Monitoring Local Comprehension Monitoring in Sentence Reading

Christian Vorstius and Ralph Radach
University of Wuppertal

Michael B. Mayer and Christopher J. Lonigan
Florida State University

Abstract. Comprehension monitoring is considered a key issue in current debates on ways to improve children' reading comprehension. However, processes and mechanisms underlying this skill are currently not well understood. This article describes one of the first attempts to study comprehension monitoring using eye-tracking methodology. Students in fifth grade were asked to read sentences for comprehension while also checking whether the meaning of the sentence was generally correct or incorrect. Items required the processing of conjunctive relations between two clauses that were either causally consistent or inconsistent. In addition, the polarity of the relation was varied by replacing the conjunction "because" with "although, " creating an additional level of processing difficulty. Inconsistency played a minor role and was dominated by polarity effects that were also modulated by the correctness of the answer. The present task represents an effective tool to study local comprehension monitoring and highlights the importance of conjunctive relations for maintaining textual coherence during reading.

Reading comprehension can be defined as the active extraction and construction of meaning from all kinds of text (Snow, 2001), requiring the reader to fluently decode and then understand the material they are reading (Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007; Scarborough, 2001). In today's society, reading comprehension can certainly be seen as a core component of becoming (and being) a successful individual. However, numbers from the National Assessment of Education Progress indicate that a large percentage of students (>30% of fourth-graders and ~25% of eighth-graders) read below a basic level of competence (National Assessment of Education Progress, 2007). Despite the obvious importance of literacy and decades of work aimed at improving students' performance in the reading domain, relatively little is known about the core component processes

and cognitive mechanisms that underlie reading comprehension.

In the present study, we will focus on comprehension monitoring, an ability frequently mentioned in discussions about techniques to improve children's reading comprehension. The idea of comprehension monitoring is certainly not new (Dewey, 1910; Locke, 1975; Stauffer, 1969; Thorndike, 1917), and there have been numerous studies looking at its relation to reading performance. Typically, comprehension monitoring is studied by embedding erroneous information in the text materials and then assessing participants' ability to detect these deviations from normal text. The type of error introduced varies between studies and has included internal inconsistency (Berthiaume, Lorch, & Milich, 2009; Chan, Cole, & Barfett, 1987; Ehrlich, Remond, & Tardieu, 1999; Markman & Gorin, 1981; Markman, 1979; Oakhill, Cain, & Bryant, 2003; Oakhill & Cain, 2012), lexical inconsistency (Garner, 1981), incomplete information (Markman, 1977), violation of prior knowledge (Markman & Gorin, 1981), syntactical errors (Paris & Myers, 1981), propositional incohesiveness (Flavell, Speer, Green, & August, 1981), and structural incohesiveness (Harris, Kruithof, Terwot, & Visser, 1981). Some error types (e.g., nonsense words or violation of prior knowledge) are easier to identify than others, with internal consistency errors being the most difficult to detect (Baker, 1984; Garner, 1981; Markman & Gorin, 1981; Reis & Spekman, 1983) and being the best candidate to differentiate between readers with different abilities.

Almost all of these studies viewed comprehension monitoring as a reading strategy that is employed deliberately and with conscious reflection. Participants are usually informed that there are errors in the reading material and asked to indicate where these errors occur, e.g., by underlining the respective words or passages in the text. Alternatively, self-corrections, repetitions, and hesitations during reading out aloud are recorded and taken as indicators for monitoring efforts when reading aloud (see Oakhill, Hartt, & Samols, 2005, for an excellent example). It may be a consequence of this dominating

methodology that comprehension monitoring is generally seen as the process by which readers constantly and actively engage in the deliberate evaluation of how well they understand newly incoming information.[1]

Another difficulty affecting real progress in this area of research can be attributed to a certain degree of ambiguity and fuzziness in terminology. Afflerbach, Pearson, and Paris (2008) noted that the inconsistent use of concepts like "reading skill" and "reading strategy" in the pertinent literature confuses students, teachers, and of course researchers. This can potentially render interventions and instruction less effective. To reduce the terminological inconsistency, these authors suggested that the terms *skill* and *strategy* should be used to distinguish between automatic and deliberately controlled processes. Notably, with ongoing reading development, strategies are more and more automatized and eventually become skills. A proficient reader should be able to switch between automatic and deliberate processing when the situation—for example, a high text difficulty—calls for it. The adequate processing mode (and switching between modes) should therefore be a function of reader, text, task, and contextual variables.

Being able to make necessary switches between automatic and deliberate processing modes seems to require the reader to continuously monitor his/her comprehension. Therefore, the concept of "comprehension monitoring" appears to find its natural place within this theoretical framework. In our view, the research on comprehension monitoring mentioned earlier has provided many useful insights, but it is also limited in regard to at least the following two aspects: First, explicitly asking participants to search for deviant information induces a kind of control behavior that developing readers may not always recruit in their normal classroom reading. Indeed, findings from early research suggest that poor readers' lack of comprehension monitoring may be a performance deficit rather than a skill or strategy deficit (Baker, 1984; Flavell, 1970; Chan & Cole, 1986; Torgesen, 1977). Second, making the existence of deviant in-

formation explicit may obscure a very important research question: Does comprehension monitoring necessarily have to be deliberate and reflexive? When, as argued earlier, reading strategies have a tendency to become more and more automatic, why should this not also be true here? This would suggest that in the course of reading development, comprehension monitoring becomes less of a "strategy" and more of a "skill. " In other words, the process of continuous checking of text understanding may become so automatic that its deliberate quality disappears and "monitoring" becomes simply part and parcel of skilled comprehension.

We believe that further research on these questions would profit from shifting the methodological focus from relying on results of linguistic processing (such as error scores) to fine-grained analyses of information processing during reading. To make a further step in this direction, the present study used eye-tracking methodology to examine the time course of processing causal inconsistencies during sentence reading. Advances in eye movement monitoring technology over the last 30 years combined with a substantially increased theoretical understanding of eye movement control during reading (see Radach & Kennedy, 2004, 2012; Rayner, 2009; Rayner & Kliegl, 2012, for recent introductions and reviews) have laid the foundation for routine use of this methodology in research on comprehension. Eye movement patterns and temporal parameters can be taken as online indicators of ongoing cognitive processes, providing a tool for looking at reading comprehension as it is occurring (Radach, Schmitten, Glover, & Huestegge, 2009; Rayner, Chace, Slattery, & Ashby, 2006). Mapping out the basic processes that are at work can help draw a clearer picture of how comprehension is monitored. Ultimately, this may enable the development of targeted interventions to increase monitoring ability and/or use in readers who show inferior comprehension monitoring.

