Children's comprehension processes and outcomes across different media formats: a think-aloud study comparing expository texts and videos

van Zeijts, B.E.J.; Ganushchak, L.Y.; Tabbers, H.K.; Hickendorff, M.; de Koning, B.B.

Routledge
Taylor & Francis Group

Discourse Processes

ISSN: 0163-853X (Print) 1532-6950 (Online) Journal homepage: www.tandfonline.com/journals/hdsp20

Children's comprehension processes and outcomes across different media formats: a think-aloud study comparing expository texts and videos

Brechtje E. J. van Zeijts, Lesya Y. Ganushchak, Huib K. Tabbers, Marian Hickendorff & Björn B. de Koning

To cite this article: Brechtje E. J. van Zeijts, Lesya Y. Ganushchak, Huib K. Tabbers, Marian Hickendorff & Björn B. de Koning (30 Jul 2025): Children's comprehension processes and outcomes across different media formats: a think-aloud study comparing expository texts and videos, Discourse Processes, DOI: 10.1080/0163853X.2025.2522639

To link to this article: https://doi.org/10.1080/0163853X.2025.2522639

Published online: 30 Jul 2025.

DISCOURSE PROCESSES
https://doi.org/10.1080/0163853X.2025.2522639

Routledge
Taylor & Francis Group

Children's comprehension processes and outcomes across different media formats: a think-aloud study comparing expository texts and videos

Brechtje E. J. van Zeijts[a], Lesya Y. Ganushchak[a], Huib K. Tabbers[a], Marian Hickendorff[b], and Björn B. de Koning[a]

[a]Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Rotterdam, Netherlands; [b]Institute of Education and Child Studies, Leiden University, Leiden, Netherlands

ABSTRACT

This study compared the comprehension processes and outcomes of 81 fourth-grade children when understanding expository texts and videos. All children were instructed to think-aloud while reading two texts and watching two videos, providing insight into their comprehension processes. Comprehension outcomes were measured with open-ended questions after each text and video. Results showed no significant differences in comprehension outcomes and underlying processes for text and video. Additional latent profile analyses identified four different comprehension profiles based on children's think-aloud responses, which were very similar for text and video. We found that 69% of children fell in the same profile in the text and video condition, suggesting that the majority of children tended to use a similar approach for understanding different media. To conclude, our findings provide empirical support for a *unitary process view* by showing that comprehension of our texts and videos was associated with similar cognitive processes.

Introduction

Understanding expository texts is an important skill for academic performance, work life, and when navigating society more broadly. However, many children fail to reach a sufficient level of reading comprehension by the end of primary education (Mullis et al., 2023; National Assessment of Educational Progress, 2022), and they are unmotivated to read and to practice with comprehension strategies (Swart et al., 2023). Instead, children spend more time online, with watching videos on YouTube being among their favorite activities (Dutch Media Literacy Network, 2021). Videos are also increasingly used in the classroom. This raises the question whether and how (instructional) videos can be used within reading education. Several scholars have suggested that children could practice their reading comprehension skills using other media than written text, for example, using video materials (e.g., Kendeou et al., 2020). An important theoretical assumption of this approach is that understanding text and video relies on similar comprehension processes, suggesting the existence of a general comprehension skill (see Gernsbacher et al., 1990). Yet, despite the popularity of videos, both in and outside of the classroom, there is little research on how children understand video materials compared to written text (List, 2018). Several studies compared comprehension outcomes of texts and videos (e.g., Wannagat et al., 2017), but less is known about the comparability of the

CONTACT Brechtje E. J. van Zeijts b.e.j.vanzeijts@uu.nl

An earlier version of this paper was presented at the ORD (the Dutch Educational Research Days) in Hasselt and at the Special Interest Group (SIG) 2 meeting of EARLI in Kiel.

underlying comprehension processes. The present study aimed to address this gap in the literature by directly comparing how children understand expository videos and texts, looking at comprehension processes as well as comprehension outcomes.

Theory on constructing comprehension from text and video

Reading comprehension requires the construction of a mental representation of what the text is about (McNamara & Maglino, 2009). The Construction-Integration Model posits that a reader can attain three different levels of comprehension: the surface level, the text base level, and the situation model level (Kintsch, 1988). A mental representation of a text at the surface level only contains the literal wording of a text, whereas a text base representation also contains the meaning of phrases and implicit relations between parts of information in the text. A situation model representation integrates the text base with the reader's prior knowledge. In contrast to the surface and text base levels, the situation model representation moves beyond the information stated in the text. This coherent mental representation includes pieces of information from the text and prior knowledge that are connected via implicit relations, also known as inferences (Cain & Oakhill, 1999).

