{\gamma}_{50} = 86.14$, $SE = 6.70$, $p < .001$. There was a marginally significant effect such that reading times increased when there was no argument overlap, $\hat{\gamma}_{40} = 39.35$, $SE = 15.11$, $p < .05$. All of the dummy variables coded for texts were significant, $\hat{\gamma}_s \geq 106.64$, $SEs < 27.90$, $ps < .001$. There was significant variation among the mean intercepts, $\chi^2 = 144.27$, $p < .001$, with $\rho = 27\%$ attributable to between-person variation, providing evidence that there was still variation to be accounted for. Adding these variables improved the fit of the model, $D(52) = 5,642.21$, $p < .01$.

Similar to the adults, when the text controls were added to the level-1 model for the children, reading times increased when there were more syllables per clause, $\hat{\gamma}_{10} = 193.37$, $SE = 9.20$, $p < .001$, at the beginning of the text, $\hat{\gamma}_{20} = -6.19$, $SE = 2.01$, $p < .01$, when there was a lack of argument overlap, $\hat{\gamma}_{40} = -80.05$, $SE = 29.07$, $p < .01$, and with infrequent vocabulary words, $\hat{\gamma}_{50} = 146.75$, $SE = 16.36$, $p < .001$. There were no significant effects for new argument nouns and the text dummy variables, $\hat{\gamma}_s \leq 108.58$, $SEs < 56.12$, $ps > .04$. There was significant variation among the reading time mean intercepts, $\chi^2 = 130.28$, $p < .001$, with $\rho = 43\%$ attributable to between-person variation. Adding these variables improved the fit of the model, $D(52) = 6,519.35$, $p < .01$.

Next, the dimension shifts were added to the level-1 model. (See Table 2 for the results of the adult and child models.) For the adults, time, space, and causal shifts predicted increases in reading times ($ps \leq .001$), but character and goal shifts did not. There was significant variation among the mean intercepts, $\chi^2 = 150.25$, $p < .001$, with $\rho = 29\%$ attributable to between-person variation. Adding these variables improved the fit of the model, $D(65) = 226.77$, $p < .01$.

For the children, shifts in causation predicted increases in reading times, $\hat{\gamma}_{120} = 157.24$, $p < .001$, whereas shifts in space ($p < .001$) predicted decreases in reading times. In addition, shifts in time, characters, and goals were marginally significant predictors of decreases in reading times at the $\alpha = .01$ level ($ps < .05$). There was significant variation among the intercepts of the mean reading times, $\chi^2 = 128.72$, $p < .001$, with $\rho = 46\%$ attributable to between-person variation. Adding these variables improved the fit of the model, $D(65) = 132.63$, $p < .01$.

In order to directly address differences in dimension monitoring between the adults and children, t -tests were run comparing the slopes of the adults versus children for each dimension. Adults were more likely to show increased reading times for shifts in time, $t(143) = 3.98$, $p < .001$, $D = 0.66$, and space, $t(143) = 8.02$, $p < .001$, $D = 1.32$. There were marginally significant effects such that adults were more likely to show increased reading times for shifts in characters, $t(143) = 1.86$, $p < .05$, $D = 0.31$, and goals, $t(143) = 2.00$, $p < .05$.

Table 2 Experiment 2: Summary of HLM analyses for variables predicting reading times

	Controls + Dimensions				Controls, Dimensions, Working Memory, Gates			
	Adults		Children		Adults		Children	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Fixed effects								
Intercept, γ_{00}	865.45***	58.42	1,125.33***	96.54	1,034.63**	367.45	1,838.05***	408.70
Gates, γ_{01}					−3.59	5.60	−11.03*	4.82
WM, γ_{02}					4.16	23.32	−6.10	20.20
Syllables, γ_{10}	99.26***	3.08	199.33***	9.47	157.01***	14.49	354.94***	31.03
Gates, γ_{11}					−.42*	.19	−2.55***	.30
WM, γ_{12}					−1.36	.88	−.89	1.58
Clause position, γ_{20}	−5.61***	1.07	−4.70*	1.91	−6.66	7.51	−7.22	7.83
Gates, γ_{21}					−.09	.06	.16	.09
WM, γ_{22}					.38	.38	−.34	.41
New nouns, γ_{30}	86.73***	7.96	6.61	14.97	105.42*	41.03	112.14	67.73
Gates, γ_{31}					−0.36	.57	.09	.67
WM, γ_{32}					.33	2.39	−6.25	3.87
Argument overlap, γ_{40}	11.01	15.55	−93.32**	29.62	118.09	100.98	−108.94	149.98
Gates, γ_{41}					−1.05	1.30	−.16	1.68
WM, γ_{42}					−1.55	5.31	1.39	7.42
Infrequent words, γ_{50}	96.72***	6.71	142.50***	16.65	42.58	33.37	−100.46	52.12
Gates, γ_{51}					.33	.40	1.90**	.59
WM, γ_{52}					1.46	1.86	7.84**	2.54
Dummy1, γ_{60}	197.59***	29.90	152.18**	55.18	−8.54	198.33	−69.98	185.28
Gates, γ_{61}					.12	2.00	−3.37	3.92
WM, γ_{62}					9.54	8.26	22.97*	10.75
Dummy2, γ_{70}	229.21***	27.12	65.83	55.06	−73.27	176.17	−551.73*	208.10
Gates, γ_{71}					.60	2.04	2.86	3.01
WM, γ_{72}					12.53	8.75	26.02*	11.22
Dummy3, γ_{80}	153.89***	27.57	126.49**	42.01	−60.55	182.22	−58.49	169.48
Gates, γ_{81}					−.13	2.16	−.08	1.99
WM, γ_{82}					10.77	8.61	10.69	8.75
Time, γ_{90}	100.17***	22.25	−79.59*	39.30	198.01	150.17	−100.22	168.43
Gates, γ_{91}					.68	1.73	1.38	1.95
WM, γ_{92}					−7.03	7.23	−3.11	8.84
Space, γ_{100}	190.99***	21.50	−169.60***	39.48	524.64***	116.74	−73.13	189.93
Gates, γ_{101}					−2.22	1.73	.54	2.13
WM, γ_{102}					−8.50	6.54	−7.12	10.56
Characters, γ_{110}	5.06	29.42	−95.02*	45.24	−470.69**	149.38	−279.59	190.03
Gates, γ_{111}					−.56	1.70	.65	2.52
WM, γ_{112}					24.86**	8.81	8.42	10.34
Causation, γ_{120}	142.08***	14.67	157.24***	24.58	222.25*	108.96	30.54	136.77
Gates, γ_{121}					−.27	.96	1.01	1.30
WM, γ_{122}					−2.96	4.95	4.04	6.48
Goals, γ_{130}	3.27	15.94	−62.48*	28.68	−.90	100.21	−361.22*	134.79
Gates, γ_{131}					−.46	.90	1.15	1.40
WM, γ_{132}					1.79	5.25	13.31	6.87

Coef. = slope coefficient; SE = standard error

* $p < .05$, ** $p < .01$, *** $p < .001$

.05, $D = 0.33$. There was no significant effect for causation, $t(143) = 0.53$, $p > .05$, $D = 0.09$.