So far, only one study has specifically looked at eye movements during comprehension monitoring in children. Van der Schoot, Reijntjes, and van Lieshout (2011) used internal inconsistencies to study comprehension monitoring in a story-reading task. On top of the consistency manipulation, the distance between the original information and the consistent/inconsistent target sentence was varied to be either short (local inconsistency) or long (global inconsistency). Their results showed that overall, both good and poor readers spent more time on the target sentence in the inconsistent condition. However, only for good readers was the effect present in both "local" and "global" conditions, whereas poor readers did not show an inconsistency effect in the "global" condition. Van der Schoot et al. (2011) interpreted their findings in terms of poor comprehenders' failure to construct an elaborate situational model. Although this approach can be very informative, it relies on a very high level of comprehension with substantial demands on memory and reasoning. In our view, this type of task would be at the very end of a spectrum of complexity in the continuous checking of comprehension. The purpose of the present work was therefore to make a first step into examining the dynamics of comprehension monitoring within a single sentence (see also the distinction between word level, sentence level, and intrasentence inconsistencies in Oakhill et al., 2005).

Our approach has some similarities to recent work on the processing of plausibility and anomaly during sentence reading (Rayner, Warren, Juhasz, & Liversedge, 2004), where target words appear atypical within the constraints of a specific sematic relation. A major difference is that in research on plausibility, participants are asked to read as they usually do so that the processing of any inconsistency or implausibility occurs as part of normal reading routines. In contrast, research on comprehension monitoring typically involves some level of information signaling the existence of inconsistent information so that the reading process is modified to search for these inconsistencies and to accommodate the resulting difficulties in comprehension.

The present study was designed as part of a series of eye movement experiments examining different aspects of reading comprehension in fifth-grade students. This age group

Table 1
Examples of Sentence Materials Used in All Four Conditions

	Region 1	Region 2	Region 3
Consistent positive	Daniel was *shivering* Erica *blushed*	because because	he was cold. she was nervous.
Inconsistent positive	Daniel was *shivering* Erica *blushed*	because because	he was hot. she was confident.
Consistent negative	Daniel was *shivering* Erica *blushed*	although although	he was hot. she was confident.
Inconsistent negative	Daniel was *shivering* Erica *blushed*	although although	he was cold. she was nervous.

Note. **Bold italics** highlight the embedded target word in Region 1. Note that spacing by regions and highlighting of targets are shown for demonstration purposes only.

appeared to be an appropriate starting point for a number of reasons, including the ability to follow relatively advanced task instructions and the fact that an appropriate set of skills and strategies related to comprehension monitoring can be expected to be developed by this grade level. For this experiment, we focused on the processing of conjunctive relations between clauses within a single sentence. This choice was motivated by the fact that the more general class of coherence relations represents a cornerstone of comprehension, serving to organize text content by connecting substantive ideas and guiding their interpretation within text (Graesser, McNamara, & Louwerse, 2003). Conjunctive relations often link adjacent clauses, with causal relations being one of the strongest ways to connect descriptions of subsequent events. Our materials comprised sentences like "Erica blushed because she was nervous/confident" (see Table 1), where a critical word in the final clause creates a consistent versus inconsistent relation. This type of construction has the advantage that by changing the conjunction word from *because* to *although*, the polarity of the relationship can be changed as in the example "Erica blushed although she was nervous/confident." This creates an additional level of processing difficulty, effectively transforming the causal rela-

tion in to a much more complex adversative relation (Halliday & Hasan, 1976).

In positive relations, the situation presented in the first clause is continued in the conjoined situation (e.g., "Jenny was glad because she passed her class."). In negative relations, on the other hand, the expected relation is discontinued (e.g., "Jenny was glad although she failed her class."). The underlying assumption (e.g., people are glad if they pass a class) is therefore confirmed in positive polarity sentences and negated in negative polarity sentences. Adding this polarity manipulation should substantially increase the difficulty of detecting inconsistencies and may therefore allow for differentiating between different levels of processing.

Our working hypothesis was that although detection of consistencies may often be accomplished within automated routines of comprehension, the added layer of complexity in adversative relations may require a more deliberate mode of cognitive control that would come close to the standard definition of comprehension monitoring discussed earlier. More specifically, we expected to find substantial increases in viewing times in the sentence final region, when readers need to determine whether the conjunctive relation is consistent or not. It was also expected that this

decision would sometimes require a reinspection of the sentence initial region. Considering the polarity manipulation, an increase of initial viewing time was expected, along with a tendency to reread the sentences' initial region because of the reverse allocation of processing resources in the adversative relation. These hypotheses characterize optimal performance when the task was solved successfully (i.e., inconsistencies were detected when present); the pattern should be attenuated for trials in which inconsistencies were not detected.

Method

Participants

Seventy-six 5th-grade students from three schools in the Bay County School District in North Florida and the Florida State University School in Leon County, Florida, were tested in this experiment as part of a larger study within the Reading for Understanding project with the Florida Center for Reading Research. Participants were drawn randomly from a large sample of students recruited at schools in North Florida, representing the racial and socioeconomic mix of this region. Written consent to participate in this work was obtained from a parent or legal guardian. In addition, verbal assent from participating students was confirmed prior to beginning the experimental session. All participants were native speakers of English and had normal or corrected to normal visual acuity as determined by a brief vision screening using an OPTEC 5000 vision tester. Average age at time of participation was 11.3 years. Information about the purpose and content of the experiment was provided to all parents or legal guardians.

Materials

Participants were asked to read a total of 28 single-line sentences consisting of 4 practice trials and 24 experimental trials. Sentences were constructed so that they contained a causal relationship. Specifically, the first clause of each sentence (referred to as Region 1) was a simple statement, followed by a con-

junction word (Region 2) and a final clause (Region 3) that created either consistent or inconsistent sentences (see Table 1). The consistency manipulation was introduced by changing one word in Region 3 in half of the trials so that it was compatible or incompatible with the meaning of Region 1 (e.g., "The vase shattered because it was made of glass." vs. "The vase shattered because it was made of plastic."). The verb in Region 1 was specified as a target word for supplementary analyses, as the verb meaning together with the meaning of the sentences' final clause established the critical semantic relation (see Table 1).

The conjunction word determined the polarity of a given sentence and was either *because* for positive polarity or *although* for negative polarity. Half of the trials contained a positive conjunction; the other half contained a negative conjunction. These manipulations amounted to a total of four conditions (Consistent-Positive, Consistent-Negative, Inconsistent-Positive, and Inconsistent-Negative). Four counterbalanced lists of 28 passages with 6 experimental items (and 1 practice item) per condition in each list were created.