The Construction-Integration Model was originally aimed at comprehension of written text, but several researchers have argued that it can be seen as a general model that extends to other media (List & Ballenger, 2019; Magliano et al., 2013; Wannagat et al., 2017). The notion that the same cognitive processes underlie comprehension of text, audio, and video is known as the *unitary process view* (Sinatra, 1990). This aligns with the theory of a *general comprehension skill* (Gernsbacher et al., 1990), stating that individuals rely on a shared set of comprehension skills for understanding different media formats. For example, making inferences is crucial for understanding written text but also for other forms of language such as narrated text (e.g., when watching a film, Graesser et al., 2001). This is also consistent with the Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990), which states that reading comprehension is based on decoding skills plus listening comprehension. Contrasting this perspective is the *dual process view* (see Sinatra, 1990), which states that the processes involved in comprehension depend on medium. Evidence for this view mainly comes from studies showing that a text format can result in better comprehension outcomes compared to an audio or video format (e.g., Salmerón et al., 2020; Verlaan et al., 2017).

These seemingly opposing views may be reconciled by distinguishing between higher-level processes and lower-level processes. Most researchers agree that understanding different media is similar in terms of higher-level cognitive processes such as inference generation (also called *back-end processes*, Magliano et al., 2013), which refers to how comprehension is constructed (Kendeou et al., 2014). However, regarding lower-level cognitive processes (or *front-end processes*), there are differences across media in how information is presented and encoded (Kendeou et al., 2014; Magliano et al., 2013). For example, an important modality-specific aspect is that reading requires text decoding (Hoover & Gough, 1990). This is especially challenging for beginning readers, whose basic reading skills are not yet automated. Thus, understanding written texts presents learners with greater linguistic demands compared to understanding videos. On the other hand, a challenging aspect of understanding videos is that the information is transient and presented at a certain pace that usually cannot be controlled by the learner (Magliano et al., 2013). This may impose additional processing demands and prevent learners from actively processing and elaborating on the incoming information (Fyfield et al., 2022). Moreover, with videos, content often is simultaneously presented via the visual and auditory channel (Mayer, 2005). The amount of visual information may differ per video, but even the social presence of an instructor (e.g., when watching a video lecture) can influence the comprehension process, for example, by increasing motivation or by distracting from important information (Beege et al., 2023). These differences in how the information is presented and encoded across different media formats are important to consider, as they could influence comprehension processes and outcomes.

Comparing comprehension outcomes for text and video

Several empirical studies have investigated the relation between children's understanding of different media, but, thus far, these studies mainly focused on outcome measures (e.g., Verlaan et al., 2017; Wannagat et al., 2017). Overall, these studies indicate that children's reading comprehension skills, listening comprehension skills, and audiovisual comprehension skills are strongly related. This relation strengthens with age, as decoding skills develop and become more automated (Diakidoy et al., 2005). Moreover, longitudinal research showed that children's television and listening comprehension skills in preschool and the early grades of primary school were predictive of their reading comprehension in later grades (Catts et al., 2015; Kendeou et al., 2008). Hence, children who are well able to understand spoken language at a young age are typically also more skilled in understanding written language at a later age. Similarly, children's ability to draw inferences in a listening context or when watching television is predictive of their inference skills when reading (Kendeou et al., 2008). These findings support the idea of a general comprehension skill as they suggest that the skills that are involved in understanding spoken language are also relevant for understanding written text. This provides indirect evidence that there may be similarities between media in terms of cognitive processes. Nevertheless, similar comprehension outcomes may result from very different underlying processes. Therefore, this prior research does not provide information on what actually happened *during* comprehension, because it did not use any process measures.

Complementing comprehension outcomes with process measures

In the field of text comprehension, steps were taken to provide insight into comprehension processes by complementing outcome measures with process measures such as eye-tracking (e.g., Kraal et al., 2019) and think-aloud protocols (e.g., Denton et al., 2015). However, research investigating the cognitive processes associated with understanding films or videos is scarce. We found only a few studies that attempted to measure comprehension processes for audiovisual materials. Magliano et al. (2001) studied how students comprehend events when watching a narrative film. They showed that the way students perceive changes in situation, such as a new location or shift in time, is consistent with findings in the field of text comprehension. Other studies focused on whether viewers generate specific types of inferences while watching films, such as emotional inferences (Diergarten & Nieding, 2015), predictive inferences (Magliano et al., 1996, 2001), and local bridging inferences (Tibus et al., 2013). Their findings demonstrated that viewers do generate such inferences while watching videos, which indicates that viewers actively elaborate on the incoming information. However, it remains unknown how these findings relate to inference generation and other comprehension processes if the same information is presented in a textual format, as these studies did not make a direct text-to-video comparison.

