How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor reading comprehenders

Menno van der Schoot · Albert Reijntjes ·
Ernest C. D. M. van Lieshout

Published online: 18 August 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract In two experiments, we investigated comprehension monitoring in 10–12 years old children differing in reading comprehension skill. The children's self-paced reading times (Experiment 1) and eye fixations and regressions (Experiment 2) were measured as they read narrative texts in which an action of the protagonist was consistent or inconsistent with a description of the protagonist's character given earlier. The character description and action were adjacent (local condition) or separated by a long filler paragraph (global condition). The self-paced reading data (Experiment 1), the initial reading and rereading data (Experiment 2), together with the comprehension question data (both experiments), are discussed within the situation model framework and suggest that poor comprehenders find difficulty in *constructing* a richly elaborated situation model. Poor comprehenders presumably fail to represent character information in the model as a consequence of which they are not able to detect inconsistencies in the global condition (in which the character information is lost from working memory). The patterns of results rule out an explanation in terms of impaired situation model *updating* ability.

Keywords Reading comprehension · Comprehension monitoring · Inconsistency detection · Situation models · Eye fixations · Children

M. van der Schoot (✉) · E. C. D. M. van Lieshout
Department of Educational Neuroscience, Vrije Universiteit, Van der Boechorststraat 1,
1081 BT Amsterdam, The Netherlands
e-mail: M.van.der.Schoot@psy.vu.nl

A. Reijntjes
Department of Psychosocial Development in Context, Utrecht University, Utrecht, The Netherlands

Introduction

An important component skill of reading comprehension in children is comprehension monitoring (e.g., Cain, Oakhill, & Bryant, 2004; Eme, Puustinen, & Coutelet, 2006; Rubman & Waters, 2000; Zinar, 2000). Comprehension monitoring refers to the processes by which readers evaluate their understanding of a text. Skilled readers who evaluate their comprehension constantly ask themselves if what they are reading makes sense. If it does not, they apply repair strategies to restore comprehension. In practice, comprehension monitoring often comes down to detecting and, if possible, resolving inconsistencies such as contradictory sentences, or statements conflicting with world knowledge. Measures of comprehension monitoring therefore usually involve the analysis of verbal and nonverbal responses immediately following a consistency violation (Skarakis-Doyle, 2002) or answers to questions following a text containing inconsistent or conflicting information (Zinar, 2000).

Although some studies have shown that children with comprehension difficulties have problems with dealing with internal inconsistencies in a text (e.g., Ehrlich, 1996; Ehrlich, Remond, & Tardieu, 1999; Vosniadou, Pearson, & Rogers, 1988; Yuill & Oakhill, 1991; Zabucky & Moore, 1989), there is no full agreement yet on the locus of this effect. One hypothesis is that poor comprehenders have difficulty in *detecting* the inconsistencies because they are impaired in the ability to construct a richly elaborated and coherent situation model from the text (e.g. Albrecht & O'Brien, 1993; Myers, O'Brien, Albrecht, & Mason, 1994; Rinck, Hahnel, & Becker, 2001; Rubman & Waters, 2000). Supposedly, this prevents them from (accurately) interpreting later text information in the light of earlier text information. Another hypothesis is that they are especially poor at *resolving* the inconsistencies. In this account, poor comprehenders are capable of detecting inconsistencies, but instead do not engage in, or lack knowledge of, appropriate comprehension-repair strategies (e.g. Gersten, Fuchs, Williams, & Baker, 2001; Hyönä, Lorch, & Rinck, 2003; Rinck, Gamez, Diaz, & de Vega, 2003; Wagoner, 1983). The main aim of this study was to replicate the finding of comprehension monitoring difficulties in children, and then to decide between the two above hypotheses. In addition, the role of verbal working memory capacity in comprehension monitoring was investigated. For these purposes, we adopted the situation model approach (Zwaan & Radvansky, 1998) to serve as our framework in which we set up the experiment and interpreted the data.

The situation model framework

It is generally assumed that reading comprehension involves the reader's ability to construct a mental representation of the situation described in a (narrative) text, rather than a representation of the text itself (e.g. Zwaan & Radvansky, 1998). Successful comprehenders build such a representation, called a situation model, by monitoring a number of key situational dimensions such as time, space, causation, and the protagonists' characteristics, goals and emotions (Zwaan, Langston, & Graesser, 1995a, Zwaan, Magliano, & Graesser, 1995b). By integrating information

from these dimensions, readers gradually update their representation and build a coherent and richly-connected situation model. Within the situation model framework, updating thus refers to the process of incorporating a new sentence or clause into the evolving model. For present purposes, it is important to recognize that inconsistencies interfere with the updating of the situation model as new incoming information is more difficult to integrate into the evolving model if it is inconsistent, or less consistent, with the information in the current state of the model.

The inconsistency detection task

To measure construction and updating of situation models, we used the inconsistency detection task described earlier by, among others, Albrecht and O'Brien (1993), Huitema, Dopkins, Klin and Myers (1993), Long and Chong (2001), O'Brien, Rizella, Albrecht and Halleran (1998) and Poynor and Morris (2003). In this task, participants typically read a text in which the action of a protagonist (e.g. eating a hamburger at McDonald's) is either consistent or inconsistent with the description of the protagonist's character (fast-food addict vs. vegetarian) or goals (*He wanted to have a quick bite before going home* vs. *He wanted to go wining and dining in a five-star restaurant*) presented earlier in the text. The character/goal description and the performed action can be adjacent (local condition) or can be separated by a substantial amount of intervening text (global condition). The purpose of the long filler paragraph in the global condition is to ensure that the character/goal description is eliminated from working memory by the time readers encounter the target action (Albrecht & O'Brien, 1993; Long & Chong, 2001). This means that if readers want to keep the character/goal information accessible for further processing they have to represent it in the situation model which is assumed to be developed and stored in long-term memory or long-term working memory (Ericsson & Kintsch, 1995). In the local condition, on the other hand, the character/goal elaboration is assumed to be still active in (short-term) working memory when participants read the target action.

Importantly, the inconsistency detection task can be used to explore the different types of information incorporated into situation models. The main argument behind this has been recently formulated by Hyönä, Lorch and Rinck (2003, p. 324): "If readers represent a certain type of information in the situation model, they should exhibit comprehension difficulties upon encountering a sentence that is inconsistent with regard to this information". Obviously, the reverse is true as well. That is, if readers do *not* build some piece of information into their situation model and a later sentence contradicts this information, the problem arises not so much in solving the inconsistency (updating the situation model and, if possible, restoring its coherence) as it does in detecting it (an inconsistency will simply go unnoticed if the information that is contradicted has been lost from memory). By applying these lines of argument, studies using (different versions of) the inconsistency detection task revealed not only the representation of character information (Albrecht & O'Brien, 1993; de Vega, Diaz, & Leon, 1997) and goal information (Huitema et al., 1993; Poynor & Morris, 2003) in situation models constructed from narrative texts

but also the representation of emotional information (Gernsbacher, Goldsmith, & Robertson, 1992), spatial information (de Vega, 1995; O'Brien & Albrecht, 1992) and temporal information (Rinck et al., 2001).

The present study

To our knowledge, most of the above-cited studies as well as other studies on this topic have been conducted with (young) adults. This prompted us to investigate the use of situation models in good and poor comprehenders who are between 10 and 12 years old, thereby extending a series of our previous studies on reading comprehension in children attending fifth and sixth grade at a regular primary school (van der Schoot, Vasbinder, Horsley, & van Lieshout, 2008, van der Schoot, Bakker Arkema, Horsley, & van Lieshout, 2009a, van der Schoot, Vasbinder, Horsley, Reijntjes, & van Lieshout, 2009b, van der Schoot, Horsley, & van Lieshout, 2010). Specifically, the goal was to find out whether poor comprehending children differ from good comprehending children in the extent to which they *construct* a richly elaborated situation model, or whether the difference between the two groups mainly resides in the ability to *update* the situation model (or working memory) representation.