Finally, the Gates and working memory scores were added to the level-2 models. (Again, see Table 2 for the results for both the adults and children.) For the adults, the Gates scores and working memory scores did not predict reading times as main effects. However, there was one significant cross-level interaction (at an alpha level set to .01): Higher working memory was associated with reductions in the effects of character shifts on reading times, $\hat{\gamma}_{112} = 24.86$, $p < .01$. There was significant variation among the mean intercepts, $\chi^2 = 149.86$, $p < .001$, with $\rho = 26\%$ attributable to between-person variation. Adding these variables did not improve the fit of the model, $D(28) = 41.89$, $p > .01$ (although $p < .05$).

For the children, the Gates and working memory scores did not predict reading times as main effects at the alpha level of .01. However, children with higher Gates scores had shorter readings times, $\gamma_{01} = -11.03$, $p < .05$. There were three significant cross-level interactions (at $\alpha = .01$). Higher Gates scores reduced the effects of the number of syllables on reading times, $\hat{\gamma}_{11} = -2.55$, $p < .001$. Also, higher working memory, $\hat{\gamma}_{52} = 7.84$, $p < .01$, and Gates, $\hat{\gamma}_{51} = 7.84$, $p < .01$, scores reduced the effects of vocabulary on reading times. There was significant variation among the intercepts of the reading time means, $\chi^2 = 117.45$, $p < .001$, with $\rho = 25\%$ attributable to between-person variation. Adding these variables improved the fit of the model, $D(28) = 75.90$, $p < .01$.

Discussion

The data replicated Zwaan, Magliano and Graesser's (1995; Zwaan et al., 1998) findings that adult reading times increase when dimension shifts are encountered. Contrary to the hypotheses, however, adults' reading times did not increase for character or goal shifts. One potential explanation is that these dimension shifts may not have been as salient in the texts as the other dimension types. Previous research has shown that manipulating the salience of a dimension interacts with reading behaviors for other dimensions (Levine & Klin, 2001). The stories used in this experiment were adapted from *Aesop's Fables*, which tend to be relatively short and simple, contain one or two main characters, and have a structure with one main goal and few subgoals. Therefore, shifts in these dimensions may have been less distinct or provided relatively minor event boundaries. The texts used in Zwaan et al. (1998) were also adapted from *Aesop's Fables* (although different fables were used), and those texts showed minor effects for the spatial dimension.

Like the adults, the children's reading times increased at causal shifts. In contrast to the adults, children's reading times decreased at spatial, goal, character, and time shifts (although the latter three were marginally significant). The finding that children's readings times actually decreased at dimension shifts was surprising. One explanation could be that dimension monitoring can be captured whenever there is a change in reading times in response to a shift, regardless of whether those changes signify speedups or slowdowns. However, this contradicts the structure-building framework (Gernsbacher, 1990, 1997): Dimension shifts represent new information that is inconsistent with the reader's current representation of the text. The difficulty associated with integration should result in increased processing times. That the children's reading times actually decreased would suggest, according to the framework, that the information was more easily integrated and, thus, processed more quickly, or that no integration was attempted. Consistent with this interpretation, decreases in reading times at dimension shifts have been found in previous work (Radvansky, Zwaan, Curiel, & Copeland, 2001), in which adults would intentionally choose to not update their situation models along certain dimensions. Readers can also ignore or remain unaware or unconcerned with those dimensions, and similar activities may have occurred with the child participants here.

As in Experiment 1, comprehension and working memory scores, in general, did not interact with children's or adults' propensities to monitor dimension shifts during moment-by-moment reading. Rather, comprehension and working memory interacted with surface-level predictors of reading times, such as the number of syllables and vocabulary.

Experiment 3

The results from Experiment 2 indicate potential differences between adults and children with regard to moment-by-moment dimension monitoring processes. However, the texts used in Experiment 2 were at a seventh-grade reading level, making them developmentally appropriate for the children but possibly too easy for the adults. Text difficulty, unsurprisingly, plays a crucial role in comprehension (Linderholm, Everson, van den Broek, Mischinski, Crittenden, & Samuels, 2000; McDaniel, Hines, & Guynn, 2002; McNamara, 2001; O'Connor, Bell, Harty, Larkin, Sackor, & Zigmond, 2002). For example, when texts are too difficult, readers must focus on both decoding and determining meaning; if texts are too easy, readers may gloss over or ignore elements of texts, potentially influencing their processing of event shifts. This could influence the cognitive resources available for making inferential connections and tracking dimensions during comprehension.

Because of the mismatch between reading skill and text level for adult participants, it remained an open question whether the results from [Experiment 2](#) could be attributed to developmental differences in dimension monitoring, rather than to a mismatch in the relative difficulty of the text. [Experiment 3](#) addressed this issue by presenting texts to adults that were appropriately difficult. If the adult patterns of reading times are similar to those obtained in [Experiment 2](#), it would suggest that the differences are indeed developmentally driven. However, if the adult patterns of reading times change, and perhaps look more like children's reading times, it would indicate that the previously obtained results may have been a function of relative text difficulty.

Method

Participants

Forty-one adults participated in this experiment. All participants were native English speakers.

Materials

Stories The four narratives from *Aesop's Fables* (1975) used in [Experiment 2](#) were adapted to be more difficult and age appropriate for adults by adding more difficult vocabulary words and grammatical complexities (mean Flesch–Kincaid grade level = 13.6). These texts contained all of the same information as the texts in [Experiment 2](#), with the content presented in the same order and the length of each text remaining at approximately 300 words. The stories were broken down into the same clauses as in [Experiment 2](#), and thus the dimension shifts occurred in the same clauses (Zwaan, Magliano, & Graesser, 1995). The clauses were also coded for control variables known to predict reading times, including the number of syllables per clause, clause position within the text (which remained the same), argument overlap, the number of new argument nouns, and the number of infrequent vocabulary words (low frequency < 250) (Kučera & Francis, 1967).