Apparatus

Sentences were displayed one-by-one in black on a light gray background using a 21-inch flat-panel monitor with a display resolution of 1024×768 pixels, running at a 120 Hz refresh rate. Texts were presented in 15 pt Courier New font and the viewing distance between each reader's eyes and the monitor was set to 68 cm. At this distance, each letter subtended 0.33° of visual angle laterally. Eye movements were recorded using the Eye-Link1000, a state-of-the-art video-based pupil and corneal reflection tracking system (SR Research Ltd.), sampling at 500 Hz. This recording system has a relative spatial resolution in the order of a few minutes of arc and its absolute accuracy is better than 1/3°, depending on calibration. A 3-point calibration was performed every 4 trials. Mean average position error in an accuracy validation routine was not to exceed 0.33° of visual angle. A drift-check before every trial ensured accuracy

of the measurement between calibrations. If the drift check showed a deviation from more than 0.33° of visual angle, an additional calibration was performed. The online saccade parser of the eye tracking system was set to detect saccades with an amplitude of 0.15° or greater, using an acceleration threshold of $8,000°/s^2$ and a velocity threshold of 30°/s. These settings have proven to produce highly accurate and reliable data in multiple reading studies across different laboratories.

Procedure

The present experiment was the third in a session of different eye tracking tasks so that participants were already familiar with the set up and calibration routines at the start of the experiment. Participants were seated in a height-adjustable chair at a desk in front of the stimulus display monitor. At the beginning of the experiment, participants were presented with a set of directions on the screen. Participants were instructed to read each of the sequentially presented sentences at their normal reading speed so that they understood its meaning. In an effort to avoid inducing an error-searching reading mode, we decided to not use terms like *correct* or *consistent* in the instructions for our fifth-grade sample. Instead, students were asked to indicate via button press whether a sentence seemed generally fine or rather weird. Although this instruction does not completely eliminate the chance of inducing a search-biased reading mode, a pilot study with small number of students suggested that this instruction was less likely to do so than a standard inconsistency detection instruction. After the directions were read, the experimenter asked if there were any questions and answered them. The experiment then started with a 3-point calibration and participants were guided through four practice sentences to guarantee complete understanding of procedure and instructions. The 24 experimental trials were presented, interspersed with calibration on every 4th trial to ensure data accuracy. After the experiment, participants were thanked for their participation and escorted back to their classroom.

Dependent Variables

In the following, brief definitions of the eye movement parameters that were used as dependent variables are given. Saccades are rapid eye movements that serve to bring the retinal area of highest visual acuity in alignment with a region of interest, whereas fixations represent phases of relatively stable visual axes. Reading-relevant visual information is only acquired during fixations, which take approximately between 100 and 500 ms, with mean durations in the order of 280–300 ms for children. Notably, the duration and number of fixations reflect the mental workload associated with the processing of linguistic information (see Inhoff & Radach, 1998; Rayner, 1998, for discussions of oculomotor measures).

More specifically, *first pass reading time* is the summed duration of all fixations made within a region (or on a target word) before the first saccade left that region. *Rereading Time* refers to additional time spent within a region (or on a target word) after the first pass has ended, irrespective of whether the first incoming saccade in later passes is progressive (from left to right) or regressive (from right to left). *Total number of passes* is the number of times a region or word was looked at with one or multiple fixations (in analyses on the word level the term *gaze* is often used instead of *pass*). *Total number of region crosses* is the number of times that a boundary between regions was crossed between successive fixations. Unless otherwise specified, all variables were calculated for the three sentence regions (see Tables 2 and 3). Additional analyses were carried out to test further hypotheses.

Data Analysis

Saccades and fixations were classified online using EyeLink software. In further steps of data analysis, raw eye movement data were converted to pairs of fixations and incoming saccades and then aggregated to a target word based matrix using the custom built software suite EyeMap (Tang, Reilly, & Vorstius, 2011) as well as SPSS. During this

Table 2
Means and *SE* for Sentence Reading Times as a Function of Polarity, Consistency, and Correctness

	Correct Response				Incorrect Response			
	Consistent		Inconsistent		Consistent		Inconsistent	
	Positive Polarity	Negative Polarity	Positive Polarity	Negative Polarity	Positive Polarity	Negative Polarity	Positive Polarity	Negative Polarity
Sentence reading time	2689 (83)	3449 (169)	3101 (103)	3369 (156)	3051 (218)	3357 (116)	3318 (206)	3535 (114)

Note. Values are given in milliseconds.

Table 3
Means and *SE* for First Pass Reading Time by Sentence Region as a Function of Polarity, Consistency, and Correctness

	Correct Response				Incorrect Response			
	Consistent		Inconsistent		Consistent		Inconsistent	
	Positive Polarity	Negative Polarity	Positive Polarity	Negative Polarity	Positive Polarity	Negative Polarity	Positive Polarity	Negative Polarity
Region 1	979 (33)	1010 (49)	987 (29)	934 (44)	1062 (73)	1040 (38)	1163 (65)	1072 (38)
Region 2	278 (6)	301 (14)	286 (7)	304 (15)	327 (21)	314 (9)	308 (18)	333 (15)
Region 3	906 (30)	946 (52)	991 (35)	837 (42)	893 (75)	970 (43)	1213 (91)	960 (35)

Note. Values are given in milliseconds.

procedure, all data were also inspected visually to detect any possible problems. Ten data sets had to be excluded, as participants did not follow instructions and/or always pushed the same response button for all trials. Fixations with durations shorter than 70 ms or longer than two standard deviations of the participant's mean were excluded from analysis. This resulted in the exclusion of less than 4% of all remaining cases. Temporal variables were log transformed for inferential statistics to better fit a normal distribution. Figures and tables, however, are presented in nontransformed values.

Inferential statistics for continuous variables are based on linear mixed models specifying subjects and sentences as crossed random effects (Baayen, Davidson, & Bates, 2008; Kliegl, Masson, & Richter, 2010). Binary variables were analyzed with generalized linear mixed models using the binomial distribution with a logit link function. Rather than dealing with $F1$ and $F2$ analyses of variance, differences between participants and sentences (items) are accounted for in a single analysis in linear mixed models. In addition, linear mixed models lose less statistical power, especially with unbalanced designs that are typical for eye movement experiments (Baayen, 2008; Quené & van den Bergh, 2008). If not stated otherwise, all analyses included Consistency (consistent vs. inconsistent), Polarity

(positive vs. negative), Correctness (correct vs. incorrect), and their interactions as fixed effects.

