



# Children’s surface, textbase, and situation model representations of written and illustrated written narrative text

Benedikt T. Seger<sup>1</sup> · Wienke Wannagat<sup>1</sup> · Gerhild Nieding<sup>1</sup>

Accepted: 21 December 2020 / Published online: 17 January 2021  
© The Author(s) 2021

## Abstract

According to the tripartite model of text representation (van Dijk & Kintsch, 1983), readers form representations of the text surface and textbase, and construct a situation model. In this study, an experiment was conducted to investigate whether these levels of representation would be affected by adding illustrations to narrative text and whether the order of text and illustrations would make a difference. Students aged between 7 and 13 years ( $N = 146$ ) read 12 narrative texts, 4 of them with illustrations presented before their corresponding sentences, 4 with illustrations presented after, and 4 without any illustration. A sentence recognition task was used to assess the accuracy for text surface, textbase, and situation model. For the text surface and situation model, neither the presence of illustrations nor the order of text and illustrations influenced accuracy. However, the textbase was negatively affected by illustrations when they followed their corresponding sentences. We suggest that illustrations can initiate model inspection after situation model construction (Schnotz, 2014), a process that can make substantial changes to the textbase representation.

**Keywords** Children · Narrative text · Picture comprehension · Text comprehension

## Introduction

Generations of children have been exposed to illustrated storybooks, with tales read aloud by the children’s caregivers. To date, much research has been conducted showing a functional link between reading from storybooks and children’s language comprehension and literacy development (e.g., Duursma, Augustyn, & Zuckerman, 2008; Isbell, Sobol, Lindauer, & Lowrance, 2004; Klein & Kogan, 2013).

---

✉ Benedikt T. Seger  
benedikt.seger@uni-wuerzburg.de

<sup>1</sup> Department of Psychology, Developmental Psychology, Julius-Maximilians-Universität Würzburg, Röntgenring 10, 97070 Würzburg, Germany

Illustrations in storybooks appear to play a crucial role during the activity of reading aloud, and young children have been supposed as relying heavily on the information conveyed by the illustrations during story retelling (Isbell et al., 2004). Books and novels for older, literate children also often include illustrations, albeit to a lesser extent than storybooks for younger children. Certainly, these illustrations have an ornamental function, but it is still worth investigating whether and how they may contribute to understanding narrative content during silent reading.

Several experiments reveal that children recall narrative text better and generate more appropriate inferences when the verbal text is accompanied by appropriate illustrations (e.g., Beagles-Roos & Gat, 1983; Beentjes & van der Voort, 1991; Gambrell & Jawitz, 1993; Gibbons, Anderson, Smith, Field, & Fischer, 1986; Greenhoot & Semb, 2008; Guttman, Levin, & Pressley, 1977; Hayes, Kelly, & Mandel, 1986; O'Keefe & Solman, 1987; Pike, Barnes, & Barron, 2010; Ricci & Beal, 2002; Salomon & Leigh, 1984; for a review see Pressley, 1977). In some studies, differences between verbal-and-visual and verbal-only text are more pronounced in younger than in older children (Gibbons et al., 1986; Guttman et al., 1977; Pike et al., 2010). The research goal of the present study is to specify how illustrations are related to both superficial and deeper comprehension levels of written narrative text. To this end, we refer to a theoretical account that provides three levels of text representation and to models of multimedia learning that use this theory to explain the comprehension of both unillustrated and illustrated narrative text.

We use *text* as an umbrella term for every presentation modality (written, auditory, and audiovisual) and genre (narrative and expository).<sup>1</sup> If applicable, *text* refers to the combination of words (verbal text) and pictures. We define *stories* as coherent units of verbal narrative text of any length. The term *picture* encompasses any nonverbal, visual text elements that can have different functions in connection with verbal text (e.g., schematic representation, metaphor, additional information, illustration). The term *illustration* is exclusively used in the context of narrative text and refers to pictures that repeat what happens in the story. Accordingly, illustrations do not use information that is relevant to understand the situation; however, they may contain details of the scene that are not specified verbally.

### Text surface, textbase, and situation model

The tripartite model of text comprehension (van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998) holds that text recipients form three different mental representations of verbal text: text surface, textbase, and situation model. The *text surface* refers to the exact wording, whereas the *textbase* covers the semantic content that can be seen as a network of propositions (Kintsch, 1988, 1998). Propositions are the smallest meaning units to which a truth value can be assigned and are usually

---

<sup>1</sup> At this point, it is worth clarifying that none of the research referred to in the present study systematically uses text that can be assigned to the linguistic concept of *spoken language*, which has, among other features, a different communicative function than written text. Auditory verbal text is usually conceived as a spoken version of written text.

outlined using predicate-argument structures (e.g., Engelkamp, 1980). A sentence such as “Jane is watering the flowers in the garden” may be expressed as WATER (agent: Jane; object: flowers; location: garden). If the sentence is framed in the passive voice, like “The flowers in the garden are being watered by Jane,” the textbase remains identical, while the text surface is different.

The *situation model* is a coherent representation of the situation referred to in the text and is constructed by drawing inferences. For example, if one reads the sentence mentioned above, one may infer that Jane feels responsible for the flowers or that it has not rained for several days. Embodied cognition accounts (e.g., Barsalou, 1999; Zwaan, 1999, 2014) further suggest that situation models may contain analogous, multidimensional, and modality-specific *simulations* of real-world events. While reading the sentence “Jane is watering the flowers in the garden,” one may easily imagine seeing the flowers’ colors, smelling their fragrance, or hearing water pouring out of the watering can. Such simulations are supposed to be largely based on the recipient’s perceptual and motor experience (Glenberg & Robertson, 2000; Stanfield & Zwaan, 2001; Taylor & Zwaan, 2009). There are a considerable number of empirical findings confirming that text recipients simulate features of the situation through their perceptual and motor systems (e.g., de Koning, Wassenburg, Bos, & van der Schoot, 2017; Engelen, Bouwmeester, de Bruin, & Zwaan, 2011; Glenberg & Kaschak, 2002; Seger, Hauf, & Nieding, 2020; Zwaan, Stanfield, & Yaxley, 2002; Zwaan & Taylor, 2006). In Zwaan et al.’s (2002) study, for example, participants read a sentence (e.g., “The ranger saw the eagle in the sky”) and had to decide whether a subsequent picture referred to an object that was included in that sentence. Pictures that matched the participant’s situation model (e.g., an eagle with spread wings) were associated with shorter response times than pictures that did not match (e.g., an eagle with folded wings). Arguably, a merely linguistic representation (“eagle”) would be insufficient to explain this effect for which the embodied cognition hypothesis does account.

## Sentence recognition method

A sentence recognition method has been developed to establish all three representations of verbal text at once (Fletcher & Chrysler, 1990; Schmalhofer & Glavanov, 1986). The above-cited researchers found that surface, textbase, and situation model representations occurred simultaneously among adults. The participants were able to discriminate between an original sentence and a *paraphrase*, where the exact wording, but not the propositional structure, had changed. They were observed to discriminate better between paraphrases and *meaning changes*, where the propositional structure was also altered while remaining true to the situation (e.g., “Jane is watering the flowers *outside*”). However, discrimination was best when a *situation change* was presented in a sentence that was also incompatible with the recipient’s situation model (e.g., “Jane is watering the flowers *on the balcony*”). Nieding (2006) replicated this pattern of results in a sample of 5–11 year-old children, so there is

evidence that the tripartite model appropriately describes text comprehension in childhood.