In addition, a fifth text, entitled “The City of Political Distinction” (Bierce, 2007), was added to increase the number of observations and the potential generalizability of any obtained results. This text, like the other materials, was a fable with a similar structure that contained shifts in each of the five dimensions. It was modified slightly to bring it to a comparable word length and difficulty as the other texts (340 words, Flesch–Kincaid grade level = 12.6). The story was broken down into clauses, and two raters coded for both the control variables and dimension shifts in the

same manner as with the other texts. Interrater reliability was good, $.81 < ks < .95$.

The texts were counterbalanced to be presented in different orders across five text order conditions. Following each story, participants answered two true/false comprehension questions. These questions assessed participants' explicit memory for the main events in each story and were included to ensure that participants read each text.

Apparatus

The experiment was run on a Dell computer using E-Prime software. Participants were seated in front of a color monitor with their right hand resting on the mouse. The text was centered on the screen in standard upper- and lower case type.

Procedure

Participants were randomly assigned to one of the five text order conditions. They read each story clause-by-clause and at their own pace on a computer screen. They proceeded to the next screen by pressing the mouse button, and reading times were collected. They could not go back and reread any section of the story. After each text, participants answered two true/false questions about the text on the computer, without feedback. The task took participants between 15 and 30 min.

Results

The average reading time per clause was $M = 2,312.27$ ms ($SD = 1,483.85$). Reading times greater than three standard deviation units above the overall mean or shorter than 200 ms were removed. Reading times were also removed if a participant answered fewer than 80% of the comprehension questions correctly. This resulted in a loss of less than 2% of the data.

As in [Experiment 2](#), a series of multi-level analyses (clauses nested within participants) were performed using HLM, (Raudenbush & Bryk, 2002). The dependent and predictor variables remained the same, with the exception that this analysis included four dummy variables coded for text. As in [Experiment 2](#), an alpha level of $\alpha = .01$ was set, and full maximum likelihood was utilized. Bartlett's test of level-1 variance ($ps < .001$ for all steps of the models) indicated unequal variances between variables even after removing outliers and transformation attempts. However, no serious evidence of non-normality or non-linearity was found, so we opted to use the original variables for ease of interpretation.

To ensure that there was enough variability to justify further modeling, an unconditional model was fit, and this

was significant, $\hat{\gamma}_{00} = 2,315.97$, $SE = 111.15$, $p < .001$. There was significant variation among the reading times ($\chi^2 = 2,240.12$, $p < .001$), with $\rho = 23\%$ of the variation attributable to between-person variation. Further modeling was justified.

First, the text control variables were added to the level-1 model. All of the controls significantly predicted reading times ($ps < .001$). Reading times increased when there were more syllables per clause, $\hat{\gamma}_{10} = 92.10$, $SE = 4.83$, $p < .001$, as the clause position decreased, $\hat{\gamma}_{20} = -8.67$, $SE = 1.45$, $p < .001$, when new argument nouns were introduced, $\hat{\gamma}_{30} = 72.76$, $SE = 15.67$, $p < .001$, when there was no argument overlap, $\hat{\gamma}_{40} = 183.28$, $SE = 32.10$, $p < .001$, and when infrequent vocabulary words were included, $\hat{\gamma}_{50} = 157.32$, $SE = 12.81$, $p < .001$. All of the text dummy variables were significant, $-486.70 \leq \hat{\gamma}s \leq 628.47$, $SEs < 68.24$, $ps < .001$. There was significant variation among the mean intercepts, $\chi^2 = 211.20$, $p < .001$, with $\rho = 28\%$ attributable to between-person variation, providing evidence that there was still variation to be accounted for. Adding these variables improved the fit of the model, $D(63) = 5,107.15$, $p < .001$.

Next, the dimension shifts were added to the level-1 model (see Table 3 for the results). Time, space, and causal shifts predicted increases in reading times ($ps \leq .01$), but character and goal shifts did not. Character shifts predicted decreases in reading times ($p < .01$). There was significant variation among the mean intercepts, $\chi^2 = 189.73$, $p < .001$, with $\rho = 28\%$ attributable to between-person variation.

Table 3 Experiment 3: Summary of HLM analyses for variables predicting adults' reading times

	Coef.	SE
Fixed Effects		
Intercept, γ_{00}	954.16	98.29***
Syllables, γ_{10}	88.03	4.44***
Clause-position, γ_{20}	-7.98	1.37***
New nouns, γ_{30}	66.83	13.93***
Argument overlap, γ_{40}	140.80	26.38***
Infrequent words, γ_{50}	169.98	12.25***
Dummy1, γ_{60}	-307.80	60.13***
Dummy2, γ_{70}	-444.30	46.71***
Dummy3, γ_{80}	-433.21	68.34***
Dummy4, γ_{90}	-623.28	67.89***
Time, γ_{100}	84.14	29.66**
Space, γ_{110}	73.23	30.12**
Characters, γ_{120}	-146.84	48.65**
Causation, γ_{130}	184.23	27.44***
Goals, γ_{140}	-64.08	-64.08

Coef. = slope coefficient; SE = standard error

* $p < .05$, ** $p < .01$, *** $p < .001$

Adding these variables improved the fit of the model, $D(70) = 161.25$, $p < .001$.

Discussion

The data replicated the findings from Experiment 2, such that adult reading times increased at shifts in time, space, and causality. This also replicates Zwaan, Magliano and Graesser's (1995; Zwaan et al., 1998) findings that adult reading times increase when dimension shifts are encountered. Additionally, reading times did not increase at shifts in characters¹ or goals, replicating the findings from Experiment 2. These data suggest that text difficulty did not play a role in the monitoring patterns of adults.

Experiment 4

To further support the previous developmental findings, we sought to replicate the children's data. We also took the opportunity to utilize the same set of materials in order to examine participants' verb-clustering and reading time behaviors in a single experiment. Besides offering an additional means of ensuring that any obtained effects were not simply the result of the different materials employed across Experiments 1 and 2, a combined experiment of this type offered the opportunity to examine both process- and product-oriented measures with a single group of participants.