Relatively few studies have directly compared the processes underlying comprehension of different media (e.g., List & Ballenger, 2019; Tibus et al., 2013). The few studies that did make a direct comparison between text and other media often used retrospective measures to capture comprehension processes. For example, List and colleagues asked students to report their strategy use after comprehending expository videos and texts (Lee & List, 2019; List, 2018; List & Ballenger, 2019). The results were mixed, with List (2018) reporting no differences in the prevalence of strategy use across media, but also finding some medium-specific strategies (e.g., rereading, highlighting content during reading, or looking at visuals presented in the video). Also, Lee and List (2019) and List and Ballenger (2019) found that students used more higher-level strategies in the text condition compared to the video condition, possibly due to higher familiarity with this medium for educational purposes. In interpreting these findings, it important to note that it may have been difficult for students to report in retrospect what strategies they used *during* comprehension. Other studies compared how students learn from digital expository text and videos and retrospectively measured cognitive processes, such as attention to the task and metacognitive calibration (Delgado et al., 2022; Mason et al., 2022; Tarchi et al.,

2021). These studies reported no differences between text and video in terms of students' perceived on-task attention (Delgado et al., 2022) and in how well students were able to estimate their own comprehension (Delgado et al., 2022; Mason et al., 2022; Tarchi et al., 2021). These findings suggest that students' comprehension processes are comparable across different media formats.

Importance of online process measures

Thus far, studies using retrospective measures have provided valuable information on how videos are understood in relation to texts, but given their disadvantages (e.g., error prone), *online* process measures are needed to gain more insight into the comparability of comprehension processes for text and video. Online measures are less dependent on the learner's memory and judgment and better able to capture what happens *during* comprehension. To our knowledge, only two studies comparing comprehension processes across media formats have used online process measures. Diergarten and Nieding (2016) used a reaction-time paradigm to compare how children and adults generate emotional inferences while reading and listening to narrative texts. They found that fourth graders and adults generated more emotional inferences during reading than during listening, whereas the inference processes of second graders were not influenced by modality. The researchers speculated that the fourth-grade children and adults might have considered reading as more difficult than listening, therefore investing more effort in generating inferences. It is, however, good to note that Diergarten and Nieding focused specifically on narratives and the generation of emotional inferences (i.e., inferences about the protagonist's mental state). The generation of emotional inferences heavily depends on children's theory of mind and emotional knowledge (Diergarten & Nieding, 2015; Gernsbacher & Robertson, 1992), so these findings may not generalize to other inference types or comprehension processes.

Neuman (1992) looked at a broader range of inferences using a think-aloud protocol, such as reiterating information, binding story elements together, or emphasizing with characters. Fifth graders were instructed to share their thoughts while watching a television series and while reading a narrative text. In contrast to Diergarten and Nieding, the results of Neuman showed no differences in the number of inferences that children generated when understanding narratives in textual and video format. They concluded that the medium is merely a conveyer of content and that the process of constructing understanding is similar. All in all, the exact relation between different media may depend on inference type (Diergarten & Nieding, 2016; Neuman, 1992), but for now it remains inconclusive to what extent comprehension processes across different media are comparable.

We contribute to this research using online process measures (Diergarten & Nieding, 2016; Neuman, 1992) by making a comparison between *expository* instead of narrative texts and videos. In the later grades of primary school, the focus shifts toward expository materials (Best et al., 2008), with the goal of learning from texts, so for this age group it would be relevant to compare comprehension of expository texts and videos. Expository texts follow a different structure than narratives and different competencies contribute to understanding expository versus narrative materials (Best et al., 2008). Because emotional inferences do not frequently occur in expository texts, we focus on the results of Neuman (1992) and studies using retrospective measures (e.g., Delgado et al., 2022; Lee & List, 2019). Given the described differences between narrative and expository texts, it is of relevance to investigate whether these findings generalize to the expository genre. Moreover, whereas previous research was largely focused on inferencing (Diergarten & Nieding, 2016; Neuman, 1992), we aimed to look at a broader range of comprehension processes. Besides inference generation, children also engage in text repetitions, comprehension monitoring, and elaborations based on personal experiences when constructing an understanding of texts (as studied in think-aloud research; see Karlsson et al., 2018; Kraal et al., 2018). This makes it relevant to also include these processes in our text-to-video comparison.