Based on the above, we examined whether illustrations would make a difference in elementary school students' comprehension of auditory narrative text (Seger, Wannagat, & Nieding, 2019; Wannagat, Waizenegger, & Nieding, 2017; Wannagat, Waizenegger, Hauf, & Nieding, 2018). Wannagat et al. (2018) asked their 7, 9 and 11 year-old participants to listen to stories that comprised six sentences each before they completed a sentence recognition task. This task included original sentences, paraphrases, meaning changes, and situation changes. In one experimental condition, the participants received these stories in an auditory-only version; in the other condition, every sentence was accompanied with a static illustration. Similarly, Seger et al. (2019) scrutinized text surface, textbase, and situation model representations of auditory and audiovisual stories in the same group and with roughly the same stimulus material, with the exception that they added a third experimental condition that used animated rather than static illustrations.

In both studies, the situation model was significantly improved when illustrations were present rather than absent; likewise, text surface representations appeared to benefit from illustrations. One study (Wannagat et al., 2018) revealed an opposite pattern of results at the textbase level, indicating that semantic representations of text are less accurate when the text is illustrated; this was not replicated in Seger et al. (2019) research. In the latter study, dynamic illustrations produced similar results as static ones when accompanying auditory narrative text. To our knowledge, the effect of illustrations on the comprehension of written narrative text has not yet been investigated with reference to the tripartite model.

## Theories of text-and-picture learning

Based on a large body of research on expository text comprehension, Mayer (1997, 2009) formulated the *multimedia principle*, which holds that people learn better from verbal text accompanied by pictures than from verbal text alone. In this research tradition, all media that present words and pictures are referred to as multimedia, and multimedia learning is defined as building mental representations from words and pictures. In her review, Butcher (2014) showed that the multimedia principle is applicable to a variety of learning forms, including both superficial and deep levels of learning, and to a variety of media types.

Comparisons between expository text with and without pictures were shown to favor the multimedia principle, especially with regard to deep level learning. Glenberg and Langston (1992), for example, found that mental models based on written expository text improved when corresponding pictures were provided. Similar effects were obtained in a training study with hypermedia (Cuevas, Fiore, & Oser, 2002); the participants performed better in an integrative knowledge task—but not in a declarative knowledge task—when pictures were included in the hypermedia. The pictures in both studies were schematic diagrams that organized the information provided by the text without containing additional information. Butcher (2006) additionally varied between simplified (conceptually true) and complex (physically

true) diagrams. Her results suggested that pictures improve the mental modeling of expository text and that simple diagrams do more so than complex ones. The latter effect is explained by the notion that pictures have a beneficial effect on mental modeling when they can highlight essential information by providing a visual summary. In addition, participants in the simple diagram condition outperformed those in other conditions regarding memory of details.

The integrated model of text and picture comprehension (ITPC; Schnotz, 2014; Schnotz & Bannert, 2003) uses van Dijk and Kintsch's tripartite model to explain the multimedia principle. It assumes that processing text-picture units involves two channels: (1) a descriptive one proceeding from verbal text and (2) a depictive one proceeding from pictures. Accordingly, the text surface representation arises from sub-semantic processing, and the textbase representation emerges from semantic processing on the descriptive path. In contrast, the situation model is a depictive representation of the text and can be acquired in two ways: The first is *situation model construction* (van Dijk & Kintsch, 1983), which is based on semantic information gathered from descriptive processing (textbase) and one's own knowledge of the world. The second is *analog structure mapping* (Gentner, 1989), which is based on a picture surface representation directly gathered on the depictive path. If the picture reproduces central features of its corresponding verbal text (this includes *illustrations*, according to our definition), analog structure mapping can be used to match a constructed situation model with the picture surface representation because they, as depictive representations, share structural properties.

Analog structure mapping can explain why situation models improve when audiovisual text rather than auditory-only text is presented (e.g., Seger et al., 2019). It can also be argued that analog structure mapping reduces the need for semantic processing, which may result in lower textbase representations in the presence of pictures (Wannagat et al., 2018). Moreover, Schnotz and Bannert (2003) proposed that text recipients can apply *model inspection* processes after they have constructed a situation model. In doing so, they obtain new information from the situation model and encode this information in a propositional format. Such new information can have its origin in an illustration of verbal text. As a consequence, pictorial information may be encoded into propositions via model inspection, so illustrations may interfere with textbase representations.

## Impact of pictures on the comprehension of written text

In the domain of narrative text, there is empirical evidence that illustrations support the comprehension of written stories. Gambrell and Jawitz (1993) examined the recall of four-page stories with and without illustrations in a sample of 10 year-old children. Participants who read illustrated stories outperformed those reading verbal-only stories in both free and probed recall measures. Similar results were obtained by O'Keefe and Solman (1987) using stories comprising about 470 words (approximately one typed A4 page). Recall accuracy was higher when the presentation of the story and illustrations was sequential (experiments 1 and 2) or simultaneous

(experiment 3). According to Pike et al. (2010), readers aged between 7 and 10 years draw more correct inferences in short narrative texts with five sentences each that include an illustration than in those that are not illustrated.

Whereas the multimedia principle is insensitive to the modality (auditory or written) in which verbal text is presented, the *modality principle* (Low & Sweller, 2014; Moreno & Mayer, 1999; Mousavi, Low, & Sweller, 1995) claims that multimedia learning benefits more from a text that employs two sensory channels (auditory–visual) rather than one (visual–visual). A somewhat intuitive explanation of the modality principle would be that audiovisual text can be simultaneously encoded on two sensory channels, whereas the early visual processing of written text and pictures has to be successive, which can create a bottleneck. However, there has been a debate about precisely where this bottleneck occurs. The *split-attention effect*, for instance, claims that the bottleneck concerns attentional focus and can thus be overcome by spatially integrating written text and pictures (e.g., using diagram labeling; Ayres & Sweller, 2014). Alternatively, Rummer, Schweppe, Fürstenberg, Scheiter, and Zindler (2011), Rummer, Schweppe, Fürstenberg, Seufert, and Brünken (2010) introduced a *sensory register* hypothesis (see also Penney, 1989) claiming that a pre-attentive integration of verbal text and picture would be easier with auditory than with written text (for a critical discussion, see Reinwein, 2012). Ascribing the visual–visual bottleneck to early sensory processing would also be in line with the ITPC (Schnotz, 2014). Accordingly, the sub-semantic but not the semantic processing stage could be affected by such a bottleneck because phonetic decoding of written text has taken place before.

Nonetheless, it can be helpful to examine the possible effects of the processing order when investigating the comprehension of illustrated written text, and experimentally varying the presentation order of text and pictures is a plausible investigative technique. In the field of expository text comprehension, Eitel and Scheiter (2015) conducted a systematic review of studies that used this variation. They reported that the number of findings indicating that comprehension was better when the text preceded the picture (e.g., Canham & Hegarty, 2010) was almost equal to the number of findings that revealed the opposite pattern (e.g., Baggett, 1984; Eitel, Scheiter, Schüler, Nyström, & Holmqvist, 2013). As far as we know, in the domain of narrative text, only one attempt has been made to directly assess whether the order of text and pictures affects comprehension. A combined analysis of the first two experiments in O’Keefe and Solman’s (1987) study indicated that illustrations presented before or after their corresponding story improved recall compared with verbal-only text. However, no difference was found as to the order of verbal text and illustrations.

## This study

The aim of the present study is to understand how illustrations affect children’s comprehension of written stories and to examine whether the processing order of verbal text and illustrations makes a difference in that regard. More specifically, we investigated how each level of representation according to the tripartite model (text surface,

textbase, and situation model) would be affected (Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). To obtain separate measures for each level, we employed a sentence recognition task similar to the one introduced by Schmalhofer and Glavanov (1986) and used in several later experiments (Fletcher & Chrysler, 1990; Nieding, 2006; Seger et al., 2019; Wannagat et al., 2017, 2018). The stories in our study reflected possible daily-life situations of school children in Western countries. We varied three story versions experimentally: written stories without illustrations (*sentence-only*, *SO*), written stories with illustrations presented beforehand (*picture-sentence*, *PS*), and written stories with illustrations presented afterward (*sentence-picture*, *SP*). Another purpose of our study was to examine whether beginning readers (age 7) would differ from more advanced readers (up to age 13) in their comprehension of written narrative text with and without illustrations. Finally, we studied the effects of illustrations and the text-illustration order on reading times.