¹ One intriguing finding was that reading times decreased for character shifts, similar to the effects observed for children in Experiment 2. To assess why character shifts potentially predicted decreases in reading times, a brief follow-up study was conducted. Twenty-nine adults read each of the texts and rank ordered the dimensions in terms of how difficult it was to pay attention to each dimension, such that 1 was *easiest* and 5 was *hardest*. The order of the stories was counter-balanced across participants. The Friedman test indicated that there were differences between the rankings of the dimensions, $\chi^2(4) = 43.67$, $p < .001$. Wilcoxon post hoc tests indicated that characters were rated as easier to follow than all of the other dimensions, $zs > 3.70$, $ps < .001$. There were no other differences, $zs < 1.69$, $ps > .05$. This evidence is consistent with the notion that readers may have strategically focused their attention on dimensions that they considered to be more difficult to track. This is also consistent with the work of Radvansky et al. (2001), who found that readers would deliberately skim and, therefore, not update their event models along certain dimensions. Readers may opt to ignore elements of texts that they find uninteresting or less than useful for their processing (e.g., van den Broek, Lorch, Linderholm, & Gustafson, 2001), which may involve modifications to the tracking processes associated with event indexing. These modifications can therefore result in null or negative associations between event shifts and reading times (Radvansky & Copeland, 2010).

Child participants read narratives and afterward grouped verbs from each story into related pairings, as in [Experiment 1](#). These children also read the *Aesop's Fables* (1975), previously employed in [Experiment 2](#), and their reading times were analyzed in relation to shifts in time, space, characters, goals, and causation. In addition, the same children also grouped verbs from *Aesop's Fables*. We hypothesized that children would pair verbs that represent continuities across dimensions, consistent with [Experiment 1](#). We also hypothesized that children's reading times would increase for causal shifts, but either would not increase or would decrease for shifts in the other dimensions, consistent with [Experiment 2](#).

Method

Participants

Forty-one children (mean age = 12.82 years, $SD = 0.35$) participated in this experiment. However, three were not native English speakers, so their data were removed.

Materials

For the verb-clustering task, each participant read the four stories from [Experiment 1](#) and the four stories from [Experiment 2](#). For the reading time task, each participant read the four stories from [Experiment 2](#).

With regard to the verb-clustering task using the materials from [Experiment 2](#), 10 unique verbs were selected from each story. Two trained coders rated every possible pairing of the 10 verbs (a total of 45 pairs per story) for continuity in each event-indexing dimension (1 = *shift*, 0 = *no shift*), surface distance, surface connections, and argument overlap. Inter-rater agreement was high: $.88 < ks < 1.00$.

In addition, we asked 25 adults to pair the verbs without reading the narratives. These pairings provided an index of the likelihood that verbs would be paired on the basis of general lexical knowledge, rather than story comprehension (Zwaan, Langston, & Graesser, 1995), and were used as a control in our examination of participant performance.

Procedure

All participants were tested in groups of 10–15 students in their language arts classroom at school over three sessions (to accommodate the teachers' course schedules). In the first session, the participants completed the verb-clustering task, using the same materials and method as in [Experiment 1](#). In the second session, which took place the following day, the participants completed the reading time task, using the same materials and method as in [Experiment 2](#). In the

third session, which occurred 2 weeks after the second session, the participants completed the verb-clustering task (using the method from [Experiment 1](#)) with the materials from [Experiment 2](#).

Results

Verb clustering

As in [Experiment 1](#), HGLM was utilized to account for the multi-level structure of the verb pairings nested within participants. All analyses were conducted in the same way as in [Experiment 1](#), such that the independent and dependent variables were the same, with the exception that working memory and fluency were not included in the analyses. The same steps in the model were utilized as in [Experiment 1](#). We first ran the analyses with the texts used in [Experiment 1](#). Next, we ran the same analyses with the texts used in [Experiment 2](#).

Experiment 1 Materials To assess the variation, unconditional models were estimated. The average log-odds of making a pairing was $\hat{\gamma}_{00} = -1.82$, $SE = .04$, $p < .001$. This result indicated that there was variability, thus justifying further modeling.

To assess whether the text controls predicted pairings, we added the control variables to the level-1 model. Children were more likely to pair verbs that had fewer intervening words, $\hat{\gamma}_{10} = -.01$, $SE = 0.003$, $p < .01$, occurred in the same sentence $\hat{\gamma}_{20} = 0.96$, $SE = 0.14$, $p < .001$, and were consistent with general lexical knowledge, $\hat{\gamma}_{40} = 1.41$, $SE = 0.18$, $p < .001$. The effect of argument overlap was not significant (although $\hat{\gamma}_{30} = 0.15$, $SE = 0.09$, $p = .10$).

To assess whether shifts in each dimension predicted pairings over and above the control variables, the five dimensions were added to the level-1 models (see [Table 4](#) for the results). Verbs that took place within a continuous event were more likely to be paired than verbs that were separated by shifts in time ($p < .05$), space ($p < .01$), characters ($p < .01$), and causation ($p < .001$). The effect of goals was not significant ($p > .05$, although the pattern was in the same direction).

Experiment 2 Materials To assess the variation, unconditional models were estimated. The average log-odds of making a pairing was $\hat{\gamma}_{00} = -1.74$, $SE = 0.03$, $p < .001$. This result indicated that there was variability, thus justifying further modeling.

To assess whether the text controls predicted pairings, we added the control variables to the level-1 model. Children were more likely to pair verbs that had fewer intervening words

$\hat{\gamma}_{10} = -0.01$, $SE = 0.001$, $p < .001$, occurred in the same sentence, $\gamma_{20} = 0.45$, $SE = 0.26$), (although $p < .09$), and were consistent with general lexical knowledge, $\hat{\gamma}_{40} = 2.74$, $SE = 0.20$, $p < .001$. The effect of argument overlap was not significant, $\hat{\gamma}_{30} = 0.02$, $SE = 0.06$, $p > .05$.

To assess whether shifts in each dimension predicted pairings over and above the control variables, the five dimensions were added to the level-1 models. (Again, see Table 4 for the results.) Verbs that took place within a continuous event were more likely to be paired than verbs that were separated by shifts in time ($p < .001$), space ($p < .001$), goals ($p < .01$), and causation ($p < .001$). The effect of characters was not significant ($p > .05$, although again the pattern was in the same direction).

Reading times

As in Experiment 2, a series of multi-level analyses (clauses nested within participants) were performed using HLM (Raudenbush & Bryk, 2002). All analyses were conducted in the same way as in Experiment 2, with the exception that working memory and comprehension were not included in the analyses. The same steps in the model were utilized as in Experiment 2. Reading times greater than three standard deviation units above the overall mean or shorter than 200 ms were removed. Reading times were also removed if a participant answered fewer than 80% of the comprehension questions correctly. This resulted in a loss of less than 3% of the data. The average reading time per clause was $M = 3,362.52$ ms ($SD = 1,795.82$).