All effects were estimated using the lmer program from the lme4 package (Bates & Mächler, 2009) in the R environment for statistical computing (Version 2.15.3; R Core Team, 2012). Regression coefficients, standard errors (*SE*), and *t* values are reported. For all tests, we used the two-tailed criterion ($t > = 1.96$ *SE* or $z > = 1.96$ *SE*), corresponding to a 5% error criterion for significance.

Results

Before reporting results by dependent variables for each region, it should be noted that the proportion of correct responses varied greatly between conditions. The means amounted to 83% correct for the consistent positive condition, 80% for the inconsistent positive condition, 35% for the consistent negative condition, and 41% for the inconsistent negative condition. Statistically, this was expressed in a main effect for polarity ($b = 2.15$, $SE = 0.13$, $z = 16.82$) and a marginally significant interaction between consistency and polarity ($b = -0.48$, $SE = 0.25$, $z = -1.89$), indicating that negative polarity led to fewer correct answers and this effect was slightly more pronounced for consistent items. Because of the overall low rate of correct responses in the negative polarity conditions, we did not exclude data sets with incorrect answers, but included Correctness (of the answer) as a fixed effect in our models. The following report of our results will start with an analysis of sentence reading time (similar to a response time analysis) to examine the overall affect of the three factors Consistency, Polarity and Correctness. In a second step, we will then divide reading time per region (sentence initial, conjunction word, and sentence final) between first pass and later passes to pursue more specific hypotheses.

Sentence Reading Time

Overall sentence reading time with correct responses varied between 2,689 ms for

consistent trials with positive polarity and 3,369 ms for inconsistent trials with negative polarity. Mean sentence reading times for incorrect responses were slightly longer (165 ms) than those for correct answers, but this numerical difference did not materialize as a main effect of correctness. Instead, the negative polarity manipulation led to an inflation of sentence reading times as reflected in a strong main effect of Polarity ($b = -185.78$, $SE = 102.40$, $t = -3.74$). In addition, we found an interesting interaction between Polarity and Correctness ($b = -525.64$, $SE = 195.00$, $t = -2.69$), indicating that the polarity effect was even more pronounced for correct answers, an effect driven mainly by shorter sentence reading times in the positive polarity condition (Table 2).

First Pass Reading Time per Region

First pass reading time for Region 1 describes the time used for the initial reading of the first clause of the sentence. There was a main effect for Correctness ($b = 0.05$, $SE = 0.03$, $t = 2.08$), based on the fact that readers were 95 ms faster when an accurate response was given in a trial. This result was repeated for Region 2 ($b = 0.05$, $SE = 0.02$, $t = 2.09$), corresponding to the conjunction word *because* for positive and *although* for negative polarity. Here, the difference amounted to 34 ms, which is approximately the same proportion of first pass reading time as in Region 1. Surprisingly, in Region 3, corresponding to the sentences' final clause, there were no significant main effects or interactions for first pass reading time (Table 3).

Rereading Time per Region

Rereading time includes all fixations made during returns when another region was previously fixated. Focusing on Region 1, there was a significant main effect for Polarity ($b = -0.80$, $SE = 0.25$, $t = -3.23$), indicating that negative polarity caused readers to spend more rereading time in the sentences' initial region. However, when the analysis was restricted to only include incoming saccades from Region 3, this effect disappeared.

Table 4
Means and *SE* for Rereading Time by Sentence Region as a Function of Polarity, Consistency, and Correctness

| | Correct Response | | | | Incorrect Response | | | |
| | Consistent | | Inconsistent | | Consistent | | Inconsistent | |
	Positive Polarity	Negative Polarity	Positive Polarity	Negative Polarity	Positive Polarity	Negative Polarity	Positive Polarity	Negative Polarity
Region 1	458 (37)	614 (75)	660 (50)	712 (73)	523 (97)	628 (46)	502 (93)	684 (54)
Region 2	229 (21)	550 (66)	256 (23)	608 (51)	399 (82)	444 (41)	320 (55)	491 (38)
Region 3	312 (38)	633 (89)	439 (49)	691 (81)	419 (108)	554 (61)	437 (108)	532 (57)

Note. Values are given in milliseconds.

Considering rereading time on Region 2 (conjunction word), there was again a significant main effect of Polarity ($b = -1.18$, $SE = 0.21$, $t = -5.60$), with longer rereading times for negative polarity. Notably, there was also a significant interaction between Polarity and Correctness ($b = 0.82$, $SE = 0.37$, $t = 2.24$), indicating that in the case of correct responses substantially more time was spent viewing the conjunction word *although* in the negative polarity condition compared to *because* in the positive polarity condition. Given the relatively small extent of Region 2, it is not only remarkable how much time was spent rereading the conjunction word but also that this time nearly doubled in the negative polarity cases (see Table 4).

In contrast to Region 2, the time spent rereading Region 3 (final clause), was much shorter compared to its initial reading as expressed in first pass reading time. Still, the pattern of effects was similar to the one just described for Region 2, even though the numerical increase between positive and negative polarity conditions was not as dramatic. Again, we found a main effect of Polarity ($b = -0.92$, $SE = 0.24$, $t = -3.78$), with longer rereading times for negative polarity. This was combined with an interaction between Polarity and Correctness ($b = 1.26$, $SE = 0.44$, $t = 2.85$). The polarity effect turned out to be more pronounced for correct answers. Figure 1 highlights the effects of Polarity and Correctness combined across Region 2 and Region 3.

Supplementary Analyses

A number of supplementary analyses were performed to examine the dynamics of movements between regions. First, we computed the total number of crosses between all regions. We found a significant main effect of Polarity ($b = -0.98$, $SE = 0.24$, $t = -4.12$), showing that there were substantially more crosses from one into another region for negative polarity items. This effect was qualified by a Polarity by Correctness interaction ($b = 0.89$, $SE = 0.37$, $t = 2.40$), indicating that success in solving the task was associated with a larger frequency of crossing from one region into another.