Present study

In the present study, we compared fourth-grade children's comprehension of expository texts and videos, by measuring both comprehension processes and comprehension outcomes. Four expository texts were selected from educational practice, and four matching talking-head videos were created. These videos centered around a speaker who presented the written text in narrated form, ensuring content equivalence between the texts and videos (cf, Tversky et al., 2002). All children were presented with two texts and two videos. To gain more insight into their comprehension processes, we used an *online* process measure that is feasible in both a reading and video context, namely, a think-aloud protocol (see Karlsson et al., 2018; Neuman, 1992). In doing so, we tried to conceptually replicate the direct text-to-video comparison from previous research (e.g., Lee & List, 2019), now using a methodology better equipped to assess comprehension processes as they occur. Children were instructed to verbalize their thoughts while reading the texts and while watching the videos. In addition, we measured children's comprehension outcomes using an offline measure, that is, by asking open-ended comprehension questions targeting important implicit connections after each text and video. We formulated the following research questions:

(1) Are comprehension outcomes in the text and video condition comparable, as demonstrated by children's performance on the comprehension questions?
(2) Are comprehension processes in the text and video condition comparable, as demonstrated by children's think-aloud responses?
    (2a) At the group level, we tested comparability of comprehension processes for text and video by looking at whether the average number of responses in the text and video condition was comparable across a range of different think-aloud categories.
    (2b) At an individual level, we explored whether individual participants utilize the same approach (i.e., engage in the same cognitive processes) for understanding the texts and videos by identifying different comprehension profiles.

Based on these research questions, we formulated hypotheses which were preregistered at the Open Science Framework (https://osf.io/zxn9j).[1] Regarding the first research question, we expected comprehension outcomes in the text and video condition to be comparable, as indicated by children's scores on the open-ended comprehension questions (Hypothesis 1). We tested children in the later years of primary school, so we expected decoding skills to be automated and therefore no differences in the level of understanding of text and video (e.g., Van Zeijts et al., 2023; Verlaan et al., 2017). For the second research question, we analyzed children's think-aloud protocols. We hypothesized children's comprehension processes when reading texts and when watching videos to be comparable, as indicated by a similar number of think-aloud responses in the text and video condition at the group level (Hypothesis 2a). Despite differences in genre and characteristics of the videos, we based our expectation on the available empirical research using an online process measure to compare inference processes across media (Neuman, 1992), and on theory stating that understanding different media formats is similar in terms of *higher-level* processes (Kendeou et al., 2014; Magliano et al., 2013). We expected the influence of *lower-level* differences between media to be limited, since the text decoding skills of fourth-grade children are largely automated, and our texts and videos were equivalent in terms of content. Finally, we did not formulate hypotheses for our analysis of comprehension processes at an individual level (Hypothesis 2b). We decided to explore individual differences in how children approach expository texts and videos using Latent Profile Analysis (LPA), complementing our analysis based on group averages. Previous studies successfully used LPA to identify different profiles of comprehension processes when reading narratives and expository texts (e.g., Karlsson et al., 2018). This exploratory analysis allowed us to examine whether individual children demonstrate a similar pattern of comprehension processes, as indicated by their comprehension profile, in a text and video context.

Methods

Participants and design

Participants were 89 fourth-grade children from three elementary schools in the Netherlands. All children received written informed consent to participate from their parents or caretakers. Data of eight children were excluded due to technical problems during the experiment ($n = 4$) or because there was more than a week between sessions due to a COVID-19 lock-down ($n = 4$). This resulted in a final sample of 81 children (40 boys and 41 girls), aged between 8 and 11 years old ($M = 9.95$, $SD = 0.51$). This sample size allowed us to detect a small to medium effect of media format (text vs. video) on the number of think-aloud responses (repeated-measures analysis of variance [ANOVA], within-subjects effect, $\alpha = .05$, power = 0.80, $\eta_p^2 = 0.03$) and a small to medium effect of media format (text vs. video) on comprehension outcomes (paired-samples two tailed t-test, $\alpha = .05$, power = 0.80, $dz = 0.32$). The participants' skill scores on a standardized test of reading comprehension ranged from 129 to 248 ($M = 172.12$, $SD = 23.52$), indicating variation in reading proficiency within our sample. The mean score was representative of the average for Dutch children (see Tomesen et al., 2017). Most children (84%) were native speakers of Dutch, whereas the other 16% had age-appropriate proficiency in Dutch but indicated to speak a different language with their parents and/or siblings at home.

A within-subjects design was used with media format (text vs. video) as independent variable and comprehension processes and comprehension outcomes as outcome variables (see further details in Materials, below). All children read two texts and watched two videos. They were all presented with the same four topics (see Texts and videos, below), shown in counterbalanced order using a balanced Latin square. The experiment consisted of two sessions, with one text and one video per session. Half of the children started their sessions with reading a text followed by a video, and the other half started with watching a video followed by a text.

Materials

All texts, videos, and comprehension questions have been made available at the Open Science Framework (https://osf.io/cxk36/).