To ensure that there was enough variability to justify further modeling, an unconditional model was fit. This was significant, $\hat{\gamma}_{00} = 3,392.83$, $SE = 134.03$, $p < .001$. There was significant

variation among the reading times, $\chi^2 = 920.94$, $p < .001$, with $\rho = 17\%$ of the variation attributable to between-person variation. Further modeling was therefore justified.

First, the text control variables were added to the level-1 model. All of the control variables except argument overlap and one of the text dummy variables predicted reading times ($ps < .05$). Reading times increased when there were more syllables per clause, $\hat{\gamma}_{10} = 162.01$, $SE = 10.72$, $p < .001$, at the beginning of the text, $\hat{\gamma}_{20} = -6.19$, $SE = 2.42$, $p < .02$, when new argument nouns were introduced, $\hat{\gamma}_{30} = 59.89$, $SE = 22.03$, $p = .01$, and when infrequent vocabulary words were included, $\hat{\gamma}_{50} = 204.24$, $SE = 10.71$, $p < .001$. The effect of argument overlap was not significant, $\hat{\gamma}_{40} = 23.68$, $SE = 42.31$, $p > .05$. Two of the dummy variables coded for texts were significant, $\hat{\gamma}_{60} = -184.37$, $SE = 68.50$, $p = .01$, and $\hat{\gamma}_{70} = 120.26$, $SE = 51.00$, $p < .03$, although one was not, $\hat{\gamma}_{80} = 19.41$, $SE = 71.67$, $p > .05$. There was significant variation among the mean intercepts, $\chi^2 = 61.17$, $p = .001$, with $\rho = 22\%$ attributable to between-person variation, providing evidence that there was still variation to be accounted for. Adding these variables improved the fit of the model, $D(52) = 2,513.98$, $p < .01$.

Next, the dimension shifts were added to the level-1 model. (See Table 5 for the results of this model.) Shifts in causation predicted increases in reading times, $\hat{\gamma}_{120} = 125.77$, $p < .01$, whereas shifts in space ($p < .001$) and shifts in time ($p < .05$) predicted decreases in reading times. The effects of characters and goals were not significant ($ps > .05$, although goal shifts predicted decreases in reading times at the $p < .10$ level). There was significant variation among the intercepts of the mean reading times, $\chi^2 = 57.22$, $p < .01$, with $\rho = 22\%$ attributable to between-person variation. Adding these

Table 4 Experiment 4: Summary of HGLM analyses for variables predicting children's verb pairings

	Experiment 1 Materials		Experiment 2 Materials	
	Coef.	SE	Coef.	SE
Fixed effects				
Intercept, γ_{00}	-1.44***	.14	-1.21***	.14
Surface dist., γ_{10}	.0002	.003	.001**	.0001
Surf. conn., γ_{20}	.60***	.14	.49**	.15
Arg. overlap, γ_{30}	.13	.10	-.14**	.05
Lexical, γ_{40}	1.21***	.20	2.11***	.16
Time, γ_{50}	-.26*	.12	-.36***	.08
Space, γ_{60}	-.28**	.09	-.18***	.05
Characters, γ_{70}	-.33**	.09	-.07	.05
Causation, γ_{80}	-.35***	.10	-.31***	.08
Goals, γ_{90}	-.07	.10	-.25***	.06

Coef. = slope coefficient; SE = standard error

* $p < .05$, ** $p < .01$, *** $p < .001$

variables did not improve the fit of the model, $D(65) = 76.23$ (although $p < .10$).

Discussion

The verb-clustering data from the two sets of materials replicated Zwaan, Langston and Graesser's (1995) findings, in that children were more likely to pair verbs that represented continuous, as compared with discontinuous, events along the dimensions of time, space, characters, causation, and goals. These data are also consistent with the results reported in Experiment 1, supporting the view that middle school children have relatively sophisticated understandings of each dimension (Case et al., 2001, 1996; van den Broek, 1997). Narrative dimensions influence readers' memory organization for narratives, even in the previously untested case of younger readers.

Consistent with Experiment 2, children's reading time data did not replicate previous adult findings (Zwaan, Magliano, & Graesser, 1995; Zwaan et al., 1998), which usually show relatively uniform increases in reading latencies for sentences containing dimension shifts. Although children's reading times increased at causal shifts, their reading times actually decreased at spatial and temporal shifts. This finding is consistent with the view that these shifts may have been more easily integrated or that no integration was attempted (Radvansky et al., 2001). By examining post-reading and moment-by-moment reading processes with a sample of children completing both types of measures, the findings further support the

commonalities and differences reported in the previous experiments for child and adult indexing patterns.

General discussion

The extant literatures document some of the similarities between adults' and children's constructions of situation models and their propensities toward resolving inconsistent events during narrative experiences. The literatures outline analogous processes with respect to how readers encode narrative information to build connections in memory between the "here-and-now" of the text and previous text information (Gernsbacher, 1990, 1997; Graesser, Louwerse, McNamara, Olney, Cai, & Mitchell, 2007; Kintsch & van Dijk, 1978; Oakhill, 1994; Pascual-Leone, 1970, 2000; Piaget & Inhelder, 2000; van den Broek, 1997). Furthermore, both literatures document the importance of time, space, characters, goals, and causation in narrative comprehension (Case et al., 2001; Zwaan, Langston, & Graesser, 1995; Zwaan, Magliano, & Graesser, 1995). However, the developmental literature has focused on children's understandings of each dimension, whereas the adult literature has additionally focused on how readers apply this knowledge by spontaneously monitoring narrative shifts. The present project attempted to bridge the understandings derived from these bodies of work.

The developmental literature led us to two different hypotheses. First, neo-Piagetian theories indicate that children should have sophisticated understandings of each dimension by late elementary grades (Case et al., 2001, 1996; van den Broek, 1997). This work has suggested that middle school children should monitor dimensions in a manner similar to that of adults. However, information-processing theories indicate that children's processing capacities continue to develop well into adolescence (Conklin et al., 2007; Gathercole et al., 2004; Luciana et al., 2005), suggesting that middle school children may not necessarily monitor dimensions in the same manner as adults, due to relative processing limitations. The present project compared these possibilities using both offline (Experiments 1 and 4) and online (Experiments 2 and 4) measures.