This analysis was repeated for each of the three regions based on the total number of passes. This more fine-grained level of analysis revealed main effects of Polarity across all regions (Region 1: $b = -0.26$, $SE = 0.07$, $t = -3.53$; Region 2: $b = -0.52$, $SE = 0.10$, $t = -4.98$; Region 3: $b = -0.29$, $SE = 0.07$, $t = -3.90$). In addition, there were interaction effects identified between Polarity and Correctness in Region 2 ($b = 0.58$, $SE = 0.17$, $t = 3.42$) and Region 3
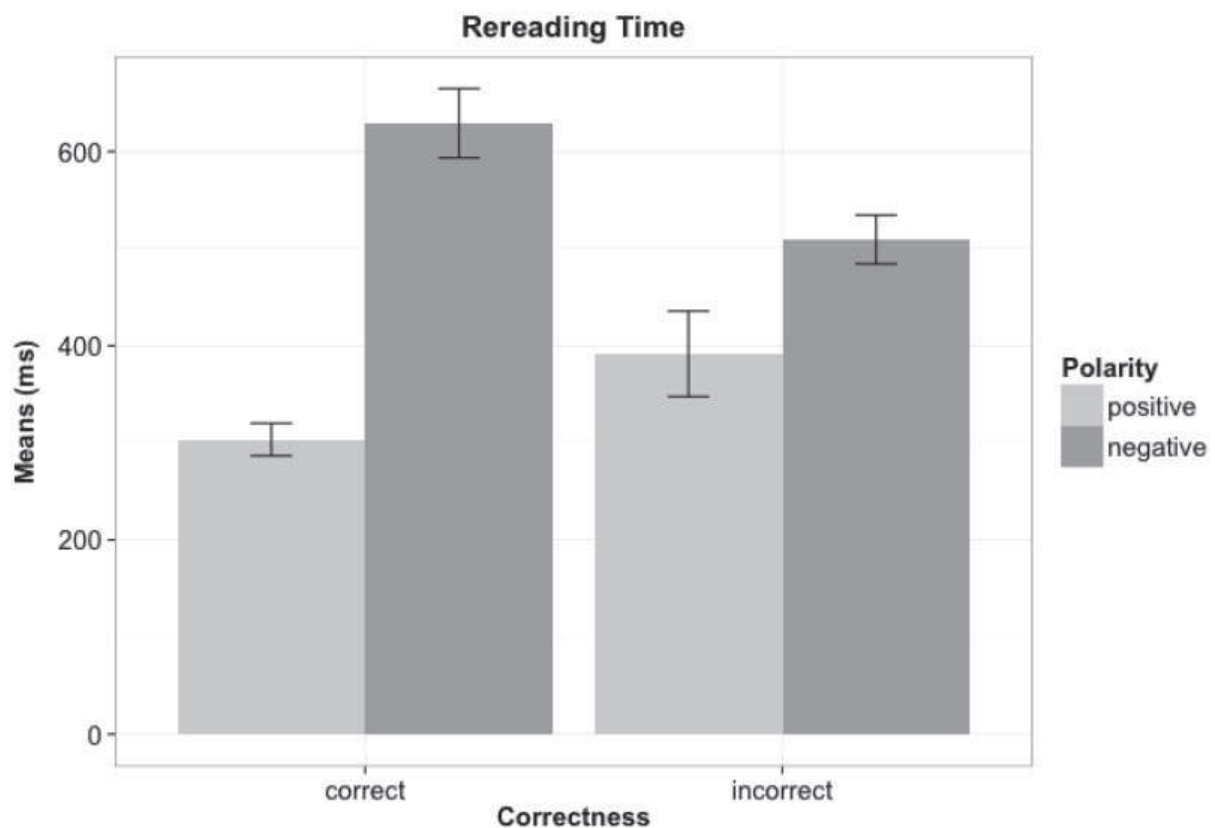
Rereading Time

Figure 1. Mean rereading times combined for Region 2 and Region 3, as a function of polarity correctness. Error bars indicate *SE*.

($b = 0.49$, $SE = 0.13$, $t = 3.75$), showing that returning more often to reread these regions was related to correct answers.

Target Word Viewing Durations

The sentences' initial clauses in all our items included verbs that strongly determined the correctness of given causal or adversative relations (e.g., "The vase *shattered* because it was made of glass."). It is reasonable to assume that a successful check of comprehension would often involve a regression back to the verb to reexamine its semantic properties in relation to the meaning of the sentences' final clause. To test this hypothesis, gaze durations and rereading times were computed for these target words. Looking at gaze durations, a main effect was found ($b = 0.06$, $SE = 0.05$, $t = 2.51$), replicating the more general finding of shorter first pass reading times for correct answers. Looking at rereading time, there was a main effect of polarity ($b = -0.55$, $SE = 0.28$, $t = -1.98$), indicating that more

time was spent rereading the critical word in items with negative polarity. However, this difference was not related to success in solving the task (see Figure 2).

Discussion

The present work constituted an initial step towards using eye tracking methodology to study comprehension monitoring at the single sentence level. Similar to prior work we manipulated the consistency of a sentence (e.g., Oakhill et al., 2005), in our case by using a causal conjunction relation. As a novel manipulation in experimental research, we included a change in polarity of the conjunction, effectively turning the causal into an adversative relation. The focus of data analysis was on viewing time measures for three regions: (1) the sentences' initial clause, (2) the conjunction word, and (3) the sentences' final clause.

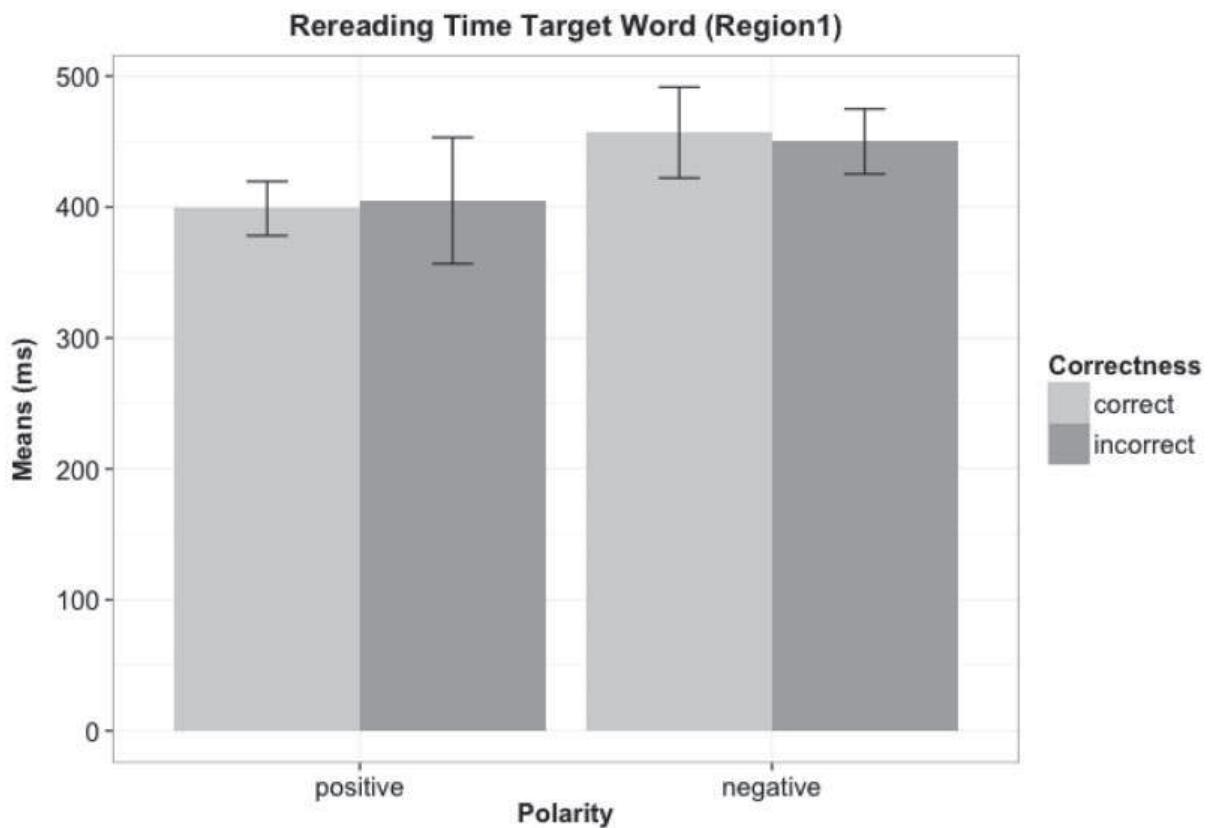The first notable result of this study is that the difficulty of the negative polarity