We anticipated that the *situation model* would benefit from illustrations in general, consistent with multimedia learning theories (Mayer, 2009) and earlier results from both auditory (Beagles-Roos & Gat, 1983; Gunter, Furnham, & Griffiths, 2000; Hayes et al., 1986; Seger et al., 2019; Wannagat et al., 2018) and written narrative text (Gambrell & Jawitz, 1993; Pike et al., 2010). We also assumed that situation model representations would be more accurate in the PS than in the SP condition. This would be in line with the ITPC (Schnotz & Bannert, 2003), which holds that an appropriate situation model could be directly obtained via analog structure mapping, which serves as a scaffold for the subsequent, more complex process of situation model construction based on verbal text. Thus, Hypothesis 1 predicted the order of accuracy for the situation model to be  $PS > SP > SO$ .

Regarding the *textbase*, we expected illustrations to have a negative effect. We derived this assumption from the ITPC. If the situation model could be directly obtained from a picture surface representation, semantic processing might become less relevant to this objective and might therefore be neglected. This effect was found in one of our earlier studies with auditory stories (Wannagat et al., 2018), but not in others (Seger et al., 2019; Wannagat et al., 2017). In addition, new information obtained from an illustration could alter textbase representations via model inspection (Schnotz & Bannert, 2003). As model inspection is presumed to take place after model construction, we thought that this effect would be more likely when the illustration was presented after the sentence rather than before. Therefore, Hypothesis 2 predicted that accuracy would be lower when illustrations were present rather than absent and that accuracy in SP would be lower than that in PS (i.e.,  $SO > PS > SP$  for the textbase).

For the *text surface*, we hypothesized that illustrations would have a positive effect, consistent with our earlier results with auditory versus audiovisual text (Seger et al., 2019; Wannagat et al., 2018). However, we made no assumption regarding the order of text and illustrations (Hypothesis 3:  $SP = PS > SO$ ). Hypothesis 4 predicted that illustrations would facilitate subsequent reading, which would be reflected in lower reading times when illustrations were present in general and when they were presented before the written text in particular ( $PS < SP < SO$  for reading time).



## Method

### Participants

We determined that an ideal sample size of  $N=144$  would enable an optimal balance across participants and conditions (see below for more details). A power analysis conducted via G\*Power (Version 3.1.9.2, Faul, Erdfelder, Lang, & Buchner, 2007) indicated that with this sample size, a true effect size of  $\eta^2=.020$  would be detected with a likelihood of more than 90% (i.e.,  $\beta < .10$ ). This effect size is remarkably below the effect sizes associated with the significant results obtained in earlier sentence recognition studies, which ranged between  $\eta^2=.040$  and  $\eta^2=.092$  (Seger et al., 2019, Wannagat et al., 2017, 2018).

In total, 146 students aged between 7.75 and 13 years (mean age=10.42,  $SD=1.25$ , median=10.58) participated in our study, with females comprising a slight majority (53%). The participants were recruited from several elementary schools and a comprehensive secondary school in Germany. All participants spoke German at the native-speaker level. The students only participated after their parents had signed a consent form.

### Sentence recognition task

We used a three-level sentence recognition task based on the method introduced by Schmalhofer and Glavanov (1986). Our task is an adapted version of the one used in earlier studies with children (Seger et al., 2019; Wannagat et al., 2017, 2018). The participants read stories composed of six sentences each. After a block of four stories, they read single sentences and were required to decide whether each was part of the story. The sentences were either presented in their original wording, requiring a positive answer, or were modified in one of three ways: as a *paraphrase*, where the wording (i.e., text surface) was changed without changing the meaning at the sentence level (e.g., by replacing one or more expressions with synonyms); as a *meaning change*, where the meaning at the sentence level (i.e., textbase) was altered but remained true to the story plot; or as a *situation change*, where the meaning of a sentence was modified in a way that was incompatible with the plot (i.e., meant to contradict the reader's situation model).

The task included 12 stories related to everyday events that might occur in a child's life in Western societies, so domain-specific knowledge or expertise would not be necessary (see Table 1 for an example). Text coherence was ensured locally by employing theme–rheme structures (e.g., pronouns that unambiguously refer to a character or object occurring in the previous sentence) and globally by providing an appropriate title in advance (Bransford & Johnson, 1972) and in capital letters.<sup>2</sup> A vast majority (91.7%) of the sentences described one or more characters' actions; some sentences (31.9%) referred to a character's emotional state. For each original

<sup>2</sup> We employed theme–rheme structures and titles to the best of our knowledge and belief, but we did not empirically test their adequacy. We also ensured that the titles did not interfere with any of the distractor sentences.



**Table 1** Sample story entitled *Beim Essen* (at lunch) and its illustrations



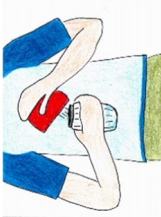


No.	Original sentence in German	English translation	Illustration
1	<i>Die Mutter ruft nach Max und Sascha, doch eigentlich wollen die beiden gerade ihre Lieblingsserie anschauen</i>	Mother calls out for Max and Sascha, but the two of them just wanted to watch their favorite serial	
2	<i>Die Oma kommt zum Mittagessen vorbei und Max und Sascha sollen deshalb den Tisch decken</i>	Grandma comes around for lunch, therefore, Max and Sascha are supposed to set the table	
3	<i>Als sie widerwillig Teller und Gläser aus der Küche herbeitragen, sieht Max auf einmal eine rote Zuckerdose auf dem Schränkchen stehen</i>	As they unwillingly fetch up plates and glasses from the kitchen, Max suddenly sees a red sugar bowl on the sideboard	
4	<i>Grinsend zwinkert er Sascha zu und schüttet dann den Zucker aus der roten Dose in den Salzstreuer hinein</i>	Smirkingly, he twinkles to Sascha and then pours the sugar from the red bowl into the salt shaker	

Table 1 (continued)

No.	Original sentence in German	English translation	Illustration
5	<i>Bald sitzen alle am Tisch und Max und Sascha lassen den Salzstreuer nicht aus den Augen</i>	Pretty soon, everyone sits at table and Max and Sascha keep their eyes glued to the salt shaker	
6	<i>Als der Vater dann endlich nach dem Salzstreuer greift, können die beiden ihr Lachen nicht mehr unterdrücken</i>	As Father finally reaches for the salt shaker, the two of them can no longer keep under their laughter	

sentence, three distractors were created that met the criterion of paraphrase, meaning change, and situation change, respectively (see Table 2 for an example). Sentence length varied between 10 and 22 words (mean = 15.23,  $SD = 2.47$ , median = 15) with negligible differences between sentence types. In the two illustrated conditions, one static illustration preceded or followed every sentence. Most depicted at least one character (90.3%) or an action or emotional state (87.5%) to which the corresponding sentence referred. We ensured that the illustrations did not include any detail that might be incompatible with the distractor sentences, especially the situation change versions.

During the task, six probe sentences were presented in scrambled order: three as original sentences, one as a paraphrase, one as a meaning change, and one as a situation change. The probe sentences were balanced as much as possible in two ways. First, we ensured that for each of the 72 sentences, every sentence type appeared equally often among all participants and in each condition. That is, each sentence appeared equally often in its paraphrase, meaning change, and situation change versions, and each sentence appeared in the original version as frequently as in all changed versions combined. Second, we ensured that the position of each sentence in the task was equally distributed. For example, the first sentence of a given story was equally often the first, third, or last sentence in its related task.