Experiments 1 and 4 revealed that adults' and 12-year-olds' memories were organized such that they were more likely to pair events that were continuous. Middle school children applied their knowledge of each dimension to their post-reading memory organization of the narratives. Experiments 2 and 4 examined whether children, like adults, monitor dimension shifts during moment-by-moment reading. While adults' reading times generally increased when the text described a dimension shift, children's reading times increased only at causal shifts

Table 5 Experiment 4: Summary of HLM analyses for variables predicting children's reading times

	Coef.	SE
Fixed effects		
Intercept, γ_{00}	987.59***	142.81
Syllables, γ_{10}	167.79***	11.23
Clause position, γ_{20}	-3.85	2.36
New Nouns, γ_{30}	54.12*	22.51
Argument overlap, γ_{40}	-24.93	41.43
Infrequent words, γ_{50}	198.24***	19.82
Dummy1, γ_{60}	-133.66†	79.14
Dummy2, γ_{70}	105.24†	52.66
Dummy3, γ_{80}	12.44	72.89
Time, γ_{90}	-117.37*	57.97
Space, γ_{100}	-285.76***	70.16
Characters, γ_{110}	84.65	78.46
Causation, γ_{120}	125.77**	39.63
Goals, γ_{130}	-57.66	37.46

Coef. = slope coefficient; SE = standard error

* $p < .05$, ** $p < .01$, *** $p < .001$

(and actually appeared to decrease for the remaining dimensions). Thus, while the experiments employing offline methodologies appeared consistent with neo-Piagetian accounts, the results from the online experiments support a view derived from information-processing theories (Conklin et al., 2007; Gathercole et al., 2004; Luciana et al., 2005).

Discrepancies between children and adult findings

Importantly, though, information-processing accounts would nevertheless be unlikely to predict the *decreases* observed in children's reading times to non-causal shifts. What might explain the conflicting results that children's monitoring behaviors match those of adults on measures of post-reading memory, but not when measured during moment-by-moment reading? One possibility is that both increases and decreases in reading times are indicative of dimension monitoring; that is, children's monitoring might reflect attempts at integration that, in some instances, lead to slowdowns and, in some instances, lead to speed-ups but, in both cases, result in established coherence between pre- and post-shift information. However, as was described earlier, the structure-building framework (Gernsbacher, 1990, 1997) indicates that processing should reflect necessary deliberative activity (i.e., processing slowdowns) when texts are difficult or complex, as they are when they contain shifts. Developmental research also contends that children build situation models using similar processes as adults. For example, children integrate new information by activating relevant knowledge and building connections between new information, prior text information, and prior knowledge (MacWhinney et al., 1989; Oakhill, 1994; Pascual-Leone, 1970, 2000; Piaget & Inhelder, 2000; van den Broek, 1997). Therefore, decreases in reading times might indicate that children did not make careful attempts to integrate these dimension shifts.

Another possible account of the conflicting adult/child effects might appeal to mismatches between the difficulty of the texts and the reading skills of adults versus children. Indeed, the texts used in Experiments 2 and 4 were at a grade level appropriate for middle school readers and, thus, would likely be less difficult for adults. To assess this possibility, in Experiment 3, adults read the same stories, albeit modified to be appropriate for their reading level. The results from Experiment 3 demonstrated that the moment-by-moment reading patterns did not change on the basis of the difficulty of the text (with the exception of character shifts). Therefore, the conflicting effects are unlikely to be a function of the ease or difficulty of the texts as a function of grade level.

Another explanation for the results obtained across the experiments, particularly with respect to children's relatively restricted moment-by-moment monitoring of dimensions,

might appeal to the index dominance hypothesis (Taylor & Tversky, 1997). Readers tend to index events along a dominant, organizing dimension and monitor events with other indices only when necessary. Perhaps unsurprisingly, causality is the dimension usually invoked as the dominant indexing framework. Children's processing has also often been considered with respect to focusing on a particular dominant feature or index; for example, consider the classic concept of decentering, in which children are able to focus on only one salient aspect of a problem at a time (Piaget & Inhelder, 2000). Previous work has shown that causality is indeed important to children's processing of the world (van den Broek, 1997). Research has also more generally supported the importance of causation as a primary organizing dimension in narrative comprehension (Trabasso, Secco, & van den Broek, 1984; Trabasso, van den Broek & Suh 1989; van den Broek, 1994). The texts used in the present study followed traditional narrative schemas in which causation and goals were the primary organizing features (e.g., Buss, Yussen, Mathews, Miller, & Rembold, 1983). Therefore, the organization of the texts and the relative importance of causation may have encouraged children to focus on that particular dimension, to the potential detriment of other dimensions.

Yet another possible explanation for the results is that children are capable of monitoring multiple dimensions but do not do so spontaneously during moment-by-moment reading. The experiments reported here assumed monitoring to be spontaneous and, therefore, did not attempt to manipulate whether participants might make strategic efforts to monitor particular shifts (Gerrig & O'Brien, 2005; Graesser et al., 1994; McKoon & Ratcliff, 1992, 1995; Singer, Graesser, & Trabasso, 1994; van den Broek, Rapp, & Kendeou, 2005). If children were instructed to monitor dimension shifts and, under such conditions, appeared similar to adults in terms of processing activity, it would provide evidence of middle school children's capacity for strategic, but not spontaneous, monitoring. Additional work from our labs plan to examine whether instructions to monitor dimensions can be influential with respect to children's strategic monitoring of dimensions (in the same way instructions can directly influence adult comprehension).

In the present study, fluency, comprehension, and working memory did not reliably predict dimension monitoring. In Experiment 1, fluency interacted with argument overlap, such that more fluent children were less likely to utilize argument overlap when making pairings. In Experiment 2, comprehension and working memory tended to interact with the surface-level predictors of monitoring among children, such as the number of syllables per clause. These findings are consistent with the work of Radvansky and Copeland (2004; Radvansky et al., 2001), which has

demonstrated effects of working memory on readers' updating of situation models.

Future directions

Since this project represents a first pass at documenting developmental differences in dimension monitoring, several crucial follow-ups will prove useful at further elucidating and informing these findings. First, different texts and samples were used across the present experiments. Future work should include participants of different ages engaging in online and offline tasks with the same texts to increase confidence in the results (along the lines of the method employed in [Experiment 4](#)). In addition, the materials in the present project were real-world texts that are commonly used in classroom settings. Since real-world texts often include all five dimensions in their content (Graesser & Ottati, 1995), the selected texts were considered appropriate materials for study. However, replication with different types of narratives, beyond fables, is needed to further confirm the generalizability of the present findings.