How does the inconsistency detection task—in which, in the present study, we varied the character of the protagonist—differentiate between these two types of problems? Above, we have already explained how the inconsistency detection task is more or less designed to reveal difficulties in situation model *construction*. Readers who do not represent the protagonist's character in their situation model have a far worse chance of *detecting* the inconsistency of the protagonist's later action. At least, this is to be expected in the global condition in which the character information is lost from working memory. In the local condition, however, even the poor situation model constructors (i.e. readers who leave the character elaboration out of their situation model) are able to detect inconsistencies since the character information against which the target action needs to be evaluated is still active in working memory. In this study, it is therefore the way in which the local inconsistencies are dealt with that can ultimately differentiate between readers who try to actively *update* their situation model and readers who do not. The former make an attempt to *resolve* the inconsistencies and restore comprehension in the situation model. The latter do not take reparative action, either because they lack knowledge of comprehension-repair strategies or because they do not know how to apply them (Gersten et al., 2001).

To investigate comprehension monitoring in 10- to 12-year-old good and poor comprehenders, we conducted two experiments both of which employed the same inconsistency detection task. In Experiment 1, we examined situation model construction and updating by measuring the participants' self-paced reading times. In Experiment 2, we used the eye tracking method and measured their eye fixations and regressions. To be able to examine group differences adjusted for working memory capacity, we included this variable as covariate in both experiments. Working memory capacity has been found to be related to comprehension skill in tasks that require the simultaneous storage and processing of verbal information

(Cain et al., 2004; Seigneuric, Ehrlich, Oakhill, & Yuill, 2000). This implies that working memory capacity may affect performance in the present task and reduce the differences between good and poor comprehenders that are hypothesized below.

Experiment 1

In addition to looking at the participants' answers to questions following the experimental texts, we addressed the research questions by looking at the pattern of reading times as determined by means of the self-paced (sentence-by-sentence) moving window method (Just, Carpenter, & Woolley, 1982). In line with the findings in experienced adult readers (e.g. Huitema et al., 1993; Long & Chong, 2001), we hypothesized that in the global and local condition, the good comprehenders would spend more time reading the target sentence when it was inconsistent with the earlier-described character than when the same sentence was consistent with the character description. It is assumed that these longer reading times reflect, at least in part, the processes involved in the detection and resolution of consistency violations (e.g. Albrecht & O'Brien, 1993).

For the poor comprehenders, two possibilities were hypothesized. The first hypothesis departs from the assumption that poor comprehenders are impaired in their ability to *construct* a richly elaborated situation model (and thus fail to incorporate the character information) but that they do try to update the model (or working memory system) when the information which they did build-in is contradicted and in need of revision. Thus, under the first hypothesis, we expected poor comprehenders to read inconsistent actions more slowly than consistent ones in the local condition (in which the protagonists' actions can be viewed in the light of their character as the character information is still active in working memory) but not, or much less so, in the global condition (in which the target actions may not be evaluated against any character information as this information is likely to be lost from memory). The starting assumption underlying the second possible hypothesis is that poor comprehenders are especially poor at *updating* the situation model. This implies that even when poor comprehenders are able to detect an inconsistency they are not expected to adjust their reading in an attempt to resolve it. Hence, under the second hypothesis, no differences in reading times were anticipated between inconsistent and consistent actions in both the global and local condition.

Method

Participants

The participants were 31 children (18 boys/13 girls) with high reading comprehension levels (good comprehension group) and 26 children (16 boys/10 girls) with low reading comprehension levels (poor comprehension group). The groups were matched on age ($M = 11.3$, $SD = .6$ vs. $M = 11.5$, $SD = .6$, respectively, $t(55) = 1.32$, *ns*) and decoding skill ($M = 81.52$, $SD = 12.50$ vs. $M = 77.19$, $SD = 12.42$, respectively, $t(55) = 1.31$, *ns*). Decoding skill was assessed by the

EMT, a standardized Dutch word reading test (Brus & Voeten, 1999). The EMT showed that the word decoding skills in all participating children were around their grade level and thus more or less automatized.

The children attended either Grade 5 or 6 at a regular primary school in the Netherlands. They were native speakers of Dutch and had normal or corrected-to-normal vision. Exclusion criteria were any neurological disorder and IQ less than 85 (IQ was estimated on the basis of two subtests (Vocabulary and Block Design) of the Dutch version of the Wechsler Intelligence Scale for Children—Revised (van Haasen, 1986)). For all participating children informed consent was obtained from their parents or care-takers.

Assessment of reading comprehension level

Children were classified as good or poor comprehenders based on their performance on the (Grade 5 and Grade 6 versions of the) standardized Test for Reading Comprehension of the Dutch National Institute for Educational Measurement (CITO) (“Toets Begrijpend Lezen”, Staphorsius & Krom, 1998). This test is part of the standard Dutch CITO pupil monitoring system and is designed to determine general reading comprehension level in primary school children.

To classify children as good or poor comprehenders, we compared the test scores with their age-appropriate norm scores. Children whose scores were among the highest 25% of the norm scores were classified as good comprehenders ($M = 70.71$, $SD = 11.24$), children scoring among the lowest 25% were classified as poor comprehenders ($M = 35.96$, $SD = 12.49$). The educational age-norms for average reading comprehension level were obtained in extensive standardization studies on reading in the Dutch population of primary school children (Staphorsius & Krom, 1998).

Materials and design

The experimental texts had the following structure. Each text began with a three- or four-sentence paragraph introducing a person. This was followed by a three- or four-sentence *elaboration paragraph* in which the characteristics (or preferences) of the person are described. These characteristics were either consistent or inconsistent with an action performed by the person later in the text. The elaboration paragraph was followed by a *filler paragraph*, which contained either one sentence (local condition: mean (M) number of words = 11.4, $SD = 1.9$) or five to seven sentences (global condition: $M = 77.8$, $SD = 5.1$). The text continued with the *target sentence* in which the person performs the action that is either consistent or inconsistent with his character described in the elaboration paragraph. For example, the target sentence *Peter ordered a cheeseburger and fries* is consistent with his description as a fast-food addict but inconsistent with his description as a vegetarian (example taken from Albrecht & O’Brien (1993) and Long & Chong (2001)). In the local condition, the character elaboration (in this case, Peter’s food preferences) was still active in working memory while reading the target sentence as the character description and the target action were only separated by one intervening sentence.

On the other hand, in the global condition, the long filler paragraph ensured that the character description was eliminated from memory (see Albrecht & O'Brien, 1993; Long & Chong, 2001). Importantly, all filler paragraphs were content-neutral with regard to the target action. The target sentence was followed by a second critical sentence (*post-target sentence*) so as to detect possible spillover effects. Each text ended with a brief closing paragraph.

Thus, in total, there were 4 within-subjects conditions formed by the crossing of 2 factors: location (local vs. global) and consistency (consistent vs. inconsistent). Each subject was presented with 8 experimental texts, 2 in each condition. The stimuli were arranged in 4 material sets, each containing the 8 texts. Each set was presented to 8 good comprehenders (with the exception of one set which was presented to 7 good comprehenders) and 7 poor comprehenders (with the exception of two sets which were presented to 6 poor comprehenders).

In order to ensure full combination of conditions and materials (and control for text effects), the different versions of each text were counterbalanced across the material sets by means of a 4×4 Latin square design. Thus, across sets and across participants, each story occurred equally often in the local/consistent, local/inconsistent, global/consistent and global/inconsistent version. The order in which the texts were presented in each set was pseudorandomized.