Texts and videos

Four expository texts and four content-matching videos were used. The texts were selected from the archive of *Nieuwsbegrip* (in English: News Comprehension), which is the most frequently used Dutch instructional method for reading comprehension. The texts covered the topics of 3D-printers, a meteorite found in the Netherlands, sperm whales that washed ashore, and the Zika virus in Central and South America. These topics were selected based on two pilot studies. In a first pilot study ($n = 12$), we tested 16 expository texts and selected 11 topics that fourth graders indicated to have little prior knowledge about. Next, we did a second pilot study ($n = 20$) and selected four topics for inclusion in the present study based on the quality and difficulty of the comprehension questions (see Comprehension questions, below). The selected four texts were between 340 and 432 words long and consisted of five to six paragraphs. Each paragraph started with a subheading showing the structure of the text. Moreover, all texts included two images, with one image depicting the main topic of the text (e.g., a 3D-printer) and one image showing something discussed in a specific paragraph (e.g., a robot arm that was made by a 3D-printer). The texts were written at the level of fourth or fifth grade in terms of decoding and comprehension difficulty, as indicated by P-CLIB software (Evers, 2008).

Based on the four texts, we created matching talking-head videos (Figure 1) that were centered around a speaker presenting the same information from the written text in narrated form. The videos were presented by three journalism students and one student of a teacher training program (two men and two women) who were all native speakers of Dutch. The four presenters had been equally positively evaluated by children in our pilot study. The duration of the videos varied from 3 minutes

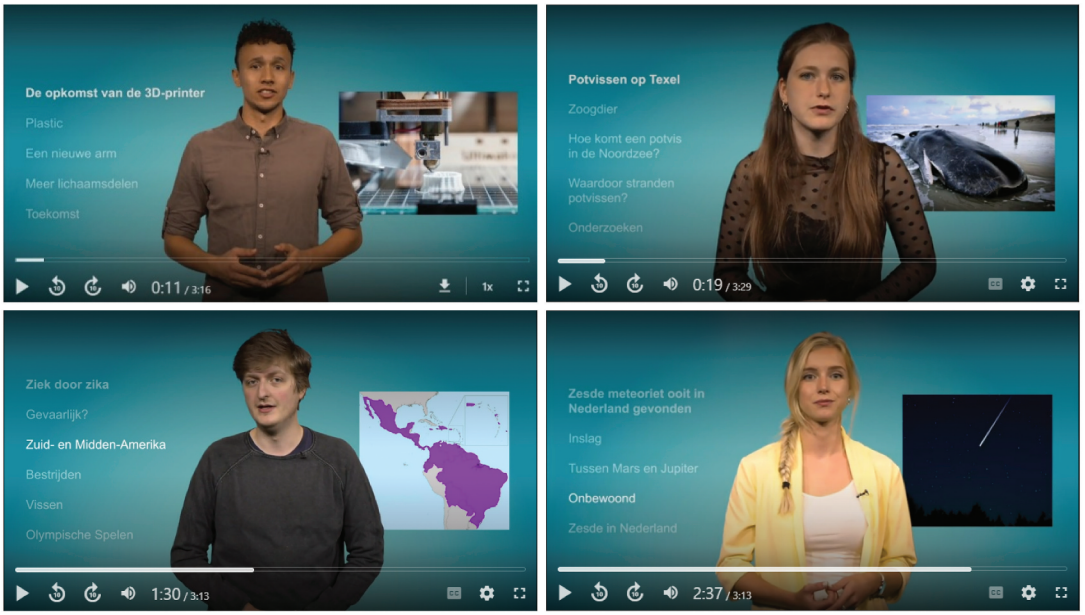

Figure 1. Screenshots of the four videos showing the presenter, subheadings, and a picture from the text.

and 13 seconds to 3 minutes and 29 seconds. When watching the videos, children were given the option to pause the video, rewind, and fast forward, as they could also read back or look ahead when reading a text. We created videos that were very comparable to the written text versions in terms of information that was being presented. This was done to avoid potential differences in comprehension between media caused by additional (visual) information in the video. Specifically, we ensured that (1) the videos did not contain any visual information that was not present in the text, except for the speaker, and (2) the subheadings that were in the texts were also added to the videos to visually display the structure of the information. When a certain paragraph was being presented, the corresponding subheading lit up, and the pictures presented in the texts were also shown in the videos, for the duration of one paragraph.

In addition, for practice purposes, two additional texts were selected from the *Nieuwsbegrip* archive. The first practice text was about the discovery of bones of a prehistoric man (337 words; 6 paragraphs), and the second practice text was about a plant called the giant hogweed (160 words; 3 paragraphs). We created two matching talking-head videos which were 2 minutes and 19 seconds, and 1 minute and 13 seconds in length, respectively.