Additionally, clause-level reading time measures, in comparison with word- or syllable-level analyses, may not be precise enough to measure dimension-monitoring behaviors. Although reading times are often used in studies of moment-by-moment comprehension (Lorch, 1993; Radvansky et al., 2001; Zwaan et al., 1998), future work could also incorporate other measures of moment-by-moment processing, such as eye movements, probe latencies, or verbal protocols. In addition, future research should address other measures of offline mental organization. For example, post-reading lexical decision tasks can be helpful for defining mental maps (e.g., Curiel & Radvansky, 2002). Free recall and targeted comprehension questions can also provide converging evidence. Taken together, these additional methodologies will prove useful for documenting the developmental trajectories that guide readers' understanding and processing of dimensions during narrative comprehension. The present study takes a first step toward this goal, by demonstrating that (1) monitoring occurs spontaneously for middle school children during reading for only a subset of text dimensions (i.e., causality), yet (2) after reading, middle school children rely on multiple dimensions to organize memory for narratives.

References

- Aesop's fables*. (1975). New York, NY: Grosset & Dunlap
 Bierce, A. (2007). *Fantastic fables*. Retrieved from <http://www.gutenberg.org/etext/374>.

- Britton, B. K., & Gulgoz, 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, 329–345.
- Buss, R., Yussen, S. R., Mathews, S. R., Miller, G. E., & Rembold, K. L. (1983). Development of children's use of a story schema to retrieve information. *Developmental Psychology*, 19, 22–28.
- Calvo, M. G. (2001). Working memory and inferences: Evidence from eye fixations during reading. *Memory*, 9, 365–381.
- Calvo, M. G. (2005). Relative contribution of vocabulary knowledge and working memory span to elaborative inferences in reading. *Learning and Individual Differences*, 15, 53–65.
- Case, R., Demetriou, A., Platsidou, M., & Kazi, S. (2001). Integrating concepts and tests of intelligence from the differential and developmental traditions. *Intelligence*, 29, 307–336.
- Case, R., Okamoto, Y., Griffin, S., McKeough, A., Bleiker, C., Henderson, B., et al. (1996). The role of central conceptual structures in the development of children's thought. *Monographs of the Society for Research in Child Development*, 61, 1–295.
- Chall, J. S. (1983). *Stages of reading development*. New York, NY: McGraw-Hill.
- Collins, W. M., & Levy, B. A. (2008). Developing fluent text processing with practice: Memorial influences on fluency and comprehension. *Canadian Psychology*, 49, 133–139.
- Conklin, H. M., Luciana, M., Hooper, C. J., & Yarger, R. S. (2007). Working memory performance in typically developing children and adolescents: Behavioral evidence of protracted frontal lobe development. *Developmental Neuropsychology*, 31, 103–128.
- Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, 99, 311–325.
- Curiel, J. M., & Radvansky, G. A. (2002). Mental maps in memory retrieval and comprehension. *Memory*, 10, 113–126.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52, 219–232.
- Duke, N., Pressley, M., & Hilden, K. (2004). Difficulties with reading comprehension. In C. A. Stone, E. Silliman, B. Ehren, & K. Apel (Eds.), *Handbook of language and literacy* (pp. 501–520). New York, NY: Guilford.
- Fletcher, H. R. (1994). Levels of representation in memory for discourse. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 589–607). San Diego, CA: Academic Press.
- Friedman, N. P., & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods*, 37, 581–590.
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, 40, 177–190.
- Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.
- Gernsbacher, M. A. (1997). Two decades of structure building. *Discourse Processes*, 23, 265–304.
- Gerrig, R. J., & O'Brien, E. J. (2005). The scope of memory-based processing. *Discourse Processes*, 39, 225–242.
- Graesser, A. C. (1981). *Prose comprehension beyond the word*. New York, NY: Springer.
- Graesser, A. C., & Clark, L. F. (1985). *Structures and procedures of implicit knowledge: Advances in discourse processes, vol 17*. Norwood, NJ: Ablex.
- Graesser, A. C., Louwerse, M. M., McNamara, D. S., Olney, A., Cai, Z., & Mitchell, H. H. (2007). Inference generation and cohesion in the construction of situation models: Some connections with