Self-paced moving window method

Reading times were collected by means of the self-paced moving window method (Just et al., 1982). In this method, subjects are presented with passages of text on a computer screen in such a way that words are masked by X's. A window reveals one word, or one sentence, at a time to the reader. When the reader is finished comprehending a word/sentence, they press a key to move the window to the next word/sentence. Assuming that information is processed as soon as it is perceived (Just & Carpenter, 1980), the key pressing latencies (i.e. reading times) in the self-paced moving window task reflect the processing of the word/sentence contained in the window.

In this experiment, we made use of the self-paced sentence-by-sentence procedure. Reading time on the target and post-target sentence was thus defined as beginning when the sentence was first revealed and lasting until the next key press.

Procedure

Before the start of the inconsistency detection task, children were informed that the study was designed to examine reading comprehension of text displayed on a screen. They were instructed to read at their normal rate and to comprehend what they were reading as well as they could. As for the text presentation procedure, seven or eight masked sentences were presented on one screen. Each sentence consisted of either one or two lines of text. The target and post-target sentences always fit on one line of the screen. At the start, the moving window was placed on the first sentence of the first text. Subjects used the down-arrow key to see each

successive sentence in a text. Pressing the down-arrow key on the last sentence on a screen caused the window to move to the first sentence on the next screen. The last sentence of each text revealed the words “NEXT TEXT” indicating the end of the text and the beginning of the next one. The first experimental sentence on the first screen was preceded by four test sentences.

The experimental texts were interspersed with short filler texts to prevent the subjects from becoming aware of the purpose of the experiment. After each text (i.e. each time the moving window revealed the words “NEXT TEXT”), a Yes or No comprehension question was asked. The question pertained to the situational content of the text just read and was asked orally by the experimenter. After answering the question, the subjects were instructed to directly jump to the next text (by pressing the down-arrow key). To prevent them from becoming tired, participants were given a break halfway through the experiment.

The children were tested at school. They completed the inconsistency detection task, sentence span task (see below), and word reading test in a silent room and the test for reading comprehension in the classroom (whole-class test taking). The experimenter was always present. In total, the experiment lasted approximately 1 h and 45 min.

Working memory capacity measure

To assess working memory capacity, an adaptation of the Sentence Span Task (Daneman & Carpenter, 1980; Swanson, 1994; Swanson, Cochran, & Ewers, 1989) was administered. In this task, participants were asked to read aloud groups of unrelated sentences (7–10 words in length). After reading, their task was to recall the last words of the sentences in the right order, and to answer a comprehension question about one of the sentences. The purpose of the question was to make sure that children read for comprehension and did not merely try to remember the target words. The number of sentences in the groups gradually increased. Working memory capacity was defined as the largest group of end words recalled (with the additional requirement that the comprehension question was answered correctly). The sentence span task measures verbal working memory capacity and predicts performance on reading tasks as well as other related tasks (Daneman & Green, 1986; Masson & Miller, 1983).

Results

Reading time on the target sentence

On the reading times, overall $2 \times 2 \times 2$ analyses of variance (ANOVA) on the subject (F_1) and item (i.e. text) (F_2) means were conducted with consistency and location as within-subject variables and with comprehension group as the between-subject variable. In Fig. 1, reading time on the target sentence (in milliseconds) is presented as a function of consistency (consistent vs. inconsistent), location (local vs. global) and comprehension group (good vs. poor). The results of the ANOVA showed that, on the subject means, reading times were faster in the global than local

Fig. 1 Reading time on the target sentence as a function of consistency (consistent vs. inconsistent) and location (local vs. global) (+SE)

condition ($F_1(1,55) = 5.40, p < .05, \eta_p^2 = .09$; $F_2(1,2) = .45, ns$), and that good comprehenders tended to read the target sentence faster than poor comprehenders ($F_1(1,55) = 3.11, p < .1, \eta_p^2 = .05$; $F_2(1,2) = 9.39, p < .1, \eta_p^2 = .82$; 0.01 is considered a small partial eta-squared effect size, 0.06 is considered a medium effect, and 0.14 is considered a large effect Stevens, 2002). More important here is that the effect of consistency varied as a function of the location of the target sentence in poor but not in good comprehenders. In the local condition, both good and poor comprehenders read the target sentences more slowly when they were inconsistent with the character description than when they were consistent with it (Consistency $\times$ Group: $F_1(1,55) = .35, ns$; $F_2(1,2) = .43, ns$). In the global condition, on the other hand, the inconsistency effect was present in the good but not in the poor comprehenders (who read inconsistent sentences as fast as consistent ones; Consistency $\times$ Group interaction: $F_1(1,55) = 4.40, p < .05, \eta_p^2 = .07$; $F_2(1,2) = 9.90, p < .1, \eta_p^2 = .83$). In other words, poor comprehenders slowed down their reading on inconsistent sentences when the target action directly followed the character elaboration, but not when the target sentence and character elaboration were interspersed with a long filler paragraph (Consistency $\times$ Location: ($F_1(1,55) = 4.44, p < .05, \eta_p^2 = .15$; $F_2(1,2) = 14.33, p < .1, \eta_p^2 = .88$). The above pattern of results was evident in the significant (F_1) and marginally significant (F_2) Consistency $\times$ Location $\times$ Group interaction ($F(1, 55) = 4.51, p < .05, \eta_p^2 = .08$; $F_2(1, 2) = 12.99, p < .1, \eta_p^2 = .87$). The three-way interaction disappeared, however, when working memory capacity was entered as a covariate ($F_1(1,54) = 2.37, p = .13$).

Reading time on the post-target sentence

In Table 1, reading time on the post-target sentence (in milliseconds) is presented as a function of consistency (consistent vs. inconsistent), location (local vs. global) and comprehension group (good vs. poor). The analyses on the reading times on the sentence immediately following the target sentence showed that good comprehenders had faster reading times than poor comprehenders ($F_1(1,55) = 7.79, p < .01, \eta_p^2 = .12$; $F_2(1,2) = 15.99, p = .057, \eta_p^2 = .89$). In addition, both groups exhibited an effect of location ($F_1(1,55) = 4.16, p < .05, \eta_p^2 = .07$; $F_2(1,2) = 77.94, p < .05, \eta_p^2 = .98$), signifying that they read the post-target sentence slower in the local than global condition. However, analyses on the post-target sentence reading times failed to show any significant differences between consistent and inconsistent sentences.

Mixed-effects modeling

In addition to ANOVA, we also analyzed the data using mixed-effects modeling with the maximum likelihood method to calculate parameter estimates. Mixed-effects modeling is a statistical technique for data repeatedly observed from the same subjects and/or materials. It is gaining popularity over the conventional methods because differences between individuals and differences between materials are modeled by means of (crossed) random effects, resulting in, among other things, increased power, a better account of heterogeneity of variance, and a better use of available data (e.g. Baayen, Davidson, & Bates, 2008; Bicknell, Elman, Hare, McRae, & Kutas, 2010; Kliegl, 2007, for an application of mixed-effects modeling in reading comprehension research using, respectively, self-paced reading and eye tracking).

In the present analyses, subjects and materials (i.e. experimental sentences) were thus treated as random effects, and consistency, location and group as fixed effects. The fixed effects were coded as follows: Group (good = 1, poor = 0), Consistency (consistent = 1, inconsistent = 0) and Location (local = 1, global = 0). To control for word-related effects, reading times were normalized by the number of words in the sentence, and the sentences' average log word frequency (obtained from CELEX, the database from the Dutch Centre for Lexical Information Baayen, Piepenbrock, & Gulikers, 1995) was added as covariate to exclude a word frequency

Table 1 Post-target sentence reading times (in milliseconds) as a function of consistency (consistent vs. inconsistent), location (local vs. global) and comprehension group (good vs. poor) (+SE)

	Consistency	
	Consistent	Inconsistent
Good comprehenders		
Local	2,735.44 (256.23)	2,791.81 (241.30)
Global	2,615.06 (194.38)	2,733.71 (298.51)
Poor comprehenders		
Local	3,914.98 (644.06)	4,055.21 (245.83)
Global	3,435.35 (304.16)	3,383.54 (352.84)

confound. We used the logarithm of word frequency because reading times are linearly related to the logarithm of word frequency, not to raw word frequencies (Haberlandt & Graesser, 1985; Just & Carpenter, 1987).