Figure 2. Rereading time on the embedded target word as a function of Polarity and Correctness. Error bars indicate *SE*.

items was quite high for our fifth-grade sample, leading to a large number of unsolved or misunderstood items. As a consequence, we included correctness of response as an additional factor in our analyses. Looking at overall sentence reading time, it is interesting to note that there was no clear overall effect of correctness. One possible outcome would have been that more successful readers were also faster in solving the task, or alternatively, it could have been the case that students who invested more time attained better performance.

A clear difference between trials with correct versus incorrect responses emerged in the analyses of first pass reading times in Region 1 and Region 2, before the nature of the critical relation became apparent. Here, faster reading predicts better performance; although, interestingly, the lack of a corresponding main effect in overall sentence reading time indicates that success in comprehension is associated with more time spent with careful

rereading of useful information (see later). The lack of any factorial effects on the initial reading of Region 3 may appear somewhat counterintuitive. However, this result may indicate that readers, aware of the potential difficulty of items, routinely preferred additional information acquisition (in Regions 1 and 2) over an early termination of trials. This interpretation is in line with the observation that only 24.4% of all trials ended after just one sequence of first pass reading over all three regions (without performance difference for these trials).

This leads to the most informative set of results, dealing with the return to previously read regions. Interestingly, the evidence for any effects on the sentences' initial region was not strong, with a significant polarity effect disappearing when only saccades from Region 3 were considered. In contrast, for the sentences' final region, a substantial amount of time was spent with additional reading, and this amount of rereading was associated with

accurate responses in the task. The same pattern also emerged for the conjunction word, but with much larger numerical differences, leading to the conclusion that focusing on this critical element is key to high performance. These results were confirmed and extended in supplementary analyses focusing on the frequency of movement between regions, again showing massive effects of polarity combined with better performance when Region 2 and/or Region 3 were revisited more often. Finally, it was shown that readers tended to find and reread the critical verb in the sentences' initial region, especially with negative polarity items. However, this did not necessarily lead to correct answers.

Although somewhat different from most prior work in this area, we see our study as an example of research on "comprehension monitoring." Successful comprehension monitoring requires the ability to notice instances in which coherence in a given text cannot be established and thus needs appropriate actions to resolve the problem. This might not have been the case with the causal coherence/incoherence manipulation, which many participants may have mastered on the basis of more or less automatic processing routines. Consequently, this led to the observed lack of a significant incoherence effect. However, the resolution of adversative incoherence required additional operations (apparent in crossing between regions) of information acquisition, transformation, and verification. We see no reason to require any level of deliberateness or conscious reflection as a precondition of success in this chain of operations. Our argument here has similarities with the one made by Radach, Huestegge, and Reilly (2008). These authors manipulated "reading intention" (or level of processing) by simply varying the type of question participants were asked after reading sentences, using a simple word verification task versus complex comprehension questions. This manipulation led to substantial changes in observed reading behavior, even modulating the microlevel of word processing without even requiring a change in instruction. In other words, if reading intention or strategy can change without con-

scious reflection, why should the monitoring of comprehension not potentially become an "under the hood" reading skill as development progresses?

Certainly, "reading intention" can also be affected by the types of instructions given to participants. It is known that the instructions provided to readers can strongly influence reading outcomes (O'Shea, Sindelar, & O'Shea, 1985, 1987). As described in the Introduction, research on comprehension monitoring usually uses an error search instruction. In the current experiment we used a more conservative monitoring instruction in an effort to minimize specific instruction effects. Moreover, the condensed format of our task with manipulations at the sentence (and not passage) level allowed us to collect an adequate amount of trials per participant and condition within the real life constraints of research in a school setting.

Despite these advantages, the instruction used in this study might have contributed to some of the effects found. Some sentences that are logically and grammatically correct might have seemed unusual to the children, and thus might have contributed to the high number of trials that were defined as incorrect. However, no items could be identified that were consistently misclassified across subjects. Taken together, these considerations lead to the conclusion that the role of instruction on reading behavior and particularly on the monitoring of comprehension should be the focus of further research.

In addition, future research should attempt to illuminate the effect of individual differences in general reading proficiency and component skills on comprehension monitoring, using the condensed tasks format introduced in this study. Although we did not detect any outliers in our sample in terms of reading speed, there was considerable variability between participants. After establishing the feasibility of the basic paradigm, the inclusion of individual difference measures is the logical next step on the way towards turning the examination of adversative causal relations into a diagnostic tool.

Consequences for Educational Practice

The present study has shown that although causal relations between clauses are moderately difficult to comprehend for fifth-grade readers, adversative relations cause major problems for many students. In addition to practicing conjunctive relations in general, some specific suggestions can be derived from our data: when dealing with difficult conjunctive relations, the focus of instruction should be on identifying the critical elements, in our case the sematic properties of the critical verb and the conjunction word spanning the relation. If these determining elements are clear, the rest of the sentence will most likely fall into place. Notably, when comprehension checking operations are necessary, these should focus on the conjunction word itself (especially when it requires an adversative transformation such as *but* or *although*) and on the sentences' final region to verify the statement's correctness.