Comprehension questions

Comprehension outcomes were measured using six open-ended questions. Of these six questions, four questions targeted a text-connecting inference and two questions targeted a gap-filling inference (Cain & Oakhill, 1999). Text-connecting inference questions were targeting an implicit relation between two pieces of information from the text or video. Gap-filling inference questions required children to combine information from the text or video with their prior knowledge. For the practice text or video about the discovery of bones of a prehistoric man, two text-connecting inference questions were used. There were no questions about the practice text or video about the giant hogweed, as this topic was only meant to rehearse the think-aloud procedure (see Think-aloud protocol, below). Questions were displayed on the computer screen and read to the children by the test leader. Questions were presented one at the time, and children could not look back at the text or video when answering the questions. Children answered verbally and did not receive feedback on their answers, and only neutral

encouragements were given (e.g., nodding). When children did not respond, the question was repeated once. If still no response was given, the child was instructed to continue with the next question.

Children's verbal responses were audio-recorded. All responses were transcribed verbatim and scored as correct (1 point), half correct (0.5 points), or incorrect (0 points) by four independent raters, using an answer model stating the key ideas. For example, one text-connecting question about meteorites was "What causes a falling star?" The correct answer was that meteorites fall down on earth (with very high speed) causing them to catch fire. A half correct answer could be only stating that meteorites are on fire. An incorrect answer could be that a star simply falls down. With six questions per topic, children could obtain a maximum of six points. For each participant, we calculated an average comprehension score for text and video based on two texts and two videos, respectively. A randomly selected subsample of participants was scored by all four raters (approximately 25% of all participants, $n = 25$). The inter-rater reliability was good (Krippendorff's $\alpha = .89$, ranging between $\alpha = .86$ and $\alpha = .94$ per topic; Hayes & Krippendorff, 2007).

Procedure

First, a paper-and-pencil questionnaire was administered in the classroom to measure children's attitudes toward reading and watching videos. It also contained questions on how often they read and watch videos in their free time. These questionnaire data are beyond the scope of this article and therefore are not reported. Next, all children were tested individually by one of three test administrators (first author and two trained undergraduate students). Children were tested in a separate room in the school. This individual phase was divided over two sessions of approximately 30 minutes, with 1 week between sessions. Children were seated in front of a laptop on which all materials were presented. In both sessions, they were presented with one experimental text and one experimental video. The first session started with questions about gender, age, and native language. Next, the think-aloud protocol was explained and modeled by the test leader (see below). Children then practiced thinking aloud and answering open-ended comprehension questions using a practice text and a practice video. In the subsequent experimental phase, all children were presented with two texts and videos. After each text or video, children were asked several questions about how much they had enjoyed reading the text or watching the video, the perceived difficulty, and their prior knowledge regarding the topic. These questions were answered on a six-point Likert scale ranging from 1 (*very little*) to 6 (*very much*). Next, they were asked open-ended comprehension questions after each text and video.

Think-aloud protocol

A think-aloud protocol was used to measure children's comprehension processes when reading texts and watching videos. As fewer prompts may encourage integration of ideas, children were prompted to think-aloud after every paragraph instead of after every sentence (Caldwell & Leslie, 2010; Denton et al., 2015). In the text condition, children were instructed to read the text aloud and click to the next page when they finished reading a paragraph. After each paragraph, three thoughts balloons were shown. When being presented with this prompt, children were asked to share as much as possible about what they were thinking. It was emphasized that there were no correct or incorrect responses. Similarly, in the video condition, a screen with three thought balloons was shown at the end of every paragraph (Figure 2). Children were instructed to pause the video, so they would have enough time to share their thoughts. Children were also allowed to share their thoughts at other points in the text or video, by pausing their reading or the video.

Based on a pilot study ($n = 12$), we introduced a practice phase to familiarize children with thinking-aloud and to make them comfortable with verbalizing their thoughts for the test leader. In the first individual session, children received a practice text or video about a prehistoric man. The first half of this practice text or video was used for modeling by the test leader. To maintain consistency across participants, the test leader modeled the think-aloud procedure following a script. This script included examples of paraphrases, text-connecting and gap-filling inferences, meta-cognitive