Results confirmed the findings reported above. The analysis on the target sentence reading times showed main effects of Group ($F(1, 57) = 4.42, p < .05$), Consistency ($F(1, 399) = 6.41, p < .05$) and Location ($F(1, 399) = 6.83, p < .01$) (The F values are produced by *Type III Tests of Fixed Effects*, and since the design is balanced, they are expressed as the ratio of the appropriate sums of squares). More importantly, the critical Consistency $\times$ Location $\times$ Group interaction was significant in predicting the reading times on the target sentence ($F(1, 399) = 9.55, p < .005$) (model fit: $-2 \log \text{likelihood} = 6,225.77$). The theoretically meaningful parameter estimates for Consistency ($PE = 68.89, SE = 39.59$) and Consistency $\times$ Location $\times$ Group ($PE = 234.67, SE = 75.92$) were, respectively, marginally significant ($p < .1$) and significant ($p < .005$). Results for the post target sentence indicated a main effect of Group ($F(1, 56.87) = 8.02, p < .01$) and Location ($F(1, 397.74) = 5.40, p < .05$), but no significant effects of Consistency, neither as a main effect, nor through its interactions with the other factors (model fit: $-2 \log \text{likelihood} = 8,078.72$). Parameter estimates for Consistency ($PE = 51.81, SE = 298.39$) and Consistency $\times$ Location $\times$ Group ($PE = 254.31, SE = 572.22$) were also not significant.

Comprehension questions

In Fig. 2, the average number of correct answers to the comprehension questions is presented as a function of consistency (consistent vs. inconsistent), location (local

Fig. 2 Number of correct answers to the comprehension questions as a function of consistency (consistent vs. inconsistent) and location (local vs. global) (+SE)

vs. global) and comprehension group (good vs. poor). As can be seen from Fig. 2, good comprehenders answered more questions correctly than poor comprehenders ($F(1,55) = 11.06, p < .005, \eta_p^2 = .17$). Of more significance is that the interaction between group and consistency depended on location, yielding a three-way interaction between those factors ($F(1,55) = 4.97, p < .05, \eta_p^2 = .08$). In the global condition, poor comprehenders gave more incorrect answers to inconsistent than consistent texts. Good comprehenders did not show this effect of consistency (Consistency $\times$ Group: $F(1,55) = 8.01, p < .01, \eta_p^2 = .13$), nor did both groups in the local condition (Consistency $\times$ Group: $F(1,55) = .00, ns$). The three-way interaction remained marginally significant after controlling for working memory ($F(1,54) = 3.22, p < .1, \eta_p^2 = .06$).

Discussion

Experiment 1 investigated comprehension monitoring in 10–12 years old children differing in general reading comprehension skill. The children's reading times were measured as they read narrative texts in which an action of the protagonist was consistent or inconsistent with a description of the protagonist's character given earlier. The character description and action were adjacent (local condition) or separated by a long filler paragraph (global condition). The goal was to find out whether poor comprehenders mainly differ from good comprehenders in situation model construction or updating. In this study, situation model construction specifically refers to the richness of the model, which here concerns the question of whether or not a reader incorporated the character information. Situation model updating refers to the adaptability and solution-readiness of the model in case of a consistency violation.

For the good comprehenders, we hypothesized that they would slow down their reading on inconsistent actions when compared to consistent ones in both the local and global condition. The target sentence reading time analysis confirmed this hypothesis at both the subject- and item-level, and in both the ANOVA and mixed model analysis, indicating that good comprehenders detected the inconsistencies and made an attempt to resolve them (e.g. Albrecht & O'Brien, 1993). Probably, the extra time good comprehenders spent on an inconsistent target sentence reflected their effort to double-check the inconsistency and/or think up possible resolutions. It is important to note that from the inconsistency effect they displayed in the global condition it can be inferred that good comprehenders must have represented the character elaboration in the situation model they constructed from the text. Otherwise, the character information would have been lost from their working memory when they read the target sentence as a consequence of which they would have missed its inconsistency (thereby leaving their above-mentioned updating skills unused).

The results for the good comprehenders are in line with the results obtained in experienced adult readers (e.g. Huitema et al., 1993; Long & Chong, 2001). Yet, this study is one of the few demonstrating the current situation model construction and updating skills in children. Apparently, skilled 10–12 years old readers who have automatized their lower-level word decoding skills (as was the case here) have

freed up enough processing capacity for the higher-level processes to enhance reading comprehension. In particular, the present study showed that they can carry out reading comprehension strategies to aid the construction and updating of a coherent and richly-connected situation model of a text.

In poor comprehenders, we found that reading times were slower on inconsistent than consistent target sentences in the local but not the global condition. This pattern of reading times, found at both the subject- and item-level, and in both the ANOVA and mixed model analysis, provides clear support for one of the two proposed hypotheses. This hypothesis says that poor comprehenders find difficulty in constructing a rich situation model. That is, they tend to leave situation-relevant information out of the model, including information that could be of use to them in interpreting later information in the text. In the present inconsistency detection task, this implies that poor comprehenders presumably failed to represent the character of the protagonist in their situation model in most stories. At least, this explains why they did not show any effect of inconsistency in the global condition. The inconsistencies simply went unnoticed as the information that was contradicted was neither present in their (short-term) working memory nor (long-term) situation model and therefore was no longer accessible.

Here, it is important to realize that the inconsistency effect which is exhibited by poor comprehenders in the local condition rules out an explanation in terms of impaired *updating* ability. Under the assumption that in this condition the character elaboration is still active in working memory, this finding indicates that when poor comprehenders understand that new information contradicts previous information they do make an attempt to resolve the inconsistency and restore comprehension in their current (working memory) representation of the text. As in good comprehenders, the extra time poor comprehenders spend reading an inconsistent action statement probably reflected their effort to double-check the inconsistency and/or elaborate possible resolutions. It should be mentioned, however, that the effects of inconsistency were limited to the target sentence reading times in that they did not extend into the spillover sentence immediately following the action sentence. This finding, obtained in both the subject and item analyses, and in both the ANOVA and mixed model analysis, is intriguing but difficult to interpret, especially since similar studies as the present one have previously demonstrated such a spillover effect (Albrecht & O'Brien, 1993; Long & Chong, 2001; Poynor & Morris, 2003).

The above-described differences in situation model construction and updating between good and poor comprehenders were not only evident in the reading time data but also in the data on the comprehension questions (which were assumed to at least partially tap the situation model representation that the reader constructed from the text). It was found that the negative effect of consistency violations on comprehension were restricted to the poor comprehenders' answers given in the global condition. As was the case with the reading time data, this can be accounted for by assuming that poor comprehenders did not build character information into their situation model as a consequence of which they had a worse chance of noticing global inconsistencies. So, the comprehension question data coincide with the reading time data, and together, they indicate that poor comprehenders found it

especially hard to build and maintain a coherent and integrated situation model from globally inconsistent texts.

This leaves us with the question of how to explain the influence of verbal working memory capacity. As was demonstrated in the analysis of covariance, the interacting effects of consistency, location and group disappeared after controlling for the participants' sentence span. The most dominant conceptions of both verbal and non-verbal working memory assume that working memory consists of a storage component, in which information is maintained during processing, and a processing component, which coordinates the mental activities that are required by the task at hand (Baddeley, 1986; Shah & Miyake, 1996). In the present sentence span task, the storage component involved reading aloud a small set of sentences and the processing component involved remembering the last word of each sentence.