Looking at the problem of textual coherence more generally, it would be very interesting to extend the list of relations included in this type of research to also include other types of complex conjunctive relations within sentences. This would gradually lead to a data-based classification of relations that deserve special attention in the classroom. Candidates for this expansion of applied research would certainly include temporal (*and then, then, when, before, after, during, while*), intentional (*in order to, by means of),* and logical (*therefore, so*) relations (see Halliday & Hasan, 1976, for a detailed overview). It appears possible that adversative relations turn out to be the most difficult on the list. It would be interesting to know the rank order of difficulty of the remaining relations, also in comparison with other mechanisms of textual coherence. This developing body of evidence could be used to inform instruction and teacher training. We completely agree with Graesser et al. (2002), who emphasized the following: "Just as there are reading programs that promote phonemic awareness, there should be those that emphasize coherence awareness. There should be a cottage industry of workbooks, computer software, and teacher training workshops that identify the different types of coherence relation" (p. 95).

One advantage of our sentence-level methodology is that, in comparison to the complex story reading task used in the interesting study of van der Schoot et al. (2011), information on performance and details of information processing can be collected in a much more economical way. Whether this difference can be turned into an advantage for diagnostic application remains a question for further research. This future research will have to show whether the detection of global, story-level inconsistences and the resolution of local coherence relation share enough functional similarity to comfortably fit under the umbrella term of *comprehension monitoring.*

Returning to the discussion of skill versus strategy in the introduction section (e.g., Afflerbach et al., 2008), we would like to emphasize that results from the present study suggest that there are aspects of comprehension monitoring that appear to fall under the category of skill and are discernible from deliberate strategies. Further studies including individual difference measures and developmental data may help to draw a clearer picture on this transition from strategy to skill.

An interesting aspect of this dimension in reading development is related to reading fluency. Theories in this domain usually build on the automaticity view of reading (LaBerge & Samuels, 1974), basically assuming that higher degrees of automatization of underlying processes free resources for meaning construction. Although there is considerable evidence that oral reading fluency is a good predictor for reading comprehension (see Kim, Petscher, Schatschneider, & Foorman, 2010, for a review), some questions arise with respect to comprehension monitoring. Our data suggest that rereading of critical parts of the text can be beneficial for comprehension, although it might decrease fluency. Therefore, once basic operations (e.g., word decoding) are mastered, fluency could be affected to some extent by comprehension monitoring capabilities, with more fluent readers focusing on the important text regions,

whereas less fluent readers engage in complete rereadings or aimless scanning of text parts. For the classroom, it is therefore important to determine the origin of a student's lack of fluency in addition to training children in identifying critical text regions. In the future, easy to handle and affordable eye tracking systems might become a useful tool in achieving these aims.

Although our findings represent only a first step towards a better understanding of "under the hood" processes of comprehension monitoring, we think that this is a promising approach. In addition to enabling us to better define comprehension monitoring, it will also help in the development of appropriate intervention tools for the classroom.

Footnotes

[1]The situation is made more complex by the fact that in educational practice there is one specific use of the term *comprehension monitoring* as "involving a process of readers self-listening (monitoring) or listening to others reading aloud" (Elliott-Faust & Pressley, 1986), so that awareness of difficulties in understanding is taught via "thinking aloud" during reading aloud (Trabasso & Bouchard, 2002, p. 178).

References

Afflerbach, P., Pearson, P. D., & Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher, 61*(5), 364–373.

Baayen, R. H. (2008). *Practical data analysis for the language sciences with R.* Cambridge, England: Cambridge University Press.

Baayen, R., Davidson, D., & Bates, M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412.

Baker, L. (1984). Children's effective use of multiple standards for evaluating their comprehension. *Journal of Educational Psychology, 76,* 588–597.

Bates, D., & Mächler, M. (2009). *lme 4, linear mixed-effects models using S4 classes* (R package version 0.999375–32) [Computer software]. Available at http://CRAN.R-project.org/package_lme4

Berthiaume, K. S., Lorch, E. P., & Milich, R. (2009). Getting clued in: Inferential processing and comprehension monitoring in boys with ADHD. *Journal of Attention Disorders*, published online October 8, 2009. doi:10.1177/1087054709347197

Chan, L. K. S., & Cole, P. G. (1986). The effects of comprehension monitoring training on the reading competence of learning disabled and normal students. *Remedial and Special Education, 7*(4), 33–40.

Chan, L. K. S., Cole, P. G., & Barfett, S. (1987). Comprehension monitoring: Detection and identification of text inconsistencies by LD and normal students. *Learning Disability Quarterly, 10*(2), 114–124.

Dewey, J. (1910). *How we think.* Boston, MA: Heath.

Ehrlich, M. F., Remond, M., & Tradieu, H. (1999). Processing of anaphoric devices in young skilled and less skilled comprehenders: Differences in metacognitive monitoring. *Reading and Writing, 11,* 29–63.

Elliott-Faust, D. J., & Pressley, M. (1986). How to teach comparison processing to increase children's short- and long-term listening comprehension monitoring. *Journal of Educational Psychology, 78,* 27–33.

Flavell, J. H. (1970). Developmental studies of mediated memory. In H.W. Reese & L. P. Lippitt (Eds.), *Advances in child development and behaviour* (Vol. 5, pp. 181–221). New York, NY: Academic Press.

Flavell, J., Speer, J. R., Green, F. L., & August, D. (1981). The development of comprehension monitoring and knowledge about communication. *Monograph of the Society for Research in Child Development, 46*(2, Serial No. 192).

Garner, R. (1981). Monitoring of passage inconsistency among poor comprehenders: A preliminary test of the "piecemeal processing" explanation. *Journal of Educational Research, 74,* 159–162.

Graesser, A. C., McNamara, D. S., & Louwerse, M. M. (2003). What do readers need to learn in order to process coherence relations in narrative and expository text? In P. Sweet & C. E. Snow (Eds.), *Rethinking reading comprehension.* New York, NY: Guilford Press.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English.* London, UK: Longmans.

Harris, P. L., Kruithof, A., Terwot, M. M., & Visser, T. (1981). Children's detection and awareness of textual anomaly. *Journal of Experimental Child Psychology, 31,* 212–230.

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29–54). Oxford, UK: Elsevier.

Kim, Y.-S., Petscher, Y., Schatschneider, C., & Foorman, B. (2010). Does growth rate in oral reading fluency matter for reading comprehension? *Journal of Educational Psychology, 102,* 652–667.

