Universiteit
Leiden
The Netherlands

Understanding narratives in different media formats: processes and products of elementary-school children's comprehension of texts and videos

Venneker, D.; Helder, A.; Broek, P.W. van den

Citation
Venneker, D., Helder, A., & Broek, P. W. van den. (2024). Understanding narratives in different media formats: processes and products of elementary-school children's comprehension of texts and videos. *Reading And Writing*. doi:10.1007/s11145-024-10573-0

Version:	Publisher's Version
License:	Creative Commons CC BY 4.0 license
Downloaded from:	https://hdl.handle.net/1887/4179771

Note: To cite this publication please use the final published version (if applicable).

Reading and Writing
https://doi.org/10.1007/s11145-024-10573-0

Understanding narratives in different media formats: Processes and products of elementary-school children's comprehension of texts and videos

Dianne Venneker[1] · **Anne Helder**[1] · **Paul van den Broek**[1]

Accepted: 27 June 2024
© The Author(s) 2024

Abstract
This study investigated similarities and differences in children's (N=83, grades 4–6) narrative comprehension between text, audio, and non-verbal video, including measures of both comprehension products and processes. The aim was to understand how children engage with information across various media and, in doing so, address inconsistent findings in the existing literature. Comprehension products were assessed through open-ended questions and recall, and comprehension processes through think-aloud protocols. Results revealed that children answered more comprehension questions correctly for video versions of the narratives than for text versions, particularly children with lower reading comprehension skills. No advantage of video over text was found for the recall task. Think-aloud responses during narrative comprehension revealed similar processing patterns for text and video, with a general tendency to report information close to the story rather than elaborate based on background knowledge. However, video versions prompted children to activate background knowledge to a greater extent than did text versions, suggesting an advantage of video at the situation-model level. Notably, differences between video and text versions cannot be attributed solely to the absence of decoding demands in video, as similar differences were found between video and audio versions. These results suggest (a) considerable similarities in both process and product across media, but (b) non-verbal videos elicit more situation-model processes than texts do, (c) non-verbal videos have an advantage over text with regard to performance on the comprehension questions, especially for less-skilled comprehenders. These findings illustrate the nuanced relationship between media affordances and comprehension processes and outcomes.

Keywords Video · Text · Narratives · Comprehension processes · Comprehension products · Reading comprehension skill

Extended author information available on the last page of the article

Published online: 18 August 2024

The ability to understand written text is an essential skill in life, and developing this ability therefore is one of the main aims of elementary education. The development of reading comprehension skills is built around two major components, namely decoding-related skills (e.g., letter recognition, phonemic awareness, word identification) and language comprehension skills (e.g., inference-making, comprehension monitoring) (Gough & Tunmer, 1986; Kendeou et al., 2014; McMaster et al., 2023; Oakhill, 2020). This implies that media formats that do not rely on decoding-related skills, such as videos, may provide promising venues for the development and teaching of comprehension skills (Kendeou et al., 2005, 2020; McMaster et al., 2019, 2023). This possibility is supported by findings that similar cognitive processes are used for text and video (Kendeou et al., 2008; Magliano et al., 2013), and that comprehension skills are positively related across these media formats (Gernsbacher et al., 1990; Kendeou et al., 2008; Paris & Paris, 2003). This suggests that comprehension is medium-independent and, therefore, a general skill. However, there also is empirical evidence for comprehension differences between text and video, presumably because of different media affordances (Beentjes & Van der Voort, 1991; Beentjes & Voort, 1993; Furnham et al., 2002; Gunter et al., 2000; Magliano et al., 2013; Walma Van der Molen & Van der Voort, 1997, 1998, 2000). To reconcile these seemingly contradictory findings, we examine if and how narrative comprehension is affected by the medium format in which the narrative is presented by comparing comprehension processes and products in written and animated narratives for children in upper elementary grades. In doing so, we aim to better understand how children engage with and process information across various media formats.

Comprehension during reading: constructing a coherent mental representation

Central to successful reading comprehension is a reader's construction of a coherent mental representation of what the text is about (Kintsch & van Dijk, 1978; Trabasso & van den Broek, 1985). This mental representation is constructed at three levels: surface structure, textbase, and situation model (Kintsch, 1988). The surface structure refers to the exact wording of a text. The textbase captures the meaning of the text and includes information that is explicitly mentioned in the text as well as relations between text elements that are directly derived from the text. The situation model includes events, situations, and scenarios that are not explicitly mentioned in the text but inferred by readers from their background knowledge. In this fashion, "a situation model is … a construction that integrates the textbase and relevant aspects of the comprehender's knowledge" (Kintsch, 1998, p. 107), enabling a deeper understanding of the text.

Constructing coherence is a dynamic and cyclical process (Kintsch, 1988; van den Broek & Helder, 2017). The mental representation gradually evolves as the reader proceeds through a text and is continually updated with each new piece of text information (Goldman & Varma, 1995; Kendeou et al., 2013; Perfetti & Helder, 2021; van den Broek et al., 1999). Thus, what happens during reading (i.e., comprehension *processes*) determines what is included in the mental representation and what the reader

understands and remembers after reading (i.e., comprehension *products*). Therefore, to understand how children successfully construct coherent mental representations of narratives, it is important to examine both comprehension processes and comprehension products. Although the construction of a coherent mental representation has been investigated primarily in the context of reading texts, it is considered important for the comprehension of other media formats as well, including visual narratives (Magliano et al., 2013). In the next sections, we discuss similarities and differences across media based on previous research.

Comprehension as a medium-invariant skill

Although the input from text and video differ, comprehension across these media formats likely involves many of the same cognitive mechanisms (Cohn, 2013; Magliano et al., 2013). Indeed, research with adults has demonstrated that causal, bridging, knowledge-based, and predictive inferences are generated not only during comprehension of texts (Graesser et al., 1994; van den Broek et al., 2001) but also during comprehension of visual narratives (Hutson et al., 2017, 2018; Magliano et al., 1996, 2016; Tibus et al., 2013). Additionally, strategies used for comprehension, such as identifying main ideas, summarizing, and comprehension monitoring, are similar for expository text and video (List, 2018). Similar findings have been reported for elementary-school children, who make inferences about characters' traits, goals, actions, and emotions when viewing videos (Diergarten & Nieding, 2015; Kendeou et al., 2008; Ricci & Beal, 2002). They are sensitive to the causal structure of audiovisual narratives, and form a coherent mental representation of their content (Lynch et al., 2008; van den Broek et al., 1996), just as they do for textual narratives. Such results underscore the common cognitive processes involved in comprehending both textual and visual information.

Besides the considerable similarities in comprehension processes, there is also evidence that the comprehension products positively correlate across media formats. For example, children with better comprehension of text also performed better in comprehending audiovisual narratives across a wide range of measurements, including recall, sensitivity to causal structure, and answers to open-ended inferential questions (Kendeou et al., 2008). Interestingly, this relation between performance on comprehension of written and non-written media even persisted when videos that did not include language were used, in both children (Paris & Paris, 2003) and adults (Gernsbacher et al., 1990). These findings suggest that comprehension skills generalize beyond written contexts.