It is generally assumed that the relationship between verbal working memory tasks, such as the present sentence span task, and reading comprehension is mediated by the processing component of verbal working memory, rather than by the storage component (e.g. Daneman, 1987; Daneman & Tardif, 1987). According to this assumption, the lower verbal memory span of poor comprehenders demonstrated in this and other studies is thus a direct reflection of their weak language comprehension skills. Interestingly, Nation, Adams, Bowyer-Crane and Snowling (1999) showed that poor comprehenders had lower spans only on memory tasks that called upon semantic processing skills. On a series of other non-verbal/non-linguistic tasks, including one tapping spatial working memory, poor comprehenders were not impaired. From this, Nation et al. (1999) concluded that the poor comprehenders' difficulties with reading and language skills can not be related to a *general* processing capacity weakness. Instead, they came to the conclusion that "...the memory difficulties associated with poor reading comprehension are *specific* to the verbal domain and are a concomitant of language impairment, rather than a cause of reading comprehension failure" (p. 139). Probably, this may account for the finding that the interaction effects of consistency and location depended on the poor and good comprehenders' verbal working memory capacity. That is to say, the effects were weakened supposedly because the covariate represented the same conceptual measure (i.e. semantic processing skill) as the patterns of the dependent variables.

Experiment 2

In Experiment 1, a self-paced reading procedure was adopted to investigate situation model construction and updating in children. Even though it may not be as ecologically valid as the eye tracking method (which provides a more natural reading experience), we assumed that the self-paced moving window method has the capacity to tap into these (and other) reading comprehension processes and reveal the hypothesized differences therein between good and poor comprehenders. This assumption is supported by results obtained by, among others, Just et al. (1982) and Rinck et al. (2003). Just et al. (1982, p. 228) qualitatively and quantitatively compared the self-paced moving window method with the eye tracking method and

concluded that “the word-level effects are generally similar”. Specifically, the correlation between the mean fixation duration on each word (averaged over subjects) and the mean button-pressing latency was .57. More recently, Rinck et al. (2003) compared two experiments on the processing of temporal text information. These experiments were the same except that one used the eye tracking method and the other collected self-paced reading times (sentence-by-sentence). Rinck et al. concluded that the results of both experiments were “in perfect accordance” (p. 81).

Nevertheless, it should be acknowledged that the self-paced moving window method has some disadvantages. One potential problem with the moving window method when a word-by-word presentation is used is that it may induce an artificial buffering strategy because readers cannot press the button as fast as they can comprehend the word (e.g. Danks, 1986). Obviously, this limitation could not have played a role here since in Experiment 1 sentence-by-sentence reading times were collected (see also Magliano, Graesser, Eymard, Haberlandt, & Gholson, 1993). On the other hand, two other limitations may have affected the present results and their generalizability. First, the self-paced moving window method tends to slow readers down (Rayner, 1998). Second and more importantly, it prevents readers from looking back, or prevents the experimenter from observing any look-backs (in case a cumulative method is used; see Rayner, 1998).

To overcome these limitations, we conducted a second experiment in which the readers’ eye movements and fixations were monitored as they read the same experimental texts used in Experiment 1. In Experiment 2, readers were therefore free to reread whenever they felt necessary. This allowed us to examine whether readers engage in any other comprehension-repair strategy than that of reading the inconsistent target sentence more slowly (see also Hyönä et al., 2003). Of course, slowing down may be the right strategic choice if they want to double-check the inconsistency or think up some resolutions for the contradiction (Yuill & Oakhill, 1991). But readers might also wonder whether they misunderstood earlier parts of the text. In that case, the right thing to do would be to look back at the possible source of the inconsistency, i.e. the character description (Hyönä et al., 2003; Zabrocky & Ratner, 1986).

Thus, Experiment 2 enabled us to distinguish between initial reading and rereading patterns while seeking converging evidence for the conclusions of Experiment 1. In a more general sense, Experiment 2 provided the opportunity of a two-method comparison as any differences in results between the two experiments will be related to the two methods of presentation and data collection (self-paced moving window method vs. eye tracking method).

Method

In Experiment 2, the materials, design and procedure were the same as in Experiment 1. The only differences between the two experiments were the method of presentation and data collection (self-paced moving window method vs. eye tracking method) and the number of subjects in each comprehension group. Here, we describe only those aspects of Experiment 2 that are related to these differences.

Participants

The participants were 16 children (7 boys/9 girls) with high reading comprehension levels (scores in the highest quartile on the standardized reading comprehension test: $M = 78.44$, $SD = 8.54$) and 15 children (7 boys/8 girls) with low reading comprehension levels (within lowest quartile: $M = 39.20$, $SD = 9.12$), all of them with measured IQ's above 85. As in Experiment 1, the good and poor comprehension groups were matched on age ($M = 11.6$, $SD = .5$ vs. $M = 11.9$, $SD = .6$, respectively, $t(29) = 1.65$ *ns*) and decoding skill ($M = 83.49$, $SD = 12.03$ vs. $M = 76.24$, $SD = 9.63$, respectively, $t(29) = 1.84$, *ns*).

Materials and design

Each material set was presented to 4 good comprehenders and 4 poor comprehenders, with the exception of one set which was presented to 3 poor comprehenders.

Eye tracking method

Eye fixations/movements were recorded during the reading of the texts with the EYELINK® II eye-tracker, an infrared video-based tracking system manufactured by SR Research Ltd. (Mississauga, Canada). The EYELINK® II system uses a corneal reflection method in combination with pupil tracking, permitting stable tracking of eye position regardless of muscle tremor, environmental vibration, or headband slippage. Although viewing was binocular, signals were recorded from one eye only. The best eye to record was automatically selected during a calibration procedure. Calibration was accepted when the worst error in gaze position was smaller than 1.5° and average error was smaller than 1.0° . The cameras sampled pupil location at the rate of 250 Hz. Participants in the experiment were seated such that the distance between the monitor (51 cm diagonal with a resolution of $1,025 \times 768$) on which the texts were displayed and their eyes was approximately 70 cm. At this distance, three letters of the presented text subtended 1° of visual angle.

Procedure

Before the experiment started, the eye-tracker was adjusted and calibrated using a 5-point calibration grid presented on the computer screen. After each text, a 1-point calibration was performed to correct for possible drifts in gaze position. A practice session preceded the reading of the first experimental text to accustom the children to the eye-tracking equipment.

On each screen, seven or eight lines of text were presented, except for the last screen of each text, on which the remaining lines were displayed. In total, each text consisted of two or three screens. The children could move from one screen to the next by pressing the space bar.

Eye fixation measures

For each text, eye fixation measures for two regions of interest were calculated: the one-sentence target action and the three- or four-sentence character description. We examined first pass duration and wrap-up times for the target sentence and regressions into the character description. First pass duration on the target sentence is defined as the sum of the duration of all fixations on the sentence from the first fixation on the sentence until the first time that the reader exits the sentence. This eye fixation measure reflects the initial processing on the target sentence, and is indicative of the comprehension-repair processes of readers who disbelieved the described action and therefore double-checked the potential inconsistency or thought up possible resolutions for the contradiction. Because the length of the target action varied across sentences, the sum of fixations was divided by the number of characters in the sentence to yield a millisecond-per-character measure. We acknowledge that, in general, there may be some problems associated with the millisecond-per-character measure (Rayner, 1998). However, since all target sentences in the present study had approximately the same number of words, those problems are unlikely to be associated with the present data (see also Rayner, Sereno, Morris, Schmauder, & Clifton, 1989; Schmauder, 1991).

The wrap-up effect refers to the finding that reading times are longer for the final one or two words of a sentence than for non-boundary words (Just & Carpenter, 1980; Rayner, Kambe, & Duffy, 2000), and is taken as an index of the integration of words within the sentence but also of the integration of the sentence with the preceding text. The wrap-up effect is relevant here because, among other things, it has been shown to result from “an attempt to handle any inconsistencies that could not be resolved within the sentence” (Just & Carpenter, 1980, p. 345). Here, the wrap-up effect is defined as the total fixation time on the sentence’s final two words.