The verbal text was presented in black Arial font in the top third of a white 800 \* 600-pixel field; the font size was 20 points for sentences and 26 points for titles. The illustrations were hand-drawn and colored (see Table 1), with a uniform size of 800 \* 600 pixels. The experiment was implemented using DMDX<sup>®</sup> software, Version 5 (Forster & Forster, 2016) on a laptop computer, with a resolution of 1280 \* 720 and frame rate of 60 Hz.

## Design and procedure

Three experimental conditions were varied within participants: one sentence-only (SO), one with illustrations presented before their corresponding sentences (PS) and one with illustrations presented after (SP). The participants read the 12 stories in 3 blocks of 4 stories each, where each block represented a single condition. All possible orders of experimental conditions were permuted and randomly assigned to the participants; however, we tried to balance them in terms of age, gender, and time of day (class hours) as far as possible.

For the experimental task, the students were instructed to read the stories and remember them as accurately as possible. Concerning the sentence recognition task, they were instructed to expect a test on which they would be presented with sentences in arbitrary order and would have to decide whether these sentences had appeared in one of the stories. For “yes,” they pressed the “3” key on the numeric keypad, which was stickered with a happy emoticon, for “no,” they pressed the “1” key, which was stickered with a sad emoticon. They completed a practice trial comprising three sentences and three probes in the following order: situation change, original, and paraphrase. We provided no feedback at any time. However, after the practice trial, we asked the participants whether they had understood how to perform

**Table 2** Original sentences, paraphrases, meaning changes, and situation changes of the third sentence from the story Beim Essen (at lunch)

Sentence type	German version used in the study	English translation
Original	Als sie widerwillig Teller und Gläser aus der Küche herbeitragen, sieht Max auf einmal eine rote Zuckerdose auf dem Schränkchen stehen	As they unwillingly fetch up plates and glasses from the kitchen, Max suddenly sees a red sugar bowl on the sideboard
Paraphrase	Als sie widerstrebend Teller und Gläser aus der Küche herbeibringen, sieht Max auf einmal eine rote Zuckerdose auf dem Schränkchen stehen	As they <i>reluctantly</i> bring up plates and glasses from the kitchen, Max suddenly sees a red sugar bowl on the sideboard
Meaning change	Als sie widerwillig <i>Geschirr</i> aus der Küche herbeitragen, sieht Max auf einmal eine rote Zuckerdose auf dem Schränkchen stehen	As they unwillingly fetch up <i>dishes</i> from the kitchen, Max suddenly sees a red sugar bowl on the sideboard
Situation change	Als sie <i>eifrig</i> Teller und Gläser aus der Küche herbeitragen, sieht Max auf einmal eine rote Zuckerdose auf dem Schränkchen stehen	As they <i>eagerly</i> fetch up plates and glasses from the kitchen, Max suddenly sees a red sugar bowl on the sideboard

the task. We also repeated the instructions if the response pattern in the practice trial raised the issue that the participants might not have correctly understood them (e.g., if they considered the order of sentences during the task). During reading, the participants always proceeded by pressing the “Enter” key (stickered with a book symbol) for the next sentence or picture to appear; thus, reading and picture viewing were self-paced, without an imposed time limit. The reading and picture-viewing times were automatically measured by the experimental software. The task phase also had no time limit except for the titles, each of which was shown for three seconds and served as a reminder for the story. No pictures were shown during the task phase. When a reading block was completed (after four stories), a short instruction in read text announced the task phase. After the task, a short instruction in green text announced the next or last block or the end of the experiment. The entire experiment usually required 25–40 min.

## Data analysis

We calculated the *acceptance rates* (i.e., the relative frequencies of “yes” responses) for originals, paraphrases, meaning changes, and situation changes to determine whether the tripartite model would be appropriate to describe text comprehension in our study. We considered this to be the case if the acceptance rates were the highest for originals and decreased with increasing change intensity.

For each level of representation, sensitivities based on the signal detection theory (Stanislaw & Todorov, 1999) were computed. We deemed this necessary because the acceptance rates of a certain change type do not unambiguously refer to the respective level of representation. For instance, accepting a situation change as being part of the story indicates that the reader had not constructed an appropriate situation model; however, rejecting a situation change can also indicate that a reader merely had a correct representation of the text surface or textbase, as situation changes necessarily imply meaning changes and meaning changes necessarily imply paraphrases. Moreover, sensitivity measures have the advantage of being independent of recipient response bias (Stanislaw & Todorov, 1999).

We used the nonparametric  $A'$  sensitivity measure that does not require normally distributed values (Donaldson, 1992) and ranges from 0 to 1, with 0.5 representing the chance level. For text surface  $A'$ , “yes” responses to originals were categorized as hits and “yes” responses to paraphrases were categorized as false alarms (i.e., false positives). For textbase  $A'$ , “yes” responses to originals and paraphrases were considered hits and “yes” responses to meaning changes were considered false alarms. Finally, for the situation model, “yes” responses to originals, paraphrases, and meaning changes were regarded as hits and “yes” responses to situation changes were regarded as false alarms. In general, we assigned the acceptance of a specific change type to false alarms, indicating that the subject had no adequate text representation at the corresponding level; moreover, we designated the combined acceptance rates at the more superficial levels as hits (see also Seger et al., 2019). For detailed formulas, see Table 3. Please note that  $A'$  cannot be expressed as a real number if the hit rate is zero or

**Table 3** Formulas for the nonparametric signal detection sensitivity measures ( $A'$ 's) used in our study

General	$A' = 0.5 + \frac{(H-F)(1+H-F)}{4H(1-F)}$
Surface	$A' = 0.5 + \frac{(Y_O - Y_P)(1 + Y_O - Y_P)}{4Y_O(1 - Y_P)}$
Textbase	$A' = 0.5 + \frac{(Y_{O,P} - Y_M)(1 + Y_{O,P} - Y_M)}{4Y_{O,P}(1 - Y_M)}$
Situation model	$A' = 0.5 + \frac{(Y_{O,P,M} - Y_S)(1 + Y_{O,P,M} - Y_S)}{4Y_{O,P,M}(1 - Y_S)}$

$H$  hit rate,  $F$  false alarm rate,  $Y$  acceptance rate,  $O$  original sentence,  $P$  paraphrase,  $M$  meaning change,  $S$  situation change

**Table 4** Acceptance rates per sentence type, mean reading and picture-viewing times

	Acceptance rate (relative frequency of “yes” responses) per sentence type ( $N=146$ )				Mean reading/viewing time in milliseconds ( $N=146$ )	
	Original	Paraphrase	Meaning change	Situation change	Text (without titles)	Illustrations
Mean	0.862	0.744	0.602	0.287	8685.18	2085.48
Median	0.889	0.750	0.583	0.250	8147.25	2025.54
$SD$	0.103	0.167	0.179	0.178	3381.01	860.19
Minimum	0.444	0.333	0.167	0.000	3408.42	870.05
Maximum	1.000	1.000	1.000	0.750	25,813.99	4825.05

the false alarm rate is one. If such a case occurred in at least one experimental condition, the participants were excluded from hypothesis testing at the corresponding text comprehension level. This applies to 37 participants (25.3%) for text surface analysis, 7 participants (4.8%) for textbase analysis, and a single participant (0.7%) for situation model analysis.