Overall, prior studies imply that children's comprehension processes and the resulting mental representations are not unique to the medium in which information is presented, and that skills acquired in one medium could transfer to another (Kendeou et al., 2020). Shared processes could be promising targets for intervention via media that do not require technical reading skills, and thus be building blocks for effective educational approaches that leverage the strengths of different media formats to enhance comprehension. McMaster et al. (2019), for example, showed that practicing inference generation in a video context improved children's language

comprehension. Again, this effect may extend beyond verbal contexts: A pilot study by Maine and Shields (2015) suggests that receiving explicit instruction on comprehension strategies with non-verbal visual narratives has the potential to improve reading comprehension skills in children.

Comprehension differences between texts and videos

The research described in the previous section demonstrates that comprehension processes and products are similar for text and video formats, but this does not necessarily mean they are identical. The finding that some processes are shared does not rule out that medium-specific processes may occur as well (Wolf et al., 2019). Medium-specific processes are generated in response to features unique to the given medium format that can differentially facilitate or hinder how information is processed, also referred to as media affordances (Magliano et al., 2013, 2019). For example, one obvious difference between text and video is that information extraction is predominantly based on verbal versus visual information, respectively. Vocabulary plays an indispensable role in text comprehension, but a smaller role in audiovisual narratives because verbal information is supported by visual information (Kendeou et al., 2020). In visual narratives without a language component, vocabulary does not play a role at all. Thus, different media formats each may uniquely afford some comprehension processes and, in so far as they share processes, they may require or support such processes to different degrees.

The majority of existing research demonstrates a comprehension advantage of video over text, at least for children in upper elementary grades: They remember information presented in videos better than information presented in text (Beentjes & Van der Voort, 1991; Furnham et al., 2002; Gunter et al., 2000; Walma Van der Molen & Van der Voort, 1997, 1998, 2000; but see Salomon, 1984 for discrepant findings). This advantage of visual narratives over written narratives occurred mostly for implicit information inferred by the reader or viewer (Beentjes & Van der Voort, 1993), especially for situation-model information (Wannagat et al., 2017). A potential explanation for these findings is that visual narratives are processed more intuitively and effortlessly than text, because our cognitive system is particularly adept at visual perception (e.g., visual ease assumption; for critical reviews, see Cohn, 2020 and Coderre, 2020, who emphasize that visual media – and, by extension, other media – require their own literacy skills). To identify factors and circumstances that contribute to an advantage of video over text, more research is needed.

Current study

The current study investigates children's narrative comprehension in the upper grades of elementary school, centered around two contrasting media formats: written text without accompanying visual information versus animated videos without accompanying linguistic information. All narratives were available in a text and a video version. A central contribution of our study is that we measure how children

process narratives *during* reading and viewing (comprehension processes, i.e. think-aloud protocol) as well as what children remember *after* reading or viewing narratives (comprehension products, i.e. open-ended questions and free recall). These complementary measures allow us to examine potential differences in cognitive processes that could lead to differential comprehension of the narratives as a function of format. We distinguish between two types of comprehension processes: story-based and knowledge-based. Story-based processes are closely related to the information provided in the narrative, analogous to the textbase described above. Knowledge-based processes extend beyond the information provided in the narrative and concern situation-model information. The study is guided by three research questions: (1) In what ways is narrative comprehension similar or different across text and non-verbal videos? (2) For aspects that differ across media formats, do these differences occur in comprehension processes, comprehension products, or both? (3) Are any such differences dependent on children's reading comprehension skill?

Most studies that reported an advantage of video over text regarding comprehension products, have used materials that combined visual information with verbal information (Beentjes & Van der Voort, 1991; Beentjes & Voort, 1993; Furnham et al., 2002; Gunter et al., 2000; Walma Van der Molen & Van der Voort, 1997, 1998, 2000; Wannagat et al., 2017). The present study adds to the existing literature by investigating if the advantage of video over text generalizes to non-verbal videos, which would suggest that visual presentation in itself bears an advantage. If so, this may also be reflected in comprehension processes, as they guide the construction of the mental representation of the narrative. There is some suggestion in prior research that differences between text and video mostly occur for implicit information (Beentjes & Van der Voort, 1993) and, possibly, mostly at the situation-model level of the mental representation (Wannagat et al., 2017). Such findings suggest that differences could be expected primarily in knowledge-based processing. Furthermore, we expect that potential differences in processes and products would be more pronounced for less-skilled than for skilled comprehenders, because non-verbal videos do not require decoding and/or language comprehension skills.

Although not the focus of the study, we also included an audio-only version in addition to the text and video versions in an attempt to separate the influence of decoding from the role of other media affordances. If potential differences between text and video mostly originate from varying decoding demands, we would expect to find no differences between the audio-only and video conditions. However, if media affordances other than decoding demands play a role as well, differences between the audio-only and video conditions are expected to be similar in direction to differences between the text and video conditions.

Method

Participants

Eighty-three children (42 girls, mean age 10.4, age range 8.7–12.1) from grade 4 ($n=30$), grade 5 ($n=26$), and grade 6 ($n=27$) participated in this study. They were

recruited from three elementary schools in the south-west of the Netherlands. Active written consent was obtained from the participating children and their caregivers according to the Declaration of Helsinki, and the study was approved by the Ethical Committee of the Institute of Education and Child Studies at Leiden University. Data from five children were excluded because of problems during task administration.

Materials

Three age-appropriate narratives were available as a video, a text, and an audio version. The video versions were existing animation videos from YouTube with a duration of 3 to 4 min (CGMeetup, 2018; TheCGBros, 2015; Southeastern Guide Dogs, 2018). They included music and sound, but no spoken language. To make the videos suitable for our study, we made minor changes, such as removing the end credits.

A text version of each narrative was constructed based on well-established story-grammar elements, such as the setting, character goals, initiating events, actions, and consequences (Mandler & Johnson, 1977; Stein & Glenn, 1979; van den Broek et al., 1996). First, a story-grammar outline of the narrative's fundamental structure was constructed for each video. The corresponding text version was written around this fundamental structure and complemented with detailed information from the video (for text versions of the three narratives, see Supplementary Information).

All texts were written from a third-person perspective, and similar in length with an average of approximately 800 words. They were matched on readability based on a reading fluency level index and a reading comprehension level index developed by the Dutch Institute for Test Development (Cito) (P-Clib; Evers, 2008). The texts were written to match the average decoding skills of a fourth-grade reader, and the average comprehension skills of a fifth-grade reader. The audio versions were voice recordings of the text versions and had a duration of five to six minutes.

The narratives consisted of seven to nine paragraphs (text version) or sections (audio and video versions). The start and end points of each paragraph or section were identical across media formats. Each narrative started with the title (e.g., Max and the Heron), displayed visually in the text and video formats and auditorily in the audio format. More detailed information about the stories can be found in Table 1. Each child was presented with three narratives, one in text, one in audio, and one in video format.

Measures

Comprehension products

Open-ended questions. After each narrative, children were asked to answer ten open-ended questions, addressing either textbase or situation-model information[1]

[1] The reliability coefficients for the comprehension questions of the three stories were Cronbach's $\alpha = 0.49$, $\alpha = 0.54$, and $\alpha = 0.61$, respectively. However, because the questions addressed both textbase and situation-model information, they did not measure a single latent construct. This violates the tau-equivalence assumption of Cronbach's alpha and, thus, leads to an underestimation of reliability (e.g., Green & Yang, 2009; Tavakol & Dennick, 2011).