- computational linguistics. In F. Schmalhofer & C. A. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes* (pp. 289–310). Mahwah, NJ: Erlbaum.
- Graesser, A. C., & Ottati, V. (1995). Why stories? Some evidence, questions, and challenges. In R. S. Wyer & T. K. Srull (Eds.), *Advances in social cognition* (pp. 121–132). Hillsdale, NJ: Erlbaum.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371–395.
- Hagaman, J. L., & Reid, R. (2008). The effects of paraphrasing strategy on the reading comprehension of middle school students at risk for failure in reading. *Remedial and Special Education*, 29, 222–234.
- Jacobs, V. A. (2008). Adolescent literacy: Putting the crisis in context. *Harvard Educational Review*, 78, 7–39.
- Jahn, G. (2004). Three turtles in danger: Spontaneous construction of causally relevant spatial situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 969–987.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394.
- Kučera, J., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Kuhn, M. R., & Stahl, S. A. (2003). *Fluency: A review of development and remedial practices*. Ann Arbor, MI: Center for the Improvement of Early Reading Achievement.
- Levine, W. H., & Klin, C. M. (2001). Tracking of spatial information in narratives. *Memory & Cognition*, 29, 327–335.
- Linderholm, T., Cong, X., & Zhao, Q. (2008). Differences in low and high working memory capacity readers' cognitive and metacognitive processing patterns as a function of reading for different purposes. *Reading Psychology*, 29, 61–85.
- Linderholm, T., Everson, M. G., van den Broek, P., Mischinski, M., Crittenden, A., & Samuels, J. (2000). Effects of causal text revisions on more- and less-skilled readers' comprehension of easy and difficult texts. *Cognition and Instruction*, 18, 525–556.
- Lorch, R. F. (1993). Integration of topic and subordinate information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1071–1081.
- Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 149–157.
- Luciana, M., Conklin, H. M., Hooper, C. J., & Yarger, R. S. (2005). The development of nonverbal working memory and executive control processes in adolescents. *Child Development*, 76, 697–712.
- MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2000). *Gates-MacGinitie reading tests* (Fourth ed.). Itasca, IL: Riverside.
- MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. (1989). Language learning: Cues or rules? *Journal of Memory and Language*, 28, 255–277.
- Magliano, J. P., & Graesser, A. C. (1991). A three-pronged approach for studying inference generation in literary text. *Poetics*, 20, 193–232.
- Marston, D. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18–78). New York, NY: Guilford.
- McDaniel, M. A., Hines, R. J., & Gynn, M. J. (2002). When text difficulty benefits less-skilled readers. *Journal of Memory and Language*, 46, 544–561.
- McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99, 440–466.
- McKoon, G., & Ratcliff, R. (1995). The minimalist hypotheses: Directions for research. In C. A. Weaver, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 97–116). Hillsdale, NJ: Erlbaum.
- McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, 55, 51–62.
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14, 1–43.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.
- O'Connor, R. E., Bell, K. M., Harty, K. R., Larkin, L. K., Sackor, S. M., & Zigmond, N. (2002). Teaching reading to poor readers in the intermediate grades: A comparison of text difficulty. *Journal of Educational Psychology*, 94, 474–485.
- Oakhill, J. (1994). Individual differences in children's text comprehension. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 849–894). San Diego, CA: Academic Press.
- Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica*, 32, 301–345.
- Pascual-Leone, J. (2000). Reflections on working memory: Are the two models complementary? *Journal of Experimental Child Psychology*, 77, 138–154.
- Piaget, J., & Inhelder, B. (2000). *The psychology of the child*. New York, NY: Basic Books.
- Radvansky, G. A., & Copeland, D. E. (2004). Working memory span and situation model processing. *The American Journal of Psychology*, 117, 191–213.
- Radvansky, G. A., & Copeland, D. E. (2010). Reading times and the detection of event shift processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 210–216.
- Radvansky, G. A., Zwaan, R. A., Curiel, J. M., & Copeland, D. E. (2001). Situation models and aging. *Psychology and Aging*, 16, 145–160.
- Radvansky, G. A., Zwaan, R. A., Federico, T., & Franklin, N. (1998). Retrieval from temporally organized situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24 (5), 1224–1237.
- Rapp, D. N., & Taylor, H. A. (2004). Interactive dimensions in the construction of mental representations for text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 988–1001.
- Rapp, D. N., van den Broek, P., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading*, 11, 289–312.
- Rasinski, T. V., Padak, N., McKeon, C., Krug-Wilfong, L., Friedauer, J., & Heim, P. (2005). Is reading fluency a key for successful high school reading? *Journal of Adolescent and Adult Literacy*, 49, 22–27.
- Raudenbush, A. S., & Bryk, S. W. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.
- Richter, T. (2006). What is wrong with ANOVA and multiple regression? Analyzing sentence reading times with hierarchical linear models. *Discourse Processes*, 41, 221–250.

- Samuels, S. J. (1987). Information processing abilities and reading. *Journal of Learning Disabilities*, 20, 18–22.
- Siegler, R. S., & Alibali, M. W. (2005). Information-processing theories of development. In R. S. Siegler (Ed.), *Children's thinking* (4th ed., pp. 65–106). Upper Saddle River, NJ: Prentice Hall.
- Singer, M., Graesser, A. C., & Trabasso, T. (1994). Minimal or global inference during reading. *Journal of Memory and Language*, 33, 421–441.
- Sundermeier, B. A., & van den Broek, P. (2005). Causal coherence and the availability of locations and objects during narrative comprehension. *Memory & Cognition*, 33, 462–470.
- Swanson, H. L. (2008). Working memory and intelligence in children: What develops? *Journal of Educational Psychology*, 100, 581–602.
- Swanson, H. L., Cochran, K. F., & Ewers, C. A. (1989). Working memory in skilled and less skilled readers. *Journal of Abnormal Child Psychology*, 17, 145–156.
- Taylor, H. A., & Tversky, B. (1997). Indexing events in memory: Evidence for index dominance. *Memory*, 5, 509–542.
- Trabasso, T., Secco, T., & van den Broek, P. (1984). Causal cohesion and story coherence. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83–111). Hillsdale, NJ: Erlbaum.
- Trabasso, T., van den Broek, P., & Suh, S. Y. (1989). Logical necessity and transitivity of causal relations in stories. *Discourse Processes*, 12, 1–25.
- Tronsky, L. N. (2005). Strategy use, the development of automaticity, and working memory involvement in complex multiplication. *Memory & Cognition*, 33, 927–940.
- Underwood, T., & Pearson, P. D. (2004). Teaching struggling adolescent readers to comprehend what they read. In T. L. Jetton & J. A. Dole (Eds.), *Adolescent literacy research and practice* (pp. 35–61). New York, NY: Guilford.
- Vadasy, P. F., & Sanders, E. A. (2008). Benefits of repeated reading intervention for low-achieving fourth- and fifth-grade students. *Remedial and Special Education*, 29, 235–249.
- van den Broek, P. (1994). Comprehension and memory of narrative texts: Inferences and coherence. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 539–588). San Diego, CA: Academic Press.
- van den Broek, P. (1997). Discovering the cement of the universe: The development of event comprehension from childhood to adulthood. In P. van den Broek, P. J. Bauer, & T. Bourg (Eds.), *Developmental spans in event comprehension and representation: Bridging fictional and actual events* (pp. 321–343). Mahwah, NJ: Erlbaum.
- van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, 29, 1081–1087.
- van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes*, 39, 299–316.
- Williams, J. P. (2005). Instruction in reading comprehension for primary-grade students: A focus on text structure. *Journal of Special Education*, 39, 6–18.
- Wilson, S. G., Rinck, M., McNamara, T. P., Bower, G. H., & Morrow, D. G. (1993). Mental models and narrative comprehension: Some qualifications. *Journal of Memory and Language*, 32, 141–154.
- Wood, D. E. (2006). Modeling the relationship between oral reading fluency and performance on a state wide reading test. *Educational Assessment*, 11, 85–104.
- Zelazo, P. D. (2004). The development of conscious control in childhood. *Trends in Cognitive Sciences*, 8, 12–17.
- Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1196–1207.
- Zwaan, R. A. (1999). Situation models: The mental leap into imagined worlds. *Current Directions in Psychological Science*, 8, 15–18.
- Zwaan, R. A., & Brown, C. M. (1996). The influence of language proficiency and comprehension skill on situation model construction. *Discourse Processes*, 21, 289–327.
- 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.
- Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 386–397.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162–185.
- Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Constructing multidimensional situation models during reading. *Scientific Studies of Reading*, 2, 199–220.
- Zwaan, R. A., & Rapp, D. N. (2006). Discourse comprehension. In M. Traxler & M. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 725–749). San Diego, CA: Academic Press.