## Results

### Preliminary analyses

The mean acceptance rate was 0.862 ( $SD=0.103$ ) for originals, 0.744 ( $SD=0.167$ ) for paraphrases, 0.602 ( $SD=0.179$ ) for meaning changes, and 0.287 ( $SD=0.178$ ) for situation changes (see Table 4). A repeated-measure analysis of variance (ANOVA) exhibited a significant effect of the sentence type,  $F(3, 143) = 321.91, p < .001, \eta^2 = .871$ . Contrast analyses showed significant differences between originals and paraphrases,  $F(1, 145) = 72.39, p < .001, \eta^2 = .333$ , paraphrases and meaning changes,  $F(1, 145) = 70.47, p < .001, \eta^2 = .327$ , and meaning changes and situation changes,  $F(1, 145) = 358.982, p < .001, \eta^2 = .712$ . Thus, we assumed that the tripartite model was applicable

**Table 5** Mean sensitivity A's for surface, textbase, and situation model, and mean reading and picture viewing times (in milliseconds) dependent on experimental conditions

	Surface A ( <i>N</i> =109)		Textbase A ( <i>N</i> =139)		Situation model A ( <i>N</i> =145)		Mean reading time ( <i>N</i> =146)		Mean picture-viewing time ( <i>N</i> =146)	
	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>
Text-only	0.593	0.257	0.649	0.237	0.833	0.143	9098.31	4139.41	–	–
Text-picture	0.600	0.243	0.579	0.243	0.822	0.168	8517.44	3307.83	2161.33	1047.93
Picture-text	0.559	0.257	0.681	0.223	0.823	0.153	8431.00	3430.53	2010.00	867.28

to the sentence recognition task in our sample. The internal consistency for the acceptance rate of originals was in the acceptable range (Cronbach's  $\alpha = .708$ ), but this was not the case for paraphrases ( $\alpha = .476$ ), meaning changes ( $\alpha = .423$ ), or situation changes ( $\alpha = .512$ ).

Descriptive statistics, including reading and picture-viewing times, are shown in Table 4. Not surprisingly, reading times were negatively correlated with age ( $r = -.322$ ,  $p < .001$ ). Sensitivity measures and acceptance rates did not correlate with reading or picture viewing times ( $|r| \leq .146$ ,  $p \geq .079$ ), indicating that there was no speed-accuracy tradeoff in our data. Sensitivity measures and acceptance rates were also unrelated to age.

### Levels of representation

Because sensitivity measures showed no correlation with age, we excluded it from the analyses for the levels of representation. We did not calculate a multivariate ANOVA that would allow for direct comparisons between levels of representation owing to the statistical interdependencies between the sensitivity measures. Thus, repeated-measure ANOVAs with the text format as predictor were separately performed for the text surface, textbase, and situation model sensitivities.

For the situation model, the effect of the text format was not significant,  $F(2, 143) = 0.272$ ,  $p = .763$ , which refutes our assumption that illustrations would enhance situation model representations of written narrative text (Hypothesis 1). However, a significant effect emerged at the textbase level,  $F(2, 137) = 7.958$ ,  $p = .001$ ,  $\eta^2 = .104$ . Planned contrasts revealed significantly higher accuracies in PS than in SP,  $F(1, 138) = 15.605$ ,  $p < .001$ ,  $\eta^2 = .102$ , whereas there was no significant difference between both illustrated conditions and the SO condition,  $F(1, 138) = 0.624$ ,  $p = .431$ . This partly supports Hypothesis 2; accuracy was significantly higher when the picture was presented before rather than after the sentence, but there was no general advantage of the SO condition over both illustrated conditions. Text surface A' was not affected by the text format,  $F(2, 107) = 1.084$ ,  $p = .342$ ; therefore, Hypothesis 3 is rejected. The descriptive statistics for sensitivities in dependence on experimental conditions are summarized in Table 5.



## Reading and picture viewing times

As reading time was significantly related to age, we ran an analysis of covariance (ANCOVA) to determine any possible interaction between experimental conditions and age. This interaction was not significant,  $F(2, 143)=1.371, p=.257$ ; therefore, we decided to perform an ANOVA instead. The effect of the text format on reading times was significant,  $F(2, 144)=5.562, p=.005, \eta^2=.072$ . Planned contrasts indicated shorter reading times in the illustrated conditions than in the SO condition,  $F(1, 145)=9.577, p=.002, \eta^2=.062$ , whereas the contrast between the PS and SP conditions did not reach significance,  $F(1, 145)=0.332, p=.565$ . These findings confirmed Hypothesis 4 insofar as reading times differed between the text-only and both illustrated conditions but not between PS and SP. Unexpectedly, illustrations were viewed longer in the SP than in the PS condition,  $t(145)=2.125, p=.035$ . For an overview of reading and picture-viewing times depending on text format, see Table 5.

## Analyses for carryover effects

Although the order of experimental conditions was balanced across participants, we were interested in any carryover effects that may have occurred between them. To this end, we re-ran our analyses of text surface, textbase, and situation model *A*'s with an additional between-participant factor indicating which text condition was completed first (SO vs. PS vs. SP). This factor yielded a significant main effect for the textbase,  $F(2, 136)=3.155, p=.046, \eta^2=.044$ ; however, Bonferroni-adjusted post hoc comparisons did not reveal significant group differences for this factor. More interestingly, a significant interaction was observed between this factor and the experimental factor for the textbase,  $F(4, 272)=3.257, p=.012, \eta^2=.046$ . Bonferroni-adjusted post hoc comparisons indicated significantly lower textbase *A*'s in the SP than in the SO condition (mean difference=0.140,  $p=.010$ ) and the PS condition (mean difference=0.204,  $p<.001$ ) in the group of participants who began with SO. Participants who started with PS yielded higher textbase *A*'s in the PS than in the SP condition (mean difference=0.122,  $p=.016$ ). Participants starting with SP did not display significant differences between the conditions.

For the situation model, this interaction was also significant,  $F(4, 284)=6.373, p<.001, \eta^2=.082$ . Bonferroni-adjusted post hoc comparisons suggested higher performance in SO than in PS (mean difference=0.081,  $p=.011$ ) for the participants who started with SO, whereas the opposite effect occurred in the group of participants starting with SP (mean difference=0.086,  $p=.005$ ). In the group starting with PS, there were no significant differences between conditions.

Importantly, the main effect of the experimental conditions was significant for the textbase,  $F(2, 274)=8.083, p<.001, \eta^2=.056$ , and no significant main effects of the experimental conditions were observed regarding the text surface,  $F(2, 212)=0.953, p=.387$ , or the situation model,  $F(2, 284)=0.363, p=.696$ . This suggests that the main results of our experiment were not affected by carryover effects.

## Discussion

The purpose of our study was to examine the effect of illustrations on text surface, textbase, and situation model representations (van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998) of written narrative text read by elementary and early secondary school children. The participants performed a sentence recognition task that allowed us to measure all three levels simultaneously (Fletcher & Chrysler, 1990; Nieding, 2006; Schmalhofer & Glavanov, 1986). The participants were forced to process verbal text and illustrations sequentially, so we were particularly interested in any possible effects of the processing order. Therefore, each participant was presented with three versions of the sentence recognition task: one with sentences presented alone (SO), one with sentences presented before their corresponding illustrations (SP), and one with the illustrations presented first (PS).

## Situation model

Our hypothesis that situation model representations would benefit from the presence of illustrations was not supported by the data. Therefore, the stable superiority of audiovisual text to auditory text with regard to the situation model (e.g., Seger et al., 2019; Wannagat et al., 2018) does not appear to pertain to illustrated compared unillustrated written text. This finding can be interpreted in the context of the modality principle (Low & Sweller, 2014), which holds that pictures have a greater beneficial impact on text comprehension if two sensory channels are involved instead of one.

Nevertheless, several studies have reported a positive effect of illustrations on the comprehension of written stories (Gambrell & Jawitz, 1993; O'Keefe & Solman, 1987; Pike et al., 2010). Three major differences between them and the study reported here must be noted. First, illustrations may be more beneficial when they appear together with the stories, as was the case in the studies of Gambrell and Jawitz (1993) and Pike et al. (2010). In O'Keefe and Solman's (1987) first two experiments, the advantage of stories with illustrations presented sequentially over stories without illustrations was smaller than the advantage of illustrations presented simultaneously with their corresponding verbal text. Situation model construction may benefit from features of concurrent text-picture units that are not shared by sequential ones. We cautiously assume that concurrent text-picture units provide the opportunity for the iterative processing of verbal text and pictures, which may lead to a more accurate representation of the state of affairs described.