Table 1 Descriptive information about the narratives

Narrative features	Max and the Heron[a]	Mouse for Sale[b]	Pip's Exam[c]
Number of paragraphs	8	7	9
Number of sentences	56	61	64
Number of words	748	781	851
Average number of words per sentence	13.4	12.8	13.3
Average number of letters per word	4.4	4.5	4.5
Proportion of frequent words	0.72	0.74	0.74
Proportion of unique words	0.51	0.47	0.48
Average reading duration (min: sec)	4:49	4:29	5:37
Audio duration (min: sec)	5:01	5:18	5:45
Video duration (min: sec)	3:15	3:32	3:41

[a]https://www.youtube.com/watch?v=1lo-8UWhVcg

[b]https://www.youtube.com/watch?v=OzYwE3Tst1Y

[c]https://www.youtube.com/watch?v=07d2dXHYb94

(for the open-ended questions, see Supplementary Information). Recordings of their answers to all questions were transcribed verbatim. Answers were assigned 1 point if fully correct and 0.5 point if partially correct. The maximum score was 10 points. Three independent raters scored all of the transcripts, resulting in an interrater agreement of $k = 0.80$. Disagreements were resolved through discussion.

Recall task. After each narrative, children were asked to retell the story in their own words as if they would to a peer who was not familiar with the story. They were allowed to use their own words and were asked to pay attention to describe the events in the correct order as much as possible. Recordings of the recall protocols were transcribed verbatim. To assess how well children remembered the content of the stories, the story-grammar structures described in the materials section were used. For each correctly recalled story element, one point was awarded. A proportion score was calculated to account for the slight variation in the maximum number of story elements that could be recalled per story (range from 21 to 25). Two independent raters scored approximately 10% of the transcripts, resulting in an interrater agreement of $k = 0.81$. Remaining transcripts were scored individually, and disagreements were resolved through discussion.

Comprehension processes: think-aloud protocol

To identify comprehension processes that take place while processing narratives, a think-aloud procedure was employed. Think-aloud procedures have been used to a considerable extent in the context of reading research (e.g., Ericsson & Simon, 1993; Magliano et al., 2007), and have shown to be informative for other media formats as well (e.g., Bezdek et al., 2013; Magliano et al., 1996, 2016). In the current study, children were instructed to say out loud whatever they were thinking after each paragraph or section. Recordings of the verbal think-aloud responses were transcribed verbatim and parsed into idea units.

Based on previous comprehension research (e.g., Carlson et al., 2014; McMaster et al., 2012; Seipel et al., 2017; Trabasso & Magliano, 1996; van den Broek et al., 2001), a coding scheme was designed to assign each idea unit to a processing category. This resulted in three main categories: (1) story-based, (2) knowledge-based, and (3) metacognitive comments, each consisting of multiple subcategories (see Table 2). The story-based category included story elements that were explicitly mentioned in the story or depicted in the video, and relations between multiple of such elements that are directly derived from the story content, within or across sections. Thus, this category included responses that stayed relatively close to the story content. The knowledge-based category included inferences that draw on the participant's background knowledge. For the story-based and knowledge-based categories, responses were also classified as valid or invalid, depending on whether they made sense in the context of the story or not. The metacognitive comments category included responses that are indicative of reflective processes, e.g., expressing one's opinion or monitoring one's comprehension. This category made up only a small percentage of all responses (ranging from 3 to 5% across media formats), which is in line with findings from previous think-aloud studies that report a paucity of such responses (Dahl et al., 2021; Karlsson et al., 2018; Seipel et al., 2017). As this category was not of main interest for the current study, it was not analyzed separately. Finally, responses that fit none of these three categories, such as task-related comments or inaudible segments, were combined into a remainder category and excluded from further analysis. The upcoming analyses focus on differences in the valid story-based and knowledge-based categories. Two independent raters coded 30% of the transcripts, resulting in interrater agreements of $k=0.74$ for the main categories and $k=0.63$ for the subcategories. Analyses were conducted only on the main categories. The remaining transcripts were coded by the raters individually and problematic cases were resolved through discussion.

For each participant, the proportions of valid responses in the story-based and knowledge-based categories were calculated, relative to the total number of responses (which also include invalid responses and metacognitive comments). This was done separately for each of the narratives the participant read, viewed, or heard. Next, a difference score was calculated for each medium for each participant, by subtracting the proportion of knowledge-based responses from the proportion of story-based responses. A difference score of zero indicates an equal focus on story-based and

Table 2 Labels, descriptions and examples of think-aloud response categories

Main category	Subcategories	Examples
Story-based	Descriptive statements based on the narrative, connecting inferences, reinstatement inferences	"The owner tells his dog to be quiet."
		"The dog and his owner are going on a fishing trip."
Knowledge-based	Elaborative inferences, associations, predictions, emotion inferences, perspective-taking	"Young birds cannot digest fish yet."
		"I think the heron is going to try to steal the worms again."
		"I think the dog feels sorry for the heron."
Metacognitive comments	Comprehension monitoring, questions, evaluations, affective responses	"I don't understand why he did that."
		"I think that's a very smart thing to do."

knowledge-based information, a positive difference score indicates a stronger focus on story-based information than on knowledge-based information, and a negative difference score indicates a stronger focus on knowledge-based information than on story-based information. Comparison of these difference scores provides insight into whether the balance between story-based and knowledge-based information varies across media formats.

Reading comprehension skill

As a measure of reading comprehension skill, results from a standardized paper-and-pencil test were used. This test was developed by the Dutch Institute for Test Development (Cito), a leading developer of national assessments in Dutch elementary education. The Cito reading comprehension test consists of a mix of approximately fifty text-based and inferential questions about sixteen to eighteen texts of various genres (e.g., narrative and expository). The quality of this test has been assessed by the Dutch Committee on Tests and Testing (COTAN; https://www.cotandocumentatie.nl), which evaluates a wide range of tests and questionnaires that are used in the Netherlands, providing them with labels ranging from insufficient to good. According to COTAN guidelines, construct validity of this test ranges from sufficient (grade 4) to good (grade 5 and 6). Reliability is good for all grades. For more information on this test, see reports by Tomesen et al. (2017, 2018, 2019).

The test was administered by the school's class teachers as part of their curriculum and children's scores were obtained from their most recent school records. The skill scores provided by schools ranged between 128 and 281 ($M = 188$, $SD = 31$). To facilitate interpretation of our analyses, we converted the skill scores to z-scores.

Design and procedure

Children were presented with each of the three stories in a within-subjects design: one as a written text, one as an audio recording, and one as a video. The medium format of a story and the order in which the stories were presented were counterbalanced following a Latin square procedure. The stories were presented on a laptop, with each paragraph or section of a text, audio recording, or video appearing on a separate screen (for screenshots, see Supplementary Information). While listening to the audio version participants saw a blank screen with a fixation cross in the middle.

Participants were tested in a single individual session, which was audio-recorded. For each narrative, three different tasks were administered in a fixed order: thinking aloud, retelling the story, and answering open-ended comprehension questions. There was a short break after each story and children received a small reward for their participation.