Kliegl, R., Masson, M. E. J., & Richter, E. M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition, 18,* 655–681.

LaBerge, D., & Samuels, J. (1974). Towards a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293–323.

Locke, E. Q. (1975). *A guide to effective study.* New York, NY: Springer.

Markman, E. M. (1977). Realizing you don't understand: A preliminary investigation. *Child Development, 48,* 986–992.

Markman, E. M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development, 50,* 643–655.

Markman, E. M., & Gorin, L. (1981). Children's ability to adjust their standards for evaluating comprehension. *Journal of Educational Psychology, 73,* 320–325.

National Assessment of Educational Progress. (2007). *The nation's report card.* NECS No. 2007496. Washington DC: National Center for Education Statistics.

Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study, *Scientific Studies of Reading, 16*(2), 91–121.

Oakhill, J., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing, 18,* 657–686.

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). Dissociation of single-word reading and text comprehension skills. *Language and Cognitive Processes, 18,* 443–468.

O'Shea, L. J., Sindelar, P. T., & O'Shea, D. J. (1985). The effects of repeated readings and attentional cues on reading fluency and comprehension. *Journal of Reading Behavior, 17*(2), 129–142.

O'Shea, L. J., Sindelar, P. T., & O'Shea, D. J. (1987). The effects of repeated readings and attentional cues on the reading fluency and comprehension of learning disabled readers. *Learning Disabilities Research, 2*(2), 103–109.

Paris, S. G., & Myers, M. (1981). Comprehension monitoring, memory, and study strategy of good and poor readers. *Journal of Literacy Research, 13*(1), 5–22.

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59,* 413–425.

R Core Team. (2012). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing (Web site: http://www.R-project.org/).

Radach, R., Huestegge, L., & Reilly, R. (2008). The role of top down factors in local eye movement control during reading. *Psychological Research, 72,* 675–688.

Radach, R., & Kennedy, A. (2004). Theoretical perspectives on eye movements in reading: past controversies, current issues and agenda for future research. *European Journal of Cognitive Psychology, 16,* 3–26.

Radach, R., & Kennedy, A. (2012). Eye movements in reading: Some theoretical context. *Quarterly Journal of Experimental Psychology.* Online publication November 16.

Radach, R., Schmitten, C., Glover, L., & Huestegge, L. (2009). How children read for comprehension: Eye movements in developing readers. In R. K. Wagner, C. Schatschneider, & C. Phythian-Sence (Eds.), *Beyond decoding: The biological and behavioral foundations of reading comprehension* (pp. 75–106). New York, NY: Guilford Press.

Rapp, D. N., van den Broek, P., McMaster, K., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading, 11*(4), 389–312.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422.

Rayner, K. (2009). Eye movements in reading: Models and data. *Journal of Eye Movement Research, 2*(5), 1–10.

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading, 10*(4), 241–255.

Rayner, K., & Kliegl, R. (2012). Eye movements and cognitive processes. In H. Cooper (Ed.), APA handbook of research methods in psychology, Vol 1: Foundations, planning measures, and psychometrics (pp. 413–427). Washington, DC: American Psychological Association. doi: 10.1037/13619-022

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1290–1301.

Reis, R., & Spekman, N. J. (1983). The detection of reader-based versus text-based inconsistencies and the effects of direct training of comprehension monitoring among upper-grade poor comprehenders. *Journal of Reading Behaviour, 15,* 49–60.

Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97–110). New York, NY: Guilford Press.

Snow, C. E. (2001). *Reading for understanding.* Santa Monica, CA: RAND Education and the Science and Technology Policy Institute.

Stauffer, R. G. (1969). *Directed reading maturity as a cognitive process.* New York, NY: Harper & Row.

Tang, S., Reilly, R., & Vorstius, C. (2011). EyeMap: a software system for visualizing and analyzing eye movement data in reading. *Behavioral Research Methods, 44*(2), 420–438.

Thorndike, E. L. (1917). Reading as reasoning: A study of mistakes in paragraph reading. *Journal of Educational Psychology, 8,* 323–332.

Torgesen, J. K. (1977). Memorization processes in reading-disabled children. *Journal of Educational Psychology, 69,* 571–578.

Trabasso, T., & Bouchard, E. (2002). Teaching readers how to comprehend text strategically. In C. C. Block & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 176–200). New York, NY: Guilford Press.

van der Schoot, M., Vasbinder, A. L., Horsley, T. M., & Van Lieshout, E. C. D. M. (2008). The role of two reading strategies in text comprehension: An eye fixation study in primary school children. *Journal of Research in Reading, 31*(2), 203–223.

van der Schoot, M., Reijntjes, A., & van Lieshout, E. C. D. M. (2011). How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor reading comprehenders. *Reading and Writing, 25,* 1665–1690.

Christian Vorstius, PhD, is an assistant professor in the School of Education and Social Sciences, Division of General and Biological Psychology, at the University of Wuppertal, Germany. His research interests include cognitive and visuomotor factors in reading development, inter- and intraindividual difference in skilled reading behavior, and relations between standardized assessments and eye movement measures. As a postdoctorate, he worked in several Institute of Education Sciences–funded projects at the Florida Center for Reading Research, managing programming, data collection, and analyses of eye tracking studies.

Ralph Radach, PhD, is a vice dean in the School of Education and Social Sciences and a professor of general and biological psychology at the University of Wuppertal, Germany. He has published widely on a range of topics in reading research, including attention and eye movement control, computational reading models, word processing in nonroman writing systems, acquired and developmental dyslexia, and most recently, reading for comprehension in children. His research is primarily funded by grants from the German Science Foundation (DFG), but he is also actively involved in projects funded by the Institute of Education Sciences and the U.S. National Institutes of Health.

Michael B. Mayer is a graduate student in the developmental psychology program at Florida State University and a fellow in the predoctoral interdisciplinary training program at the Florida Center for Reading Research. His interests center on word-, sentence-, and individual-level factors that influence language processing. More specifically, his research focuses on how these factors influence sentence processing across various syntactic structures, and the relation between syntactic parsing skills and comprehension ability.

Christopher J. Lonigan, PhD, is a distinguished research professor of psychology and an associate director of the Florida Center for Reading Research at Florida State University. His research interests include the development, assessment, and promotion of preschool early literacy skills and self-regulation. His current research projects, which are funded by the Institute of Education Sciences and the National Institute of Child Health and Human Development, include early identification of reading disability, effectiveness of response to instruction models in preschool, assessments for Spanish-speaking English learners, preschool curriculum evaluation, the development of reading comprehension, and the interaction between the development of self-regulation and academic skills.