Second, the stories used in all these studies included fewer illustrations than ours while having a comparable (Pike et al., 2010) or even larger (Gambrell & Jawitz, 1993; O'Keefe & Solman, 1987) number of words. This results in pronouncedly different picture-per-word rates (1:15 in our study, as opposed to 1:65 in Pike et al. and nearly 1:100 in the other two studies). If a single illustration refers to a portion of text larger than a hundred words, it is quite likely that the illustration would help the reader integrate the comparatively rich semantic information into a coherent situation model. By contrast, one illustration per sentence is supposed to have a more limited potential in that regard; moreover, illustrations that are presented in

alternation with sentences interrupt the flow of reading, which may have a detrimental effect on situation model construction. Therefore, we do not rule out the possibility that illustrations would enhance situation model construction if there were only one illustration per story rather than one per sentence.

Third, our sentence recognition task did not allow us to illustrate inferences that were incompatible with situation change distractors. By contrast, Gambrell and Jawitz (1993) and O'Keefe and Solman (1987) tested whether the total number of correctly recalled information units differed between the illustrated and text-only conditions. Neither approach examined the possibility that this difference might be limited to information units that were present in both verbal text and illustrations. The central result of Pike et al.'s (2010) study was that the generation of correct inferences was significantly enhanced when relevant features of the situation were shown. Therefore, it is possible that readers' situation models benefit from illustrations only with regard to the aspects displayed.

## Textbase

Our second hypothesis was that illustrations would impede textbase representations, especially when they were presented after written text. This was confirmed insofar as textbase sensitivities were lower in the SP condition than in the other two conditions. The model inspection process, which is part of the ITPC framework (Schnotz, 2014; Schnotz & Bannert, 2003), accounts for this result: readers construct a situation model based on verbal text information and then update this model based on visual information from the illustration. This is followed by model inspection, where readers encode the updated model in a propositional format, which allows them to verbalize the story plot from their own perspective. This means that illustrations presented after their corresponding verbal text can motivate readers to make substantial changes to their textbase representations.

Earlier results with auditory narrative text indicated an overall negative effect of illustrations on the textbase (Wannagat et al., 2018). The explanation was that obtaining a situation model on the depictive path of the ITPC (via analog structure mapping) would render the semantic processing of verbal text less relevant; therefore, the participants would generate a weaker text representation at the semantic level. If this were the case, we would expect lower sensitivities in both the SP and PS conditions than in the SO condition. Because textbase sensitivities were lower in SP, but not in PS, compared with SO, this explanation appears to be less suitable than the model inspection account described above. Therefore, we suggest that, on one hand, recipients form a textbase representation regardless of whether a text is illustrated. On the other, illustrations can initiate model inspection, leading to changes in the textbase representation, especially when the illustration is processed after its corresponding verbal text.

Different presentation modalities of verbal text may explain why participants appeared to neglect the textbase in Wannagat et al. (2018) study but not in this one. The auditory stories used by Wannagat et al. (2018) are recorded readings of written stories that do not resemble oral language, which means that textbase processing

might be less effective when written text is presented aurally rather than in its original written format. By consequence, illustrations may prompt listeners but not readers to apportion fewer mental resources to semantic processing and to favor analog structure mapping instead. It is notable, however, that the evidence of low textbase representations in audiovisual text lacks replication (Seger et al., 2019). To further examine this issue, future studies should include simultaneous units of written text and illustration and compare these to audiovisual text.

### **Text surface**

Unlike our prediction, there were no significant differences between written and illustrated written text formats with respect to text surface sensitivities; this contrasts with our earlier studies' findings that illustrations improved text surface representations of auditory text (Seger et al., 2019; Wannagat et al., 2018). Interestingly, these studies also reported a positive effect of illustrations on situation model construction. It may be that the memory of the exact wording profits from the same text features that facilitate situation model construction to the extent that the cognitive resources needed for the latter process can partly be spared when illustrations are present. This is in line with another finding reported in our 2019 study; accordingly, text surface representations were significantly improved when auditory text was furnished with static illustrations but not animated ones, whereas the situation model sensitivity was equally high in both conditions. We argued in 2019 that the animations demanded additional cognitive load that used up the resources left over from the situation model construction in both audiovisual text versions. In the study reported here, neither the situation model nor the text surface profited from the presence of illustrations. At this point, however, we should be aware of the danger of over-interpreting a single non-significant result. It may be worth gauging the linear relationship between text surface and situation model representations within the scope of a systematic review or meta-analysis.

### **Reading and picture-viewing times**

As expected, reading times were significantly shorter when illustrations were present rather than absent, corroborating the multimedia principle (Mayer, 2009), which holds in general that pictures facilitate reading. However, the specific version of this assumption, namely that illustrations would diminish the reading time of *subsequent text*, could not be affirmed here because there was no difference between the PS and SP conditions. One reason might be that illustrations help recipients anticipate the further course of events (i.e., they support predictive inferences; cf. Unsöld, & Nieding, 2009), which might constitute a reliable comprehension strategy for the commonplace stories in our study. In this case, whether the term "subsequent text" refers to the corresponding sentence (PS) or the following sentence (SP) would be of little relevance; both sentences might be more easily predicted in these conditions compared with the text-only condition, resulting in shorter reading times.

Alternatively, the participants might have been more confident about their task performance when illustrations were present and therefore spent less time reading. Although illustrations presented before or after written text do not appear to increase understanding, it is still possible that they increase an individual's illusion of understanding (e.g., Jaeger & Wiley, 2014; Serra & Dunlosky, 2010). Nonetheless, the total time spent on a sentence was, on average, more than a second longer in the illustrated conditions than in the SO condition (cf. Table 5). It can thus be stated that the ensemble of processes related to the situation model (i.e., model construction, model inspection, and analog structure mapping) in the two illustrated text versions was more time-consuming than the model construction process in the verbal-only version, without having a positive effect on situation model accuracy. We tentatively conclude that asynchronous units of written text and illustrations are inefficient media formats in the domain of narrative text (for scientific text, see research on the temporal contiguity principle, e.g., Mayer & Fiorella, 2014).

Our participants spent significantly more time on viewing the illustrations in the SP condition than in the PS condition. We did not expect this result, but we think that it can be ascribed to model inspection (Schnotz & Bannert, 2003), which may be more pronounced when the sentence has been processed before the illustration than vice versa. For example, imagine a participant reading, "Max pours the sugar from the red bowl into the salt shaker." If the participant has constructed a situation model that depicts Max with the sugar bowl in his right hand and the salt shaker in his left, the subsequent illustration may induce the participant to update this situation model (cf. Table 1) so that it depicts the sugar bowl in Max's left hand (perhaps together with the inference that Max may be left-handed). Thus, one may suppose that model inspection takes additional cognitive resources that are reflected in longer picture-viewing times. The embodied cognition account (e.g., Glenberg & Robertson, 2000; Taylor & Zwaan, 2009; Zwaan et al., 2002) may explain this result in a similar way: if the subsequent picture does not match a participant's perceptual and motor simulations (e.g., if he or she simulates the right hand holding the sugar bowl after reading the example sentence above), it may take longer for that participant to verify that picture (which shows the sugar bowl in the left hand).

It is noteworthy that the mean picture-viewing times vary remarkably across participants, ranging from just below one to almost five seconds (see Table 4). Exploring in detail how students use illustrations while looking at them and how far individual differences may play a role here could be informative. For example, one may imagine that those spending more time on illustrations try to create an appropriate context where the presented story may be embedded; indeed, such a strategy can support the construction of an appropriate situation model. In this sense, we encourage future research to explore more deeply what children do when exposed to illustrations of narrative text (Table 3).