At the beginning of the session, the test leader explained the procedure to the child. For the think-aloud protocol, the test leader instructed the children to verbalize everything that they were thinking in response to a visual prompt that appeared after each paragraph or section of the story (image of a thought balloon). If the child did not respond to a prompt, the experiment leader explicitly asked "What are you thinking now?" before moving on to the next paragraph or section. They emphasized that

there were no right or wrong answers, and that they were just interested in what came to the child's mind while reading, listening, or viewing the narrative. To familiarize the children with the think-aloud procedure, the session started with a short practice text consisting of two paragraphs. The test leader used a script to model examples of what to say in response to the first paragraph. The child practiced the procedure on the second paragraph. For the recall, children were instructed to retell the narrative as if they would to a peer who was not familiar with the storyline. If a child finished their retell, the experiment leader always asked "What else do you remember?" until they indicated that they could not recall anything else. For the comprehension questions, children were asked to answer verbally, and they were encouraged to always try to give an answer, even if they were not sure. They did not receive feedback about the accuracy of their answers. Since the used videos are publicly accessible on video-sharing platforms, we verified whether children were already familiar with some of the stories before participating. In order to preserve as much data as possible, we coded the observations for known stories as missing values ($N=24$) instead of fully excluding participants that were familiar with one narrative from the sample.

Data analysis

All analyses for comparisons between media formats were performed using linear mixed models in R Studio (*lme4* package; Bates et al., 2015). Wald chi-square testing (Type II) (*car* package; Fox & Weisberg, 2019) was used to assess main effects and interaction effects. Post-hoc tests were done using pairwise comparisons between the three media formats (*emmeans* package; Lenth, 2023). The models included medium format (categorical: text, audio, video), reading comprehension skill (continuous: *z*-scores of a standardized reading comprehension test), and their interaction as fixed factors[2]. A random intercept was added for each participant to account for variability between subjects. The number of correctly answered comprehension questions and the proportion of correctly recalled story elements for each participant were used as dependent variables for the comprehension products. The difference between the proportion of story-based and knowledge-based think-aloud responses was used as the dependent variable for the comprehension processes[3]. When the difference scores for text and video were significantly different, two separate follow-up analyses were performed with the proportion of story-based responses and the proportion of knowledge-based responses as dependent variables, respectively.

[2] To control for a possible effect of grade, we also ran our analyses with grade as an additional fixed factor. The outcomes showed no interaction between grade and medium format (all p's$\geq$0.12), for any measure (open-ended questions, recall, think aloud). In addition, none of the effects we report changed in their significance or non-significance in these additional analyses.

[3] For each dependent variable, correlations between media formats can be found in the Supplementary Information.

Results

Comprehension products

Open-ended questions

To examine if there are differences in what children remember after processing narratives, we tested whether their performance on open-ended questions differed across media formats. The Wald chi-square tests showed that there was a significant main effect of medium format ($\chi^2(2)=15.76$, $p<.001$), with higher scores in the video condition ($M=8.9$, $SD=1.0$) than in the text ($M=8.2$, $SD=1.6$) ($B=0.73$, $SE=0.19$, $p<.001$) and audio ($M=8.4$, $SD=1.3$) ($B=0.44$, $SE=0.18$, $p=.049$) conditions. There was no significant difference between the text and audio conditions ($B=-0.29$, $SE=0.18$; $p=.219$). There was a significant main effect of reading comprehension skill ($\chi^2(1)=35.00$, $p<.001$): Better reading comprehension skill was associated with better performance on the comprehension questions. Importantly, these main effects were qualified by an interaction between medium format and reading comprehension skill ($\chi^2(2)=11.05$, $p=.004$), see Fig. 1. The advantage of video over text was stronger for less-skilled comprehenders than for skilled comprehenders, which were defined as scoring one standard deviation below and above the sample mean, respectively. Post-hoc tests revealed that skilled comprehenders performed comparable across text, audio and video conditions (p's≥ 0.794), whereas less-skilled comprehenders performed significantly better in the video condition than in the text ($B=1.33$, $SE=0.26$, $p<.001$) and audio conditions ($B=0.76$, $SE=0.26$, $p=.010$). There

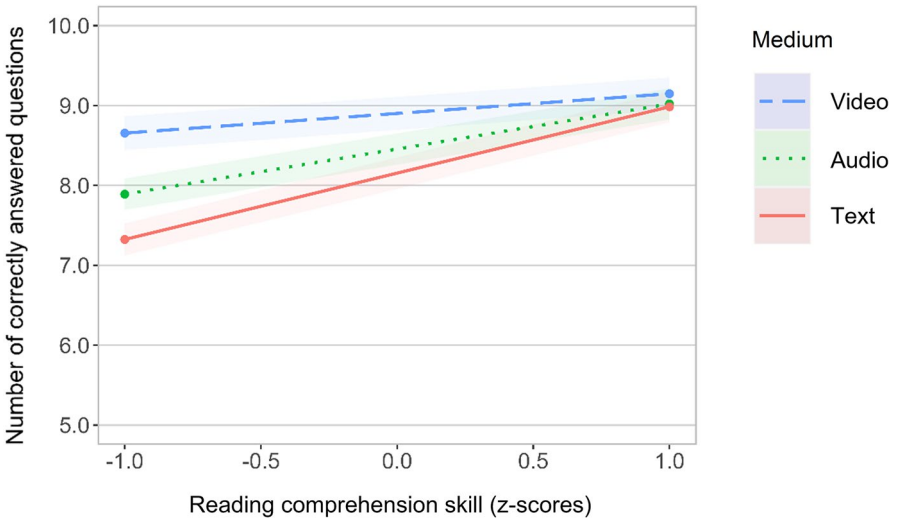

Fig. 1 Estimated marginal means of correctly answered comprehension questions in the text, audio, and video condition as a function of reading comprehension skill. *Note*. Less-skilled comprehenders were defined as scoring one standard deviation below the sample mean; Skilled comprehenders were defined as scoring one standard deviation above the sample mean. Error bands represent 1 standard error

was no significant difference between the text and audio condition for less-skilled comprehenders (B=-0.57, SE=0.25, p=.064).

Recall of story elements

We examined if performance on the recall task differed across media formats. The Wald chi-square tests showed that there was no main effect of medium format ($\chi^2(2)=4.70$, $p=.096$), indicating that the proportion of recalled story elements was similar for text ($M=0.60$, $SD=0.19$), audio ($M=0.65$, $SD=0.18$), and video ($M=0.61$, $SD=0.17$) conditions. There was a significant main effect of reading comprehension skill ($\chi^2(1)=6.25$, $p=.012$): Better reading comprehension skill was associated with better performance on the recall task. There was no interaction between medium format and reading comprehension skill ($\chi^2(2)=2.21$, $p=.331$), indicating that the consistency in recall performance across media formats occurred for both skilled and less-skilled comprehenders.

Comprehension processes

To examine whether the balance of story-based responses and knowledge-based responses differed across media formats, we compared difference scores in proportions of responses for the two categories (see Table 3). Note that all difference scores were positive, indicating a stronger focus on story-based information than on knowledge-based information. The Wald chi-square tests showed a significant main effect of medium format ($\chi^2(2)=21.33$, $p<.001$), with the difference between proportions of story-based and knowledge-based responses being significantly smaller for the video condition than for the text ($B=-0.193$, $SE=0.042$, $p<.001$) and audio conditions ($B=-0.103$, $SE=0.042$, $p=.042$), and comparable for the text and audio conditions ($B=0.091$, $SE=0.040$, $p=.063$). There was no significant main effect of reading comprehension skill ($\chi^2(1)=0.116$, $p=.733$), nor an interaction effect between medium format and reading comprehension skill ($\chi^2(2)=2.00$, $p=.368$).