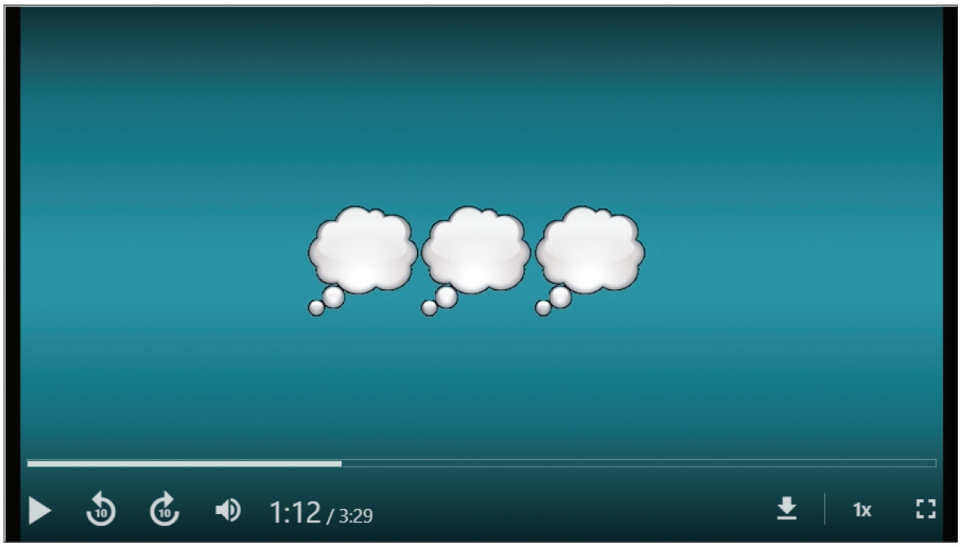

Figure 2. Screenshot of thought balloons in the video about sperm whales.

comments, and evaluations. Next, children practiced thinking aloud for the second half of the text or video. After practicing, children were presented with the first experimental topic, either as text or video. For all children, the second experimental topic was presented in a different modality (depending on counterbalancing it was either text or video). Therefore, they again practiced the think-aloud procedure but with a shorter practice text or video about the giant hogweed. In the second individual session, the instruction about the think-aloud task was repeated by the test leader. Children received encouragements from the test leader when practicing, with the primary aim to stimulate their talking. However, in the experimental phase, no feedback or content-related comments were given, as this can influence children's thought processes (e.g., Pressley & Afflerbach, 2012). When children did not respond to a prompt or indicated to have no thoughts, one neutral encouragement was given (e.g., "can you tell me what you were thinking?"). If still no response was given, children were instructed to continue reading or watching the video.

Children's responses were audio-recorded and afterward transcribed verbatim. Next, responses were parsed into verb–subject clauses, also called idea units (Trabasso & van den Broek, 1985). The idea units were categorized using a coding scheme (Table 1) by three independent raters (first author and two undergraduate students). The coding of the three raters was calibrated over three 1.5-hour training sessions. Our coding scheme was based on previous research using think-aloud protocols to measure comprehension processes during reading (e.g., Kraal et al., 2018; Linderholm & van den Broek, 2002; McMaster et al., 2012) and was adjusted to fit the texts and videos in our study. We refined the coding scheme after each training session (e.g., by merging or adding subcategories or improving category descriptions) until the inter-relater reliability was sufficient. Literal responses and inferences were coded as either correct or incorrect. A response was coded incorrect only when it contradicted with information given in the text or video or when incorrect prior knowledge had been used. As only 3.58% of these responses were coded as incorrect, it was decided to exclude incorrect responses from the analyses. After the training sessions, a randomly selected subset of transcripts not used for training and calibration (approximately 10% of all transcripts, $n = 36$) was coded by all three raters. This resulted in an inter-rater reliability of Krippendorff's $\alpha = .76$ (Hayes & Krippendorff, 2007), which is considered a sufficient reliability. For each participant, we calculated the average number of think-aloud responses in the text and video condition based on two texts and two videos, respectively.

Table 1. Description and examples of the think-aloud coding categories.

Category	Description	Examples
Literal information	Repeating or paraphrasing information literally mentioned in the text or video	"A 3D-printer costs hundreds of euros." "The people in Brazil were very worried."
Inferences	Making a connection between parts of information from the text or video, or integrating information from the text or video with background knowledge	"The whole group swam to the North Sea." "The meteorite caught fire due to the high speed."
Meta-cognitive comments	Reflecting on one's own (lack of) understanding or attempting to repair understanding, either with or without including information from the text or video	"I did not fully understand this part." "But why didn't the mosquitos fly to Europe?"
Personal comments	Adding information based on personal experiences, giving one's opinion, or sharing prior knowledge (which is not necessary for coherence)	"I feel sorry for Faith losing her arm." "I've never seen a meteorite myself." "I know sperm whales are very rare."
Irrelevant comments	Comments that are unrelated to the content of the text or video	"The image is a bit small." "I don't have anything to add."

Results

Our dataset and analysis scripts are available at the Open Science Framework (https://osf.io/cxk36/). The participants' prior knowledge varied across the four topics, $F(1, 80) = 27.69$, $p < .001$, $\eta p^2 = .45$. The lowest prior knowledge was reported for Zika virus ($M = 1.84$, $SD = 1.44$) and the highest for 3D-printers ($M = 3.44$, $SD = 1.60$). These variations were anticipated and resolved through counter-balancing, as described in the Method section.