### Limitations and further directions

One methodological drawback in the present study may originate from the instructions, which could have induced participants to learn sentences by rote and thus

focusing on the text surface instead of constructing a situation model, or what Kintsch (1998) calls a “real understanding.” In fact, our intention was that participants would not only focus on the situation model but also pay attention to the text surface and textbase, which were also within the scope of our research interest. In an earlier study (Seger et al., 2019), we followed a different approach; namely, providing a rather vague instruction of *remembering the text well* and asking afterward whether the participants employed a verbatim or plot-based memory strategy. As expected, those who indicated using a verbatim strategy outperformed those employing an exclusively plot-based strategy with regard to the text surface. However, there was no such effect concerning the situation model or textbase. We inferred from these earlier results that situation model construction unintentionally takes place during text reception at least when the text is close to the recipients' daily lives, whereas some conscious effort is required to have a more accurate memory of specific wording. In the study reported here, we thus decided to formulate an instruction that also prompted participants toward memorizing the text verbatim.

The within-participant design in this study increased the statistical power of the results (compared with a between-participant design of the same sample size) and controlled for individual differences concerning reading abilities, among others. One shortcoming of this design was that the participants read only four stories per condition and were therefore exposed to only four paraphrases, four meaning changes, and four situation changes per condition. Thus, false alarm rates of 100% were quite likely, leading to incalculable sensitivity measures with the consequence of serious drop-out rates, especially at the text surface level. Additional analyses, where missing data were replaced by 0 (no sensitivity) or 0.5 (random-level sensitivity), did not reveal significantly different result patterns, indicating that these drop-outs did not systematically bias our results. Furthermore, there were carryover effects between the presentation modalities in the course of the experiment; however, these appear to be unrelated to the main findings and thus do not constitute a serious threat to their internal validity.

The internal consistencies for acceptance rates for paraphrases, meaning changes, and situation changes are below the margin of acceptability. This means that the sensitivity measures for all three levels of representation are associated with considerable measurement errors that may limit the interpretability of our results, especially if these errors are systematic. These low reliability values may be attributable to the fact that there was only one paraphrase, one meaning change, and one situation change (as opposed to three original sentences) per participant and story. However, we deemed it important to have a 50% rate of correct acceptances (i.e., originals) to minimize the risk that participants did not respond significantly above chance level, with the consequence that each change type could occur only once in every six sentences.

We also acknowledge that the text-picture units in our work do not constitute a setting that represents typical narrative reading situations for 7–13 year-old children. First, the sequential presentation of illustrations and corresponding verbal text, especially without the opportunity to turn back to previous pages, is far from the reality of either printed or electronic books. Second, as discussed above, the picture-per-word rate of our stories was ten times higher than that employed by

O'Keefe and Solman (1987), who used real samples from fifth-grade literature. In our study, this rate is presumably closer to what would be usual for younger children's storybooks. Third, of course, children rarely read stories in expectation of a sentence recognition task; for example, reading in the school context more often requires free retelling or cued recall. However, our major research goal was related to the simultaneous examination of text surface, textbase, and situation model representations in a maximally distinct way, and the sentence recognition task introduced by Schmalhofer and Glavanov (1986) is a well-established method for this purpose. As different sentences within a story were assigned to different probe sentence types in the task, it was necessary to illustrate each of them. Another research goal was to determine whether the processing order of sentences and illustrations would have an impact on comprehension. An experimental variation of the presentation order is presumably the most effective way to do so.

Finally, a sequence of actions and utterances treated in as few as six sentences cannot easily be generalized to typical narratives in the literature of Grade 2 and higher, which exceed the length of our stories by far. Therefore, future research should make use of longer narrative texts that can also include processing themes, along with more words per picture. Eye tracking can also be a powerful tool not only to obtain reading and picture-viewing time data in text-picture units that are presented simultaneously but also to explore the sequence of reading and picture-viewing episodes. Both types of data can be related to outcomes relevant to situation model construction to gain a deeper understanding of the cognitive processes underlying the comprehension of illustrated narrative text. Further attempts to transfer our findings to more realistic reading situations should also investigate whether an iterative processing of verbal text and pictures (e.g., the opportunity to turn back to the verbal text after viewing the picture or to return to the picture after reading) would improve situation model construction compared with strictly sequential text-picture units or verbal text alone.

To the best of our knowledge, our study marks the first systematic attempt to establish the influence of illustrations on text surface, textbase, and situation model representations of written narrative text. It further contributes to understanding the impact of the processing order of written text and pictures on text comprehension, a topic that has been explored abundantly in the domain of expository text (Eitel & Scheiter, 2015) but scarcely in the area of narrative text. Although we do not generally think that the theories developed in the context of scientific text learning can simply be transferred to the field of narrative text comprehension, this study yields evidence that the ITPC framework that originated in instructional psychology (see Schnotz & Bannert, 2003) also applies well to research on narrative text. As a practical implication, we recommend that authors and typesetters of illustrated reading books place illustrations before the corresponding text passages, as long as they want readers to remember not only the state of affairs but also the meaning that the text conveys. Meaning-based representations are apparently relevant for some tasks in language teaching, such as re-narrations and content analyses.



**Acknowledgements** We thank Ekin Günaydin, Kim Hertinger, and Isabelle Weisenburger for their assistance in data collection, and the participating children, parents, teachers, and school principals for their cooperation.

**Funding** Open Access funding enabled and organized by Projekt DEAL. This research was funded by a grant from *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation) awarded to Gerhild Nieding (Ni496/9-2).

## Compliance with ethical standards

**Conflict of interest** The authors assure that no conflict of interest has to be declared.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 206–226). New York: Cambridge University Press.
- Baggett, P. (1984). Role of temporal overlap of visual and auditory material in forming dual media associations. *Journal of Educational Psychology*, 76, 408–417.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.
- Beagles-Roos, J., & Gat, I. (1983). Specific impact of radio and television on children's story comprehension. *Journal of Educational Psychology*, 75(1), 128–137.
- Beentjes, J., & van der Voort, T. (1991). Children's written accounts of televised and printed stories. *Educational Technology Research and Development*, 39(3), 15–26.
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11, 717–726.
- Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology*, 98, 182–197. <https://doi.org/10.1037/0022-0663.98.1.182>.
- Butcher, K. R. (2014). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 174–205). New York: Cambridge University Press.
- Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, 20, 155–166. <https://doi.org/10.1016/j.learninstruc.2009.02.014>.
- Cuevas, H. M., Fiore, S. M., & Oser, R. L. (2002). Scaffolding cognitive and metacognitive processes in low verbal ability learners: Use of diagrams in computer-based training environments. *Instructional Science*, 30, 433–464.
- De Koning, B. B., Wassenburg, S. I., Bos, L. T., & van der Schoot, M. (2017). Size does matter: Implied object size is mentally simulated during language comprehension. *Discourse Processes*, 54(7), 493–503. <https://doi.org/10.1080/0163853X.2015.1119604>.
- Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, 121(3), 275–277.
- Duursma, E., Augustyn, M., & Zuckerman, B. (2008). Reading aloud to children: the evidence. *Archives of Disease in Childhood*, 93, 554–557. <https://doi.org/10.1136/adc.2006.106336>.