To investigate the origins of the differences in proportion of story-based and knowledge-based responses between the text and video conditions, follow-up analyses were performed. Two separate models were used, with the proportion of story-based responses and the proportion of knowledge-based responses as the dependent variables, respectively. For the story-based responses, the Wald chi-square tests showed a significant main effect of medium format ($\chi^2(2)=12.73$, $p=.002$). Post-hoc tests revealed that text versions elicited significantly higher proportions of story-based responses than video versions ($B=0.083$, $SE=0.024$, $p=.002$). There was no difference in the proportion of story-based responses between the audio and video versions ($B=0.040$, $SE=0.024$, $p=.214$), or the text and audio versions ($B=0.043$, $SE=0.022$, $p=.130$). There was no significant main effect of reading comprehension skill ($\chi^2(1)=0.035$, $p=.852$) and no significant interaction between medium format and reading comprehension skill ($\chi^2(2)=0.922$, $p=.631$). For the knowledge-based responses, the Wald chi-square tests showed a significant main effect of medium format as well ($\chi^2(2)=31.02$, $p<.001$). Post-hoc tests revealed that video versions elicited significantly higher proportions of knowledge-based responses than the text

Table 3 Average proportion scores for story-based and knowledge-based think-aloud responses in the text, audio, and video conditions

Main category	Text	Audio	Video
	M (*SD*)	*M* (*SD*)	*M* (*SD*)
Story-based	0.68 (0.21)	0.63 (0.23)	0.59 (0.18)
Knowledge-based	0.25 (0.20)	0.30 (0.21)	0.36 (0.16)
Difference score	0.42 (0.41)	0.34 (0.43)	0.23 (0.34)

Note The difference score equals the proportion of story-based minus the proportion of knowledge-based responses. Proportions and difference scores are rounded to the nearest 0.01

($B=0.110$, $SE=0.020$, $p <. 001$) and audio versions ($B=0.063$, $SE=0.020$, $p=.006$). In addition, the audio versions elicited significantly higher proportions of knowledge-based responses than the text versions ($B=0.047$, $SE=0.019$, $p=.035$). There was no significant main effect of reading comprehension skill ($\chi^2(1)=0.242$, $p=.623$), and no significant interaction between medium format and reading comprehension skill ($\chi^2(2)=3.55$, $p=.170$).

Discussion

In this study, we examined children's narrative comprehension across media formats in the upper grades of elementary school, focusing on both comprehension products and comprehension processes. We also investigated whether similarities or differences in the processing and memory of narratives across media formats depend on children's reading comprehension skill. With respect to the comprehension *products*, the results indicate that children answered more comprehension questions correctly after viewing video versions of the narratives than after reading text versions (or listening to audio versions). This benefit of video was significant for less-skilled comprehenders but not significant for skilled comprehenders. Performance on the recall task did not differ across media but it was influenced by comprehension skill: For all formats skilled comprehenders remembered more story elements than did less-skilled comprehenders. With respect to the comprehension *processes*, think-aloud responses were dominated by story-based information in all media formats, but children did produce a more balanced distribution of story-based and knowledge-based responses for video than for text: Video elicited relatively fewer story-based responses than did text (but not audio) and relatively more knowledge-based responses than did text (and audio). This pattern was not affected by reading comprehension skill. In the next paragraphs, we will discuss these findings in more detail.

Similarities and differences between text and non-verbal videos

Comprehension products

Children's narrative comprehension was better after viewing video versions than after reading text versions, as indicated by their performance on the comprehension questions. These findings extend prior reports of an advantage of *language-based* video over text (Beentjes & Van der Voort, 1991; Beentjes & Voort, 1993; Furnham

et al., 2002; Gunter et al., 2000; Walma Van der Molen & Van der Voort, 1997, 1998, 2000), by showing that this advantage also occurs for video that is entirely visual and does not include language. Thus, it is not merely the combination of the visual and auditory input that enhances comprehension in earlier studies, for example by repeating information or by allowing dual coding of the information (Paivio, 1969; Mayer, 2014). The visual channel alone plays a pivotal role. Importantly, the advantage of video over text was stronger for less-skilled comprehenders than for skilled comprehenders. While previous proposed that this might be the case, results were mixed and limited to audiovisual narratives containing language (Beentjes & Van der Voort, 1993; Walma Van der Molen & Van der Voort, 1997, 1998, 2000). The current study indicates that non-verbal video formats particularly improve understanding by less-skilled comprehenders. A possible cause for the differences between video and text formats is that, unlike written texts, video's do not require the child to decode words. It is possible that such decoding demands while reading the text versions consumes cognitive capacity that otherwise could be used for higher-level comprehension processes (LaBerge & Samuels, 1974; Perfetti, 1985, 2007). Decoding does not seem to be a complete explanation for the current results, however, because the video version elicited better comprehension than the audio version as well. This suggests that comprehension-related processes also contribute to the observed differences.

Notably, the aforementioned patterns of comprehension differences as a function of medium were not observed in the results of the recall task. A possible explanation for this finding could be that recalling a story relies less strongly on situation-model information than answering questions (Kintsch, 1994). Indeed, previous research shows that fourth-graders predominantly reproduce information that was provided by the text when asked to recall a story, as opposed to inferred information (Kim et al., 2021).The number of individual pieces of information that are retrieved from memory may therefore be a less sensitive measure to detect differences in the mental representation than comprehension questions that require connections across story elements and between story elements and background knowledge. Another explanation could be that perhaps retelling non-verbal videos is more challenging than retelling text because retell more heavily relies on language production abilities. It seems probable that *verbally* describing what was *visually* depicted in the video requires more mental effort than reactivating what was verbally stated in the text. As a result, the advantage of video over text that we detected for the comprehension questions may be cancelled out for the recall task.

Comprehension processes

Children engaged in similar processes when viewing video versions and when reading text versions of the narratives. Both story-based and knowledge-based response categories and also each of their subcategories in the think-aloud coding scheme occurred in each of the media formats. This confirms that comprehension processes are indeed shared across media formats, at least for the overt comprehension processes we examined (for a discussion on whether it is appropriate to interpret think-aloud responses as being indicative of comprehension processes see McNamara et al., 2023). In addition, the patterns of processes were similar to a considerable extent,

with a stronger focus on story-based information than on knowledge-based information. However, the children also engaged in partially distinct patterns of processing of the non-verbal video and textual narratives. On the one hand, when viewing visual narratives, they focused more on activating information from their background knowledge than when reading textual narratives. Background knowledge is a central aspect of the generation of inferences (Cain et al., 2000; Elbro & Buch-Iversen, 2013; Graesser et al., 1994; Kendeou et al., 2008; McMaster et al., 2012) which, in turn, are essential for creating a coherent and rich representation of a narrative (Kintsch, 1988, 1994; Trabasso & van den Broek, 1985). In this sense, the advantage of video appears to be mostly at the situation-model level, as proposed by Wannagat et al. (2017). This also aligns with previous findings that semantic knowledge is more readily accessible in visual media compared to written text (Bajo, 1988; Seifert, 1997). On the other hand, children focused more on story-based information when reading textual narratives than when viewing visual narratives. This may be attributed to the lexical and syntactic specificity of text[4], which allows for a more precise representation of story-based information. Thus, our findings are in line with the notion that the advantages of different media formats may depend on the level of the mental representation being addressed, because different media have different affordances (e.g., Magliano et al., 2013, 2019; Öncel et al., 2024).