In addition to first pass duration, we determined the frequency with which readers looked back to the character description given that they had initially processed this critical region and moved on to at least the target sentence (i.e. regressions). This processing measure reflects the reparative action of readers who wondered whether they misunderstood the possible source of the inconsistency i.e. the elaboration of the protagonist’s character.

Results and discussion

First pass duration on target sentence

One child obtained values of first pass duration that were more than 3 SD from the mean. The sentence measure from this subject was subsequently removed from the data set. Additionally, short fixation durations of less than 50 ms and long fixation durations of more than 1,250 ms were excluded from the analysis.

On the first pass duration, an overall $2 \times 2 \times 2$ ANOVA on the subject (F_1) and item (i.e. text) (F_2) means were performed with consistency and location as within-subject variables and with comprehension group as the between-subject variable. In Fig. 3, first pass duration on the target sentence (in milliseconds per character) is

Fig. 3 First pass duration on the target sentence (in milliseconds per character) as a function of consistency (consistent vs. inconsistent) and location (local vs. global) (+SE)

presented as a function of consistency (consistent vs. inconsistent), location (local vs. global) and comprehension group (good vs. poor). The results of the ANOVA showed a main effect of comprehension group ($F_{1(1,28)} = 19.55$, $p < .001$, $\eta_p^2 = .41$; $F_{2(1,2)} = 135.28$, $p < .01$, $\eta_p^2 = .98$), indicating that, overall, poor comprehenders took longer to read the target sentence than good comprehenders. Of more relevance here is that the effect of consistency tended to depend on its location in poor but not in good comprehenders. In the local condition, both comprehension groups read the target sentences more slowly when they were inconsistent with the character elaboration than when they were consistent with it (Consistency $\times$ Group: $F_{1(1,28)} = .06$, ns ; $F_{2(1,2)} = .41$, ns). In the global condition, on the other hand, this inconsistency effect was present in the good comprehenders but reversed in the poor comprehenders (who read the target sentences faster when they were inconsistent) (Consistency $\times$ Group interaction ($F_{1(1,28)} = 3.16$, $p < .1$, $\eta_p^2 = .10$; $F_{2(1,2)} = 56.54$, $p < .05$, $\eta_p^2 = .97$). The above pattern of results was reflected in the marginally significant (F_1) and significant (F_2) Consistency $\times$ Location $\times$ Group interaction ($F_{1(1,28)} = 3.01$, $p < .1$, $\eta_p^2 = .10$; $F_{2(1,2)} = 22.24$, $p < .05$, $\eta_p^2 = .92$). When working memory capacity was entered as a covariate, the three-way interaction was not significant ($F(1,27) = 1.26$, ns). An explanation for the effects of working memory capacity is given in the Discussion of Experiment 1.

Sentence wrap-up times

Equivalent results as above were found when the analyses were conducted on the sentence wrap-up times (see Fig. 4). Good comprehenders displayed faster wrap-up times than poor comprehenders ($F_{1(1,28)} = 21.76$, $p < .001$, $\eta_p^2 = .44$;

Fig. 4 Target sentence wrap-up time as a function of consistency (consistent vs. inconsistent) and location (local vs. global) (+SE)

$F_{2(1,2)} = 103.37$, $p < .05$, $\eta_p^2 = .98$), and, overall, inconsistent sentences yielded longer wrap-up times than consistent sentences ($F_{1(1,28)} = 4.84$, $p < .05$, $\eta_p^2 = .15$; $F_{2(1,2)} = 560.86$, $p < .005$, $\eta_p^2 = .97$). Like in the first pass duration analyses, both comprehension groups had slower wrap-up times on inconsistent than consistent sentences in the local condition ($F_{1(1,28)} = .05$, ns ; $F_{2(1,2)} = .43$, ns), while in the global condition, this inconsistency effect was present in the good comprehenders but reversed in the poor comprehenders ($F_{1(1,28)} = 4.36$, $p < .05$, $\eta_p^2 = .14$; $F_{2(1,2)} = 244.95$, $p < .005$, $\eta_p^2 = .99$). The Consistency $\times$ Location $\times$ Group interaction was significant in the analyses on both the subject (F_1) and item (F_2) means ($F_{1(1,28)} = 4.78$, $p < .05$, $\eta_p^2 = .15$; $F_{2(1,2)} = 134.47$, $p < .01$, $\eta_p^2 = .99$).

Mixed-effects modeling

As for the self-paced reading times, the first pass durations (FPD) and wrap-up times (WT) were also analyzed using mixed-effects modeling with location, consistency and group as fixed effects and random effects for subjects and materials. Average log word frequency was added as the covariate in both analyses, and first pass duration was normalized for the number of words in the sentence. Results corroborated the above findings, showing main effects of Group (FPD: $F_{1(29.67)} = 21.73$, $p < .001$; WT: $F_{1(29.95)} = 23.30$, $p < .001$) and Consistency (FPD: $F_{1(208.76)} = 3.07$, $p < .1$; WT: $F_{1(208.38)} = 13.09$, $p < .001$). More importantly, the analyses confirmed the critical interaction between these two factors and Location (FPD:

$F(1,208.76) = 4.85, p < .05$; WT: $F(1,208.38) = 6.77, p < .05$) (model fit (FPD): $-2 \log \text{likelihood} = 2,790.97$; model fit (WT): $-2 \log \text{likelihood} = 3,168.93$). The theoretically critical parameter estimates for Consistency (FPD: $PE = 44.35, SE = 19.72$; WT: $PE = 91.02, SE = 43.91$) and Consistency $\times$ Location $\times$ Group (FPD: $PE = 84.14, SE = 38.19$; WT: $PE = 221.28, SE = 85.03$) were significant ($p < .05$) in both the FPD and WT analyses.

The above patterns of first pass durations and wrap-up times are similar to the pattern of reading times obtained in Experiment 1. Following the same line of reasoning as before, we conclude from both the subject and item analyses, and from both the ANOVA's and mixed model analysis, that poor comprehenders are impaired in their ability to construct a richly elaborated situation model. They failed to represent the character information into the model as a consequence of which they were not able to see the inconsistencies in the global condition. However, at this time, we cannot offer a satisfactory explanation for the finding that poor comprehenders spent *less* time reading the (end of the) inconsistent target sentences.

Regressions

The first thing that became apparent while looking over the results was that both poor and good comprehenders were highly unlikely to go back and reread the character description. In the four experimental conditions, the proportion of readers who looked back to the character description *at least once* ranged from 7% (poor comprehenders in the local/consistent condition) to 63% (good comprehenders in the global/inconsistent condition). Therefore, to avoid interpretative difficulties, we decided to no longer perform analyses of variance on the number (and duration) of regressions (made by the few readers who did look back) but rather loglinear analyses with regressions/no regressions (readers who did and did not make one or more regressions) as a categorical variable.

In Table 2, the number of readers who did and did not regress back to the character description at least once is presented as a function of comprehension group (good vs. poor), location (local vs. global) and consistency (consistent vs. inconsistent). From Table 2, it can be seen that the pattern of regressions coincides with the pattern of first pass durations. In the local condition, the inconsistency effect was the same in both comprehension groups. That is, more readers decided to return to the character elaboration after seeing an inconsistent than a consistent action statement (Consistency $\times$ Regressions/NoRegressions: partial $\chi^2(1) = 5.40, p < .05$). In the global condition, the inconsistency effect differed between the comprehension groups. In comparison with consistent target sentences, inconsistent target sentences encouraged more good comprehenders to make regressions back to the character description. This effect was reversed in the poor comprehension group (Group $\times$ Consistency $\times$ Regressions/NoRegressions: $\chi^2(1) = 1.92, p = .16$).