Preregistered analyses

RQ1: comprehension outcomes

For our first research question, we compared children's comprehension scores in the text and video condition. A paired-samples *t*-test was performed with media format (text vs. video) as independent variable and children's comprehension score as outcome variable. To quantify the evidence in favor of the null hypothesis, we also performed a Bayesian *t*-test. In line with Hypothesis 1, there was no significant difference between comprehension scores in the text condition ($M = 3.48$, $SD = 1.15$) and video condition ($M = 3.56$, $SD = 1.13$), $t(80) = 0.69$, $p = .492$, $d = 0.08$. This was confirmed by the Bayes factor, which provided moderate evidence in favor of the null hypothesis, $BF01 = 6.49$.

RQ2a: comprehension processes at the group level

For the second research question, we compared children's comprehension processes in the text and video condition by analyzing their think-aloud responses. Figure 3 displays the average number of think-aloud responses in the text and video condition per category. We performed a repeated-measures ANOVA with media format (text vs. video) and think-aloud category (literal information vs. inferences vs. meta-cognitive comments vs. personal comments vs. irrelevant comments[2]) as within-subject factors. The dependent variable was the absolute number (i.e., counts) of think-aloud responses. Again, we also performed a Bayesian ANOVA. Consistent with Hypothesis 2a, there was no significant main effect of media format, $F(1, 80) = 1.22$, $p = .273$, $\eta_p^2 = 0.02$, which means that the average number of responses in the text condition and video condition did not significantly differ. In line with this, the Bayes factor provided medium evidence in favor of the null hypothesis, $BF01 = 6.44$. This factor indicates that our data are more likely under the null model than under the model with media format as predictor. We did find a significant main effect of think-aloud category, $F(4, 320) = 3.11$, $p = .016$, $\eta_p^2 = 0.04$, $BF10 = 2.51$. This merely shows that the number of responses differed across think-aloud categories.

For the interaction between media format and think-aloud category, the assumption of sphericity was violated ($\varepsilon = .89$, $p = .012$). Therefore, the Greenhouse-Geisser corrected results are reported. Again, as hypothesized, there was no significant interaction effect, $F(4, 320) = 0.263$, $p = .883$ $\eta_p^2 < 0.01$. Here, the Bayes factor provided very strong evidence against the interaction effect, $BF01 = 144.52$.

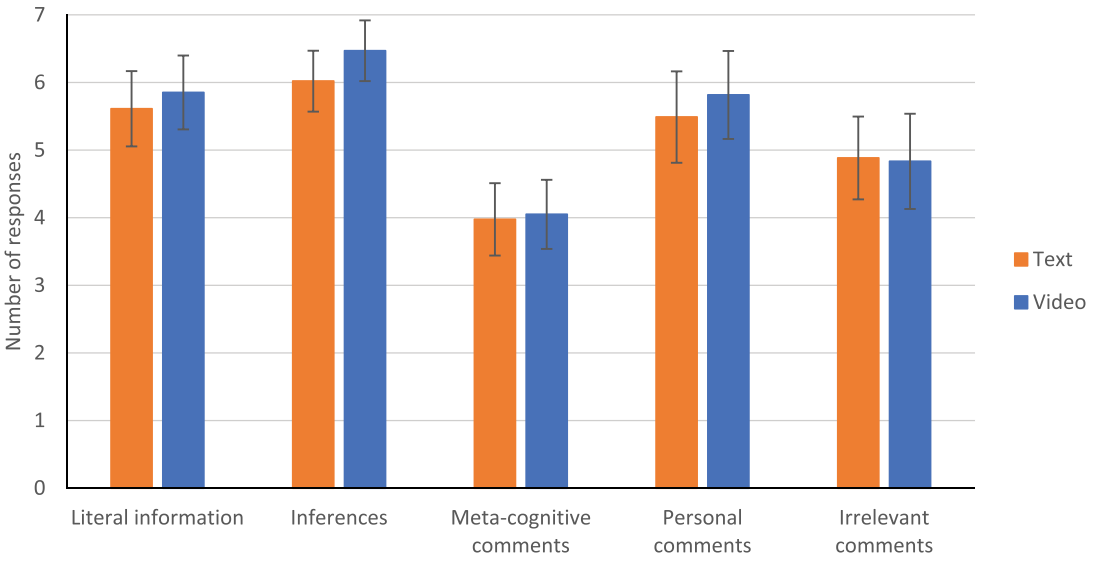

Figure 3. Absolute number of think-aloud responses for text and video per category.

This indicates that our data are far more likely under the two main effects model than under the model that adds the interaction term.

Exploratory analyses

RQ2b: comprehension processes at an individual level

Our analyses based on group averages did not reveal any differences between text and video