Combining comprehension processes and products

When we consider processes and products in conjunction, it appears that the increased processing of knowledge-based information elicited by the video format compared to text leads to the construction of a richer, more coherent mental representation at the situation-model level, as reflected in improved performance on the comprehension questions. This pattern was particularly visible for less-skilled comprehenders. Skilled readers showed the same distinct processing pattern for video compared to text but, unlike their less-skilled peers, for them this did not lead to media effects on comprehension outcomes. Thus, the differences in processing patterns across media formats are similar for less-skilled and skilled comprehenders, yet the effects on comprehension product are different. Prior studies have also reported results whereby less-skilled and skilled comprehenders are similar in processing yet differ in the comprehension outcomes (e.g., Helder et al., 2016; Kraal et al., 2018). There are several possible reasons that this may happen. One possibility specific to the current study is that a ceiling effect on the comprehension questions measure by skilled comprehenders obscured a potential beneficial effect of video – their scores for each of the three conditions were around nine out of ten questions correct. A second possibility is that the *quality* of the processes that took place was higher for skilled comprehenders than for less-skilled comprehenders (Fox, 2009), or that skilled comprehenders were more effective than less-skilled comprehenders in translating online processes into a mental representation, regardless of whether these processes are more text-based or more knowledge-based. A third, similar possibility is that skilled comprehenders engaged in additional processes that are not captured in our think-aloud method (e.g.,

[4] We thank Joe Magliano for this suggestion.

Seipel et al., 2017), leading to better comprehension outcomes. These possibilities call for further investigation, possibly with methods other than think aloud.

Limitations and future directions

Although the findings of this study contribute to our understanding of comprehension differences across media, some limitations have to be taken into account while interpreting the results. First, the absence of language in the video versions as opposed to the presence of language in the text versions makes it more challenging to directly compare the precise information conveyed in both formats. However, this contrast allows us to leverage the unique affordances of each medium format, rendering the comparisons meaningful in real-world settings.

Second, we used an audio condition to control for the role of decoding in the difference between text and video conditions, under the assumption that both were identical except for the need to decode words. Although the conveyed linguistic input was indeed identical, there may be other relevant differences between audio and text formats, such as the (im)possibility to control the pace of information presentation or the impact of the narrator's intonation or emphasis of particular words. Consequently, the audio condition serves as a useful though not perfect basis for comparison.

Third, as the goal of our study was to measure both processes and products of comprehension, we opted for a design with multiple measurements within the same participant (i.e., within-subjects). The advantage of this design is that we can connect patterns of processing with comprehension outcomes, a central contribution of our study. A disadvantage, however, is that it is possible that performance on one measure influenced performance on a subsequent measure. Because the tasks were presented in the same order for each medium and for each participant, task order is unlikely to affect our findings on the similarities and differences across media. We therefore believe that in our study the benefits of using a within-subjects design outweigh the disadvantage of potential carry-over effects across the tasks.

Fourth, we used think-aloud protocols to gain insight into the processes that take place during narrative comprehension. We used a fairly broad coding scheme to categorize the think-aloud responses, and it would be valuable to focus future research on the evaluation of how media formats affect comprehension processes at a more detailed level (e.g., Öncel et al., 2024). Additionally, although think-alouds offer valuable information, they are constrained by the conscious thoughts available to participants and their ability to articulate them clearly. To fully understand the current findings and those of prior research, it would be important to understand both the conscious (strategic) and subconscious (automatic) mechanisms by which video presentation exerts its influence on comprehension. Employing alternative or additional measurement methods, such as eye tracking, could significantly contribute to this understanding.

Finally, it is important to note that the majority of the children in our sample scored above average on the standardized test of reading comprehension. As a result, truly poor comprehenders may be underrepresented in our sample (Cain & Oakhill, 2006; Catts et al., 2006). It may also partly explain the ceiling effect for the open-

ended questions mentioned previously. Future research could focus on poor comprehenders in particular to see if the results generalize to this specific population as well.

Conclusion

In conclusion, the present study extends the existing literature on similarities and differences in comprehension across different media formats, in terms of both process and product. Furthermore, it attempted to bridge a gap in the current knowledge by specifically providing insights into videos *without* language. Our results showed that non-verbal videos elicited relatively more situation-model processing than texts and, particularly for less-skilled comprehenders, resulted in a stronger representation of the narratives. From a practical perspective, we provide insight into the extent to which non-written formats may be used to foster comprehension skills in reading. Our findings suggest video narratives may be a good context for practicing comprehension skills and strategies. If educational professionals aspire to use non-textual materials to train reading comprehension skills, the finding that medium-specific features affect less-skilled comprehenders to a larger extent than skilled comprehenders is important to keep in mind – especially because the former is usually the ability group targeted by interventions. A skill trained in one medium format might not automatically transfer to another, and explicit instruction and practice to facilitate transfer seems essential (McMaster et al., 2019). Overall, our findings call for a more advanced investigation of the unique and common affordances of different types of media and their influence on the cognitive mechanisms that underlie comprehension.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11145-024-10573-0.

Acknowledgements This research was funded by the Netherlands Initiative for Education Research (NRO) under grant number 40.5.18300.038. The authors would like to thank Brechtje van Zeijts, Lesya Ganushchak, Björn de Koning, and Huib Tabbers for their valuable comments regarding the study design.

Data Availability Data are available at https://osf.io/dgaes/

Declarations

Conflict of interest We have no conflicts of interest to disclose.

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

Bajo, M. T. (1988). Semantic facilitation with pictures and words. *Journal of Experimental Psychology: Learning Memory and Cognition*, *14*(4), 579–589. https://doi.org/10.1037/0278-7393.14.4.579.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Beentjes, J. W., & Van der Voort, T. H. (1991). Recall and language usage in retellings of televised and printed stories. *Poetics*, *20*, 91–104. https://doi.org/10.1016/0304-422x(91)90035-n.

Beentjes, J. W., & Van der Voort, T. H. (1993). Television viewing versus reading: Mental effort, retention, and inferential learning. *Communication Education*, *42*(3), 191–205. https://doi.org/10.1080/03634529309378927.

Bezdek, M. A., Foy, J. E., & Gerrig, R. J. (2013). Run for it! Viewers' participatory responses to film narratives. *Psychology of Aesthetics Creativity and the Arts*, *7*(4), 409–416. https://doi.org/10.1037/a0034083.

Cain, K., & Oakhill, J. (2006). Profiles of children with specific reading comprehension diffculties. *British Journal of Educational Psychology*, *76*(4), 683–696. https://doi.org/10.1348/000709905X67610.

Cain, K., Oakhill, J., & Bryant, P. (2000). Investigating the causes of reading comprehension failure: The comprehension-age match design. *Reading and Writing*, *12*, 31–40. https://doi.org/10.1023/A:1008058319399.

Carlson, S. E., Seipel, B., & McMaster, K. (2014). Development of a new reading comprehension assessment: Identifying comprehension differences among readers. *Learning and Individual Differences*, *32*, 40–53. https://doi.org/10.1016/j.lindif.2014.03.003.

Catts, H. W., Adlof, S., & Ellis-Weismer, S. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech Language and Hearing Research*, *49*, 278–293. https://doi.org/10.1044/1092-4388(2006/023).

CGMeetup (2018, February 12). *A Joy Story: Joy and Heron* [Video]. Youtube. https://www.youtube.com/watch?v=1lo-8UWhVcg.