At first glance, the above pattern of results reinforces the idea that poor comprehenders do not build character information into the situation model they construct from a text and therefore overlook the inconsistencies in the global condition. However, the above effects are small and the pattern of results did not yield a significant Group $\times$ Consistency $\times$ Location $\times$ Regressions/NoRegressions

Table 2 Number of readers who did and did not regress back to the character description (at least once) presented as a function of group, location and consistency

Group	Location	Consistency	Regressions/ no regressions	Number of readers
Good	Local	Consistent	Yes	4
			No	12
		Inconsistent	Yes	8
			No	8
	Global	Consistent	Yes	7
			No	9
		Inconsistent	Yes	10
			No	6
Poor	Local	Consistent	Yes	1
			No	13
		Inconsistent	Yes	5
			No	9
	Global	Consistent	Yes	4
			No	10
		Inconsistent	Yes	2
			No	12

interaction in the overall four-way loglinear analysis ($\chi^2(1) = 1.97, p = .16$). The results of this analysis did show a Group $\times$ Regressions/NoRegressions interaction (partial $\chi^2(1) = 8.03, p < .005$), signifying that regressions into the character description, albeit scarce, were made more often by good comprehenders than poor comprehenders.

Comprehension questions

Good comprehenders answered more questions correctly than poor comprehenders ($F(1,28) = 14.26, p < .001, \eta_p^2 = .23$). Across consistency, poor comprehenders answered fewer questions correct in the global condition ($M = 5.00, SD = 1.11$) than in the local condition ($M = 5.36, SD = 1.60$). Good comprehenders showed the reverse pattern (local: $M = 6.25, SD = 1.13$; global: $M = 6.63, SD = 1.09$). However, the Location $\times$ Group interaction was not significant ($F(1,28) = 1.42, ns$). In spite of this, it should be noted that the pattern of results resembles that of Experiment 1, indicating that poor comprehenders have trouble with building a coherent and richly-connected situation model in the global condition (in which the situation-relevant character information was no longer available by the time they read the target sentence).

General discussion

Taken together, the results of Experiment 1 and 2 indicate that poor comprehending children differ from good comprehending children in the extent to which they

construct a richly elaborated situation model. In contrast to good comprehenders, they tend to leave out situation-relevant information that could be used as a basis to interpret later text information. In the present study, this implies that poor comprehenders supposedly failed to represent the character of the stories' protagonists in their situation model. At least, this accounts for the finding that, in poor comprehenders in the global condition, inconsistent target actions neither led to longer reading times (Experiment 1) nor to longer first pass durations or sentence wrap-up times (Experiment 2) nor to more regressions back to the character description (Experiment 2). Probably, poor comprehenders simply did not notice the inconsistencies since, in the global condition, the character information could neither be retrieved any more from their working memory.

Importantly, the results in the local condition yielded a different picture of inconsistency effects than in the global condition. The local condition allowed us to determine whether participants adjusted their reading on the target sentence when the character information was assumed to be still active in working memory. Like good comprehenders, poor comprehenders showed longer reading times (Experiment 1) and longer first pass durations and sentence wrap-up times (Experiment 2) on inconsistent than consistent target sentences. In addition, in both comprehension groups, more readers decided to return to the character description after seeing an inconsistent than a consistent target action. These findings show that, when poor comprehenders do detect a semantic inconsistency, they resemble the good comprehenders in that they try to resolve the inconsistency and restore comprehension in their current (working memory) representation of the text.

In sum, the self-paced reading patterns (obtained in Experiment 1), the initial reading and rereading patterns (obtained in Experiment 2), together with the comprehension question data (obtained in both experiments), lead to the same conclusion: poor comprehenders differ from good comprehenders in the extent to which they enrich the situation model constructed from narrative texts. This supports the conclusion drawn in one of our previous studies that good comprehenders are strategically aware readers who build a more effective mental model of the text than poor comprehenders (e.g. van der Schoot et al., 2008). Poor comprehenders seem not to be impaired, however, in their ability, or readiness, to update the situation model in cases where they detect a consistency violation. Another, more general, methodological, conclusion that can be derived is that self-paced reading times apparently can serve as informative indicators of high-level cognitive processes underlying reading comprehension. At least, this study shows that the self-paced moving window method can provide valid insights on how a text representation is constructed during reading. Finally, the item analyses in both Experiment 1 and Experiment 2 justify the conclusion that the findings generalize beyond our specific sample of language materials, and that they are true for language in general (see Clark, 1973).

Limitations and implications

Although Experiment 2 reinforced the findings obtained in Experiment 1, it is clear that the eye tracking data were less convincing than the self-paced reading

data. In particular, the following question should be addressed. Why did even the good comprehenders hardly feel a need to look back and reread the character description after they had seen an inconsistent action statement, especially since in previous eye movement studies using similar inconsistency detection tasks as in the present study, inconsistent target sentences did cause readers to look back to the probable source of the inconsistency (Poynor & Morris, 2003; Rinck et al., 2003)? The answer may lie in the type of contradictory information. In the Poynor and Morris (2003) study, the information that was contradicted concerned the elaboration of the protagonist's goals. In the Rinck et al. (2003) study, inconsistencies related to the temporal (order) aspects of a described situation. Possibly, it is more difficult to activate—and maintain—a representation of goal and temporal information than of character information. That is, readers might wonder whether they misunderstood the narrated order of events (*Did John arrive before or after Peter?*), or question the goals they inferred for the protagonist (*Precisely, what type of evening did she have in mind again?*), but they are less likely to disbelieve the character information they were presented with. Character information may in this respect be more indisputable in that readers who once activate a representation of, let's say, a vegetarian, and encode this into their situation model, will probably not encounter compelling reason later in the text to doubt the validity of this representation. At least, they will not feel compelled to reinspect the passage about the protagonist's eating habits so as to assure themselves it was not about a meat-eater after all.

Thus, together, the previous and present findings suggest that readers try to reinstate both earlier goal information, temporal information and character information in the face of later inconsistent actions or events, but that only in the case of character information, they do so while staying with the sentence containing the inconsistent action or event. Conversely, readers often deal with inconsistencies regarding the goal of the protagonist or the temporal order of events by also regressing to the source of the inconsistency, attempting to reconcile the two conflicting pieces of information there. Obviously, future research is needed to test this suggestion more directly.