- Eitel, A., & Scheiter, K. (2015). Picture or text first? Explaining sequence effects when learning with pictures and text. *Educational Psychology Review*, 27, 153–180. <https://doi.org/10.1007/s10648-014-9264-4>.
- Eitel, A., Scheiter, K., Schüler, A., Nyström, M., & Holmqvist, K. (2013). How a picture facilitates the process of learning from text: Evidence for scaffolding. *Learning and Instruction*, 28, 48–63. <https://doi.org/10.1016/j.learninstruc.2013.05.002>.
- Engelen, J., Bouwmeester, S., de Bruin, A., & Zwaan, R. A. (2011). Perceptual simulation in developing language comprehension. *Journal of Experimental Child Psychology*, 110, 659–675. <https://doi.org/10.1016/j.jecp.2011.06.009>.
- Engelkamp, J. (1980). Some studies on the internal structure of propositions. *Psychological Research*, 41, 355–371.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G\*Power3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences [Computer Software]. *Behavior Research Methods, Instruments & Computers*, 39, 175–191.
- Fletcher, C. R., & Chrysler, S. T. (1990). Surface forms, textbases, and situation models: Recognition memory for three types of textual information. *Discourse Processes*, 13(2), 175–190. <https://doi.org/10.1080/01638539009544752>.
- Forster, K. I., & Forster, J. C. (2016). DMDX Version 5. Retrieved December 13, 2016 from <http://www.u.arizona.edu/~kforster/dmdx/download.htm>.
- Gambrell, L. B., & Jawitz, P. B. (1993). Mental imagery, text illustrations, and children's story comprehension and recall. *Reading Research Quarterly*, 28, 264–276.
- Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199–241). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511529863.011>.
- Gibbons, J., Anderson, D. R., Smith, R., Field, D. E., & Fischer, C. (1986). Young children's recall and reconstruction of audio and audiovisual narratives. *Child Development*, 57(4), 1014–1023.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558–565.
- Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language*, 31, 129–151.
- Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, 43, 379–401. <https://doi.org/10.1006/jmla.2000.2714>.
- Greenhoot, A. F., & Semb, P. A. (2008). Do illustrations enhance preschoolers' memories for stories? Age-related change in the picture facilitation effect. *Journal of Experimental Child Psychology*, 99, 271–287. <https://doi.org/10.1016/j.jecp.2007.06.005>.
- Gunter, B., Furnham, A., & Griffiths, S. (2000). Children's memory for news: A comparison of three presentation media. *Media Psychology*, 2(2), 93–118. [https://doi.org/10.1207/S1532785XMEP0202\\_1](https://doi.org/10.1207/S1532785XMEP0202_1).
- Guttmann, J., Levin, J. R., & Pressley, M. (1977). Pictures, partial pictures, and young children's oral prose learning. *Journal of Educational Psychology*, 69, 473–480.
- Hayes, D. S., Kelly, S. B., & Mandel, M. (1986). Media differences in children's story synopses: Radio and television contrasted. *Journal of Educational Psychology*, 78(5), 341–346.
- Isbell, R., Sobol, J., Lindauer, L., & Lowrance, A. (2004). The effects of storytelling and story reading on the oral language complexity and story comprehension of young children. *Early Childhood Education Journal*, 32, 157–163.
- Jaeger, A. J., & Wiley, J. (2014). Do illustrations help or harm metacomprehension accuracy? *Learning and Instruction*, 34, 58–73. <https://doi.org/10.1016/j.learninstruc.2014.08.002>.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163–182.
- Kintsch, W. (1998). *Comprehension. A paradigm for cognition*. Cambridge: Cambridge University Press.
- Klein, O., & Kogan, I. (2013). Does reading to children enhance their educational success? *Child Indicators Research*, 6, 321–344. <https://doi.org/10.1007/s12187-012-9174-2>.
- Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 227–246). New York: Cambridge University Press.
- Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32(1), 1–19.

- Mayer, R. E. (2009). *Multimedia learning*. New York, NY: Cambridge University Press.
- Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spacial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 279–315). New York: Cambridge University Press.
- Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358–368.
- Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*, 319–334.
- Nieding, G. (2006). Wie verstehen Kinder Texte? Die Entwicklung kognitiver Repräsentationen [How do children understand text? The development of cognitive representation]. Lengerich, Germany: Pabst.
- O'Keefe, E. J., & Solman, R. T. (1987). The influence of illustrations on children's comprehension of written stories. *Journal of Reading Behavior, 19*, 353–377.
- Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition, 17*, 398–422. <https://doi.org/10.3758/bf03202613>.
- Pike, M. M., Barnes, M. A., & Barron, R. W. (2010). The role of illustrations in children's inferential comprehension. *Journal of Experimental Child Psychology, 105*, 243–255. <https://doi.org/10.1016/j.jecp.2009.10.006>.
- Pressley, M. (1977). Imagery and children's learning: Putting the picture in developmental perspective. *Review of Educational Research, 47*, 585–622.
- Reinwein, J. (2012). Does the modality effect exist? And if so, which modality effect? *Journal of Psycholinguistic Research, 41*, 1–32. <https://doi.org/10.1007/s10936-011-9180-4>.
- Ricci, C. M., & Beal, C. R. (2002). The effect of interactive media on children's story memory. *Journal of Educational Psychology, 94*(1), 138–144. <https://doi.org/10.1037/0022-0663.94.1.138>.
- Rummer, R., Schweppe, J., Fürstenberg, A., Scheiter, K., & Zindler, A. (2011). The perceptual basis of the modality effect in multimedia learning. *Journal of Experimental Psychology: Applied, 17*, 159–173. <https://doi.org/10.1037/a0023588>.
- Rummer, R., Schweppe, J., Fürstenberg, A., Seufert, T., & Brünken, R. (2010). Working memory interference during processing texts and pictures: Implications for the explanation of the modality effect. *Applied Cognitive Psychology, 24*, 164–176. <https://doi.org/10.1002/acp.1546>.
- Salomon, G., & Leigh, T. (1984). Predispositions about learning from print and television. *Journal of Communication, 34*(2), 119–135.
- Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional and situational representations. *Journal of Memory and Language, 25*, 279–294.
- Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 72–103). New York: Cambridge University Press.
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction, 13*, 141–156. [https://doi.org/10.1016/S0959-4752\(02\)00017-8](https://doi.org/10.1016/S0959-4752(02)00017-8).
- Serra, M. J., & Dunlosky, J. (2010). Metacomprehension judgements reflect the belief that diagrams improve learning from text. *Memory, 18*, 698–711. <https://doi.org/10.1080/09658211.2010.506441>.
- Seger, B. T., Hauf, J., & Nieding, G. (2020). Perceptual Simulation of Vertical Object Movement during Comprehension of Auditory and Audiovisual Text in Children and Adults. *Discourse Processes, 57*(5-6), 460–472.
- Seger, B. T., Wannagat, W., & Nieding, G. (2019). How static and animated pictures contribute to multi-level mental representations of auditory text in seven-, nine-, and eleven-year-old children. *Journal of Cognition and Development, 20*(4), 573–591.
- Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12*(2), 153–156.
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments & Computers, 31*(1), 137–149.
- Taylor, L. J., & Zwaan, R. A. (2009). Action and cognition: The case of language. *Language and Cognition, 1*(1), 45–58. <https://doi.org/10.1515/LANGCOG.2009.003>.
- Unsöld, I. H., & Nieding, G. (2009). Die Bildung prädiktiver Inferenzen von Kindern und Erwachsenen bei der kognitiven Verarbeitung audiovisueller und auditiver Texte. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 41*(2), 87–95.
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

- Wannagat, W., Waizenegger, G., & Nieding, G. (2017). Multi-level mental representations of written, auditory, and audiovisual text in children and adults. *Cognitive Processing, 18*, 491–504.
- Wannagat, W., Waizenegger, G., Hauf, J., & Nieding, G. (2018). Mental representations of the text surface, the text base, and the situation model in auditory and audiovisual texts in 7-, 9-, and 11-year-olds. *Discourse Processes, 55*(3), 290–304.
- Zwaan, R. A. (1999). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes, 28*(1), 81–88. <https://doi.org/10.1080/01638539909545070>.
- Zwaan, R. A. (2014). Embodiment and language comprehension: Reframing the discussion. *Trends in Cognitive Sciences, 18*(5), 229–234. <https://doi.org/10.1016/j.tics.2014.02.008>.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162–185.
- Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168–171.
- Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General, 135*(1), 1–11. <https://doi.org/10.1037/0096-3445.135.1.1>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.