Coderre, E. L. (2020). Dismantling the visual ease assumption: A review of visual narrative processing in clinical populations. *Topics in Cognitive Science*, *12*(1), 224–255. https://doi.org/10.1111/tops.12446.

Cohn, N. (2013). *The visual language of comics: Introduction to the structure and cognition of sequential images*. Bloomsbury Academic.

Cohn, N. (2020). *Who understands comics? Questioning the universality of visual language comprehension*. Bloomsbury Academic.

Dahl, A. C., Carlson, S. E., Renken, M., McCarthy, K. S., & Reynolds, E. (2021). Materials matter: An exploration of text complexity and its effects on middle school readers' comprehension processing. *Language Speech and Hearing Services in Schools*, *52*(2), 702–716. https://doi.org/10.1044/2021_LSHSS-20-00117.

Diergarten, A. K., & Nieding, G. (2015). Children's and adults' ability to build online emotional inferences during comprehension of audiovisual and auditory texts. *Journal of Cognition and Development*, *16*(2), 381–406. https://doi.org/10.1080/15248372.2013.848871.

Elbro, C., & Buch-Iversen, I. (2013). Activation of background knowledge for inference making: Effects on reading comprehension. *Scientific Studies of Reading*, *17*(6), 435–452. https://doi.org/10.1080/10888438.2013.774005.

Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis: *Verbal reports as data* (revised ed.). MIT Press.

Evers, G. (2008). *Programma Voor berekening Cito Lees Index Voor Het Basisonderwijs. P-CLIB versie 3.0*. Cito.

Fox, E. (2009). The role of reader characteristics in processing and learning from informational text. *Review of Educational Research*, *79*(1), 197–261. https://doi.org/10.3102/0034654308324654.

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage.

Furnham, A., De Siena, S., & Gunter, B. (2002). Children's and adults' recall of children's news stories in both print and audio-visual presentation modalities. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *16*(2), 191–210. https://doi.org/10.1002/acp.777.

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning Memory and Cognition*, *75*(3), 430–445. https://doi.org/10.21236/ada226788.

Goldman, S. R., & Varma, S. (1995). CAPping the construction-integration model of discourse comprehension. In C. A. Weaver, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 337–358). Lawrence Erlbaum Associates.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10. https://doi.org/10.1177/074193258600700104.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371–395. https://doi.org/10.1037/0033-295x.101.3.371.

Green, S. B., & Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, *74*, 169–173. https://doi.org/10.1007/s11336-008-9098-4.

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.

Helder, A., Van Leijenhorst, L., & van den Broek, P. (2016). Coherence monitoring by good and poor comprehenders in elementary school: Comparing offline and online measures. *Learning and Individual Differences*, *48*, 17–23. https://doi.org/10.1016/j.lindif.2016.02.008.

Hutson, J. P., Smith, T. J., Magliano, J. P., & Loschky, L. C. (2017). The effects of narrative comprehension and viewing task on gaze control in film. *Cognitive Research: Principles and Implications*, *2*(1), 1–46. https://doi.org/10.1186/s41235-017-0080-5.

Hutson, J. P., Magliano, J. P., & Loschky, L. C. (2018). Understanding moment-to-moment processing of visual narratives. *Cognitive Science*, *42*(8), 2999–3033. https://doi.org/10.1111/cogs.12699.

Karlsson, J., van den Broek, P., Helder, A., Hickendorff, M., Koornneef, A., & van Leijenhorst, L. (2018). Profiles of young readers: Evidence from thinking aloud while reading narrative and expository texts. *Learning and Individual Differences*, *67*, 105–116. https://doi.org/10.1016/j.lindif.2018.08.001.

Kendeou, P., Lynch, J. S., van den Broek, P., Espin, C. A., White, M. J., & Kremer, K. E. (2005). Developing successful readers: Building early comprehension skills through television viewing and listening. *Early Childhood Education Journal*, *33*, 91–98. https://doi.org/10.1007/s10643-005-0030-6.

Kendeou, P., Bohn-Gettler, C., White, M. J., & van den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading*, *31*(3), 259–272. https://doi.org/10.1111/j.1467-9817.2008.00370.x.

Kendeou, P., Smith, E. R., & O'Brien, E. J. (2013). Updating during reading comprehension: Why causality matters. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(3), 854. https://doi.org/10.1037/a0029468.

Kendeou, P., van den Broek, P., Helder, A., & Karlsson, J. (2014). A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research & Practice*, *29*(1), 10–16. https://doi.org/10.1111/ldrp.12025.

Kendeou, P., McMaster, K. L., Butterfuss, R., Kim, J., Bresina, B., & Wagner, K. (2020). The Inferential Language comprehension (iLC) framework: Supporting children's comprehension of visual narratives. *Topics in Cognitive Science*, *12*(1), 256–273. https://doi.org/10.1111/tops.12457.

Kim, Y. S. G., Dore, R., Cho, M., Golinkoff, R., & Amendum, S. J. (2021). Theory of mind, mental state talk, and discourse comprehension: Theory of mind process is more important for narrative comprehension than for informational text comprehension. *Journal of Experimental Child Psychology*, *209*, 105181. https://doi.org/10.1016/j.jecp.2021.105181.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163–182. https://doi.org/10.1037/0033-295x.95.2.163.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, *49*(4), 294–303. https://doi.org/10.1037/0003-066X.49.4.294.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. https://doi.org/10.1037/0033-295X.85.5.363.

Kraal, A., Koornneef, A. W., Saab, N., & van den Broek, P. W. (2018). Processing of expository and narrative texts by low-and high-comprehending children. *Reading and Writing*, *31*, 2017–2040. https://doi.org/10.1007/s11145-017-9789-2.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2.

Lenth, R. V. (2023). Emmeans: Estimated marginal means, aka least-squares means. Retrieved from https://cran.r-project.org/package=emmeans.

List, A. (2018). Strategies for comprehending and integrating texts and videos. *Learning and Instruction*, *57*, 34–46. https://doi.org/10.1016/j.learninstruc.2018.01.008.

Lynch, J. S., van den Broek, P., Kremer, K. E., Kendeou, P., White, M. J., & Lorch, E. P. (2008). The development of narrative comprehension and its relation to other early reading skills. *Reading Psychology*, *29*(4), 327–365. https://doi.org/10.1080/02702710802165416.

Magliano, J. P., Dijkstra, K., & Zwaan, R. A. (1996). Generating predictive inferences while viewing a movie. *Discourse Processes*, *22*(3), 199–224. https://doi.org/10.1080/01638539609544973.

Magliano, J. P., Millis, K., Ozuru, Y., & McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 107–136). Lawrence Erlbaum Associates.

Magliano, J. P., Loschky, L. C., Clinton, J. A., & Larson, A. M. (2013). Is reading the same as viewing? An exploration of the similarities and differences between processing text- and visually based narratives. In B. Miller, L. E. Cutting, & P. McCardle (Eds.), *Unraveling the behavioral, neurobiological, & Genetic Components of reading comprehension* (pp. 78–90). Brookes Publishing Co.

Magliano, J. P., Larson, A. M., Higgs, K., & Loschky, L. C. (2016). The relative roles of visuospatial and linguistic working memory systems in generating inferences during visual narrative comprehension. *Memory & Cognition*, *44*, 207–219. https://doi.org/10.3758/s13421-015-0558-7.