Finally, the implications for instruction and intervention of our findings deserve mention. In our previous work (van der Schoot et al., 2008, 2009b), we argued that in general, the goal is to teach the high-level reading processes (often referred to as reading strategies) that are found to be important for comprehension in laboratory experiments such as the present ones. Specifically, the results of this study underscore the importance of situation model construction as part of the educational methods used in teaching reading comprehension to poor comprehenders (see van der Schoot et al., 2008, 2009b) for a more elaborate discussion on the educational and instructional implications of studies like this one).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model—Maintaining both local and global coherence. *Journal of Experimental Psychology-Learning Memory and Cognition*, 19, 1061–1070.
- 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.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (release 2) [CD-ROM]*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].
- Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press.
- Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*, 63(4), 489–505.
- Brus, B. T. & Voeten, M. J. M. (1999). *Eén-minuut-test: Vorm A en B: Verantwoording en handleiding: Schoolvorderingentest voor de technische leesvaardigheid, bestemd voor groep 4 tot en met 8 van het basisonderwijs* [One-minute-test: Version A and B. Justification and manual: Test for progress at school regarding decoding skill, from Grade 4 through Grade 8 of primary school]. Lisse, the Netherlands: Swets & Zeitlinger.
- Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, 96, 31–42.
- Clark, A. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological. *Journal of Verbal Learning and Verbal Behavior*, 12, 335–359.
- Daneman, M. (1987). Reading and working memory. In J. R. Beech & A. M. Colley (Eds.), *Cognitive approaches to reading*. London: Wiley.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.
- Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, 25, 1–18.
- Daneman, M., & Tardif, T. (1987). Working memory and reading skill re-examined. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading* (pp. 491–508). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Danks, J. H. (1986). Identifying component processes in test comprehension—Comment on Haberlandt and Graesser. *Journal of Experimental Psychology-General*, 115, 193–197.
- de Vega, M. (1995). Backward updating of mental models during continuous reading of narratives. *Journal of Experimental Psychology-Learning Memory and Cognition*, 21, 373–385.
- de Vega, M., Diaz, J. M., & Leon, I. (1997). To know or not to know: Comprehending protagonists' beliefs and their emotional consequences. *Discourse Processes*, 23, 169–192.
- Ehrlich, M. F. (1996). Metacognitive monitoring in the processing of anaphoric devices in skilled and less-skilled comprehenders. In C. Cornoldi & J. V. Oakhill (Eds.), *Reading comprehension difficulties: Processes and remediation* (pp. 221–249). Mahwah, NJ: Erlbaum.
- Ehrlich, M. F., Remond, M., & Tardieu, H. (1999). Processing of anaphoric devices in young skilled and less skilled comprehenders: Differences in metacognitive monitoring. *Reading and Writing: An Interdisciplinary Journal*, 11, 29–63.
- Eme, E., Puustinen, M., & Coutelet, B. (2006). Individual and developmental differences in reading monitoring: When and how do children evaluate their comprehension? *European Journal of Psychology of Education*, 21, 91–115.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working-memory. *Psychological Review*, 102, 211–245.
- Gernsbacher, M. A., Goldsmith, H. H., & Robertson, R. R. W. (1992). Do readers mentally represent characters emotional states. *Cognition & Emotion*, 6, 89–111.
- Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research*, 71, 279–320.
- Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General*, 114, 357–374.
- Huitema, J. S., Dopkins, S., Klin, C. M., & Myers, J. L. (1993). Connecting goals and actions during reading. *Journal of Experimental Psychology-Learning Memory and Cognition*, 19, 1053–1060.

- Hyönä, J., Lorch, R. F., & Rinck, M. (2003). Eye movement measures to study global text processing. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 313–334). Amsterdam: Elsevier.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading—From eye fixations to comprehension. *Psychological Review*, 87, 329–354.
- Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Boston: Allyn & Bacon.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228–238.
- Kliegl, R. (2007). Toward a perceptual-span theory of distributed processing in reading: A reply to Rayner, Pollatsek, Drieghe, Slattery, and Reichle (2007). *Journal of Experimental Psychology: General*, 136, 530–537.
- Long, D. L., & Chong, J. L. (2001). Comprehension skill and global coherence: A paradoxical picture of poor comprehenders' abilities. *Journal of Experimental Psychology-Learning Memory and Cognition*, 27, 1424–1429.
- Magliano, J. P., Graesser, A. C., Eymard, L. A., Haberlandt, K., & Gholson, B. (1993). Locus of interpretive and inference processes during text comprehension—A comparison of gaze durations and word reading times. *Journal of Experimental Psychology-Learning Memory and Cognition*, 19, 704–709.
- Masson, M. E. J., & Miller, J. A. (1983). Working memory and individual differences in comprehension and memory of text. *Journal of Educational Psychology*, 75, 314–318.
- Myers, J. L., O'Brien, E. J., Albrecht, J. E., & Mason, R. A. (1994). Maintaining global coherence during reading. *Journal of Experimental Psychology-Learning Memory and Cognition*, 20, 876–886.
- Nation, K., Adams, J. W., Bowyer-Crane, C. A., & Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology*, 73, 139–158.
- O'Brien, E. J., & Albrecht, J. E. (1992). Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology-Learning Memory and Cognition*, 18, 777–784.
- O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., & Halleran, J. G. (1998). Updating a situation model: A memory-based text processing view. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 24, 1200–1210.
- Poynor, D. V., & Morris, R. K. (2003). Inferred goals in narratives: Evidence from self-paced reading, recall, and eye movements. *Journal of Experimental Psychology-Learning Memory and Cognition*, 29, 3–9.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Rayner, K., Kambe, G., & Duffy, S. (2000). The effect of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology A*, 53(4), 1061–1080.
- Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line comprehension processes. *Language and Cognitive Processes*, 4, SI 21–SI 49.
- Rinck, M., Gamez, E., Diaz, J. M., & De Vega, M. (2003). Processing of temporal information: Evidence from eye movements. *Memory & Cognition*, 31, 77–86.
- Rinck, M., Hahnel, A., & Becker, G. (2001). Using temporal information to construct, update, and retrieve situation models of narratives. *Journal of Experimental Psychology-Learning Memory and Cognition*, 27, 67–80.
- Rubman, C. N., & Waters, H. S. (2000). A, B seeing: The role of constructive processes in children's comprehension monitoring. *Journal of Educational Psychology*, 92, 503–514.
- Schmauder, A. R. (1991). Argument structure frames: A lexical complexity metric? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 17, 49–65.
- Seigneuric, A., Ehrlich, M. F., Oakhill, J. V., & Yuill, N. M. (2000). Working memory resources and children's reading comprehension. *Reading and Writing: An Interdisciplinary Journal*, 13, 81–103.
- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, 125, 4–27.
- Skarakis-Doyle, E. (2002). Young children's detection of violations in familiar stories and emerging comprehension monitoring. *Discourse Processes*, 33, 175–197.
- Staphorsius, G., & Krom, R. (1998). *Toetsen begrijpend lezen [Reading comprehension tests]*. Arnhem, the Netherlands: CITO.

- Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Swanson, H. L. (1994). Short term memory and working memory—Do both contribute to our understanding of academic achievement in children and adults with learning disabilities. *Journal of Learning Disabilities*, 27, 34–50.
- 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.
- van der Schoot, M., Bakker Arkema, A. H., Horsley, T. M., & Van Lieshout, E. C. D. M. (2009a). The consistency effect depends on markedness in less successful but not successful problem solvers: An eye fixation study in primary school children. *Contemporary Educational Psychology*, 34, 58–66.
- van der Schoot, M., Horsley, T. M., & Van Lieshout, E. C. D. M. (2010). The effects of instruction on situation model construction: An eye fixation study on text comprehension in primary school children. *Educational Psychology*, 30, 817–835.
- van der Schoot, M., Vasbinder, A. L., Horsley, T. M., Reijntjes, A., & Van Lieshout, E. C. D. M. (2009b). Lexical ambiguity resolution in good and poor comprehenders: An eye fixation and self-paced reading study in primary school children. *Journal of Educational Psychology*, 101, 21–36.
- 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 Haasen, P. P. (1986). *Wechsler intelligence scale for children—Revised: Nederlandstalige uitgave [Dutch version]*. Lisse, the Netherlands: Swets & Zeitlinger.
- Vosniadou, S., Pearson, P. D., & Rogers, T. (1988). What causes children's failures to detect inconsistencies in text—Representation versus comparison difficulties. *Journal of Educational Psychology*, 80, 27–39.
- Wagoner, S. A. (1983). Comprehension monitoring—What it is and what we know about it. *Reading Research Quarterly*, 18, 328–346.
- Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension*. Cambridge, England: Cambridge University Press.
- Zabucky, K., & Moore, D. (1989). Children's ability to use 3 standards to evaluate their comprehension of text. *Reading Research Quarterly*, 24, 336–352.
- Zabucky, K., & Ratner, H. H. (1986). Children's comprehension monitoring and recall of inconsistent stories. *Child Development*, 57, 1401–1418.
- Zinar, S. (2000). The relative contributions of word identification skill and comprehension-monitoring behavior to reading comprehension ability. *Contemporary Educational Psychology*, 25, 363–377.
- Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995a). 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. (1995b). 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.