Magliano, J. P., Higgs, K., & Clinton, J. A. (2019). Sources of complexity in comprehension across modalities of narrative experience. In M. Grishokova, & M. Poulaki (Eds.), *Cognition and narrative complexity* (pp. 149–173). University of Nebraska.

Maine, F., & Shields, R. (2015). Developing reading comprehension with moving image narratives. *Cambridge Journal of Education*, *45*(4), 519–535. https://doi.org/10.1080/0305764X.2014.998625.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, *9*(1), 111–151. https://doi.org/10.1016/0010-0285(77)90006-8.

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed.). Cambridge University Press.

McMaster, K. L., van den Broek, P., Espin, C., White, M. J., Rapp, D. N., Kendeou, P., Bohn-Gettler, C. M., & Carlson, S. (2012). Making the right connections: Differential effects of reading intervention for subgroups of comprehenders. *Learning and Individual Differences*, *22*(1), 100–111. https://doi.org/10.1016/j.lindif.2011.11.017.

McMaster, K. L., Kendeou, P., Bresina, B., Slater, S., Wagner, K., White, M. J., Butterfuss, R., Kim, J., & Umana, C. (2019). Developing an interactive software application to support young children's inference-making. *L1-Educational Studies in Language and Literature*, *19*, 1–30. https://doi.org/10.17239/l1esll-2019.19.04.04.

McMaster, K. L., Kendeou, P., Kim, J., & Butterfuss, R. (2023). Efficacy of a technology-based early language comprehension intervention: A randomized control trial. *Journal of Learning Disabilities*, *57*(3), 139–152. https://doi.org/10.1177/00222194231182974.

McNamara, D. S., Newton, N., Christhilf, K., McCarthy, K. S., Magliano, J. P., & Allen, L., K (2023). Anchoring your bridge: The importance of paraphrasing to inference making in self-explanations. *Discourse Processes*, *60*, 337–362. https://doi.org/10.1080/0163853X.2023.2225757.

Oakhill, J. (2020). Four decades of research into children's reading comprehension: A personal review. *Discourse Processes*, *57*(5–6), 402–419. https://doi.org/10.1080/0163853x.2020.1740875.

Öncel, P., Hu, S., Ness-Maddox, H., Allen, L. K., & Magliano, J. P. (2024). Exploring the affordances of text and picture stories. *Discourse Processes*, *61*(4–5), 203–221. https://doi.org/10.1080/0163853X.2024.2352814

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, *76*, 241–263. https://doi.org/10.1037/h0027272.

Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly*, *38*(1), 36–76. https://doi.org/10.1598/rrq.38.1.3.

Perfetti, C. A. (1985). *Reading ability*. Oxford University Press.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357–383. https://doi.org/10.1080/10888430701530730.

Perfetti, C., & Helder, A. (2021). Incremental comprehension examined in event-related potentials: Word-to-text integration and structure building. *Discourse Processes*, *58*(1), 2–21. https://doi.org/10.1080/0163853x.2020.1743806.

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.

Salomon, G. (1984). Television is easy and print is tough: The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology*, *76*(4), 647. https://doi.org/10.1037/0022-0663.76.4.647.

Seifert, L. S. (1997). Activating representations in permanent memory: Different benefits for pictures and words. *Journal of Experimental Psychology: Learning Memory and Cognition*, *23*, 1106–1121. https://doi.org/10.1037/0278-7393.23.5.1106.

Seipel, B., Carlson, S. E., & Clinton, V. E. (2017). When do comprehender groups differ? A moment-by-moment analysis of think-aloud protocols of good and poor comprehenders. *Reading Psychology*, *38*(1), 39–70. https://doi.org/10.1080/02702711.2016.1216489.

Southeastern Guide Dogs (2018, October 24). *Pip* [Video]. Youtube. https://www.youtube.com/watch?v=07d2dXHYb94.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing* (pp. 53–120). Ablex.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd.

TheCGBros (2015, November 2). *Mouse for Sale* [Video]. Youtube. https://www.youtube.com/watch?v=OzYwE3Tst1Y.

Tibus, M., Heier, A., & Schwan, S. (2013). Do films make you learn? Inference processes in expository film comprehension. *Journal of Educational Psychology*, *105*(2), 329–340. https://doi.org/10.1037/a0030818.

Tomesen, M., Weekers, A., Hiddink, L., & Jolink, A. (2017). *Wetenschappelijke Verantwoording Begrijpend lezen 3.0 voor groep 6*. Cito.

Tomesen, M., Engelen, R., & Hiddink, L. (2018). *Wetenschappelijke Verantwoording Begrijpend lezen 3.0 voor groep 7*. Cito.

Tomesen, M., Engelen, R., & Hiddink, L. (2019). *Wetenschappelijke Verantwoording Begrijpend lezen 3.0 voor groep 8*. Cito.

Trabasso, T., & Magliano, J. P. (1996). Conscious understanding during comprehension. *Discourse Processes*, *21*(3), 255–287. https://doi.org/10.1080/01638539609544959.

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*(5), 612–630. https://doi.org/10.1016/0749-596X(85)90049-X.

van den Broek, P., & Helder, A. (2017). Cognitive processes in discourse comprehension: Passive processes, reader-initiated processes, and evolving mental representations. *Discourse Processes*, *54*(5–6), 360–372. https://doi.org/10.1080/0163853x.2017.1306677.

van den Broek, P., Lorch, E. P., & Thurlow, R. (1996). Children's and adults' memory for television stories: The role of causal factors, story-grammar categories, and hierarchical level. *Child Development*, *67*(6), 3010–3028. https://doi.org/10.2307/1131764.

van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading. In van H. Oostendorp, & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71–98). Lawrence Erlbaum Associates.

van den Broek, P., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory and Cognition*, *29*(8), 1081–1087. https://doi.org/10.3758/bf03206376.

Walma van der Molen, J. H., & Van der Voort, T. H. (1997). Children's recall of television and print news: A media comparison study. *Journal of Educational Psychology*, *89*(1), 82. https://doi.org/10.1037/0022-0663.89.1.82.

Walma van der Molen, J. H., & van der Voort, T. H. (1998). Children's recall of the news: TV news stories compared with three print versions. *Educational Technology Research and Development*, *46*(1), 39–52. https://doi.org/10.1007/bf02299828.

Walma Van Der Molen, J. H., & Van Der Voort, T. H. (2000). The impact of television, print, and audio on children's recall of the news. A study of three alternative explanations for the dual-coding hypothesis. *Human Communication Research*, *26*(1), 3–26. https://doi.org/10.1093/hcr/26.1.3.

Wannagat, W., Waizenegger, G., & Nieding, G. (2017). Multi-level mental representations of written, auditory, and audiovisual text in children and adults. *Cognitive Processing*, *18*(4), 491–504. https://doi.org/10.1007/s10339-017-0820-y.

Wolf, M. C., Muijselaar, M. M. L., Boonstra, A. M., & de Bree, E. H. (2019). The relationship between reading and listening comprehension: Shared and modality-specific components. *Reading and Writing*, *32*(7), 1747–1767. https://doi.org/10.1007/s11145-018-9924-8.

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

Authors and Affiliations

Dianne Venneker[1] · Anne Helder[1] · Paul van den Broek[1]

✉ Dianne Venneker
d.venneker@fsw.leidenuniv.nl

[1] Institute of Education and Child Studies, Leiden University, Wassenaarseweg 52, Leiden 2333 AK, The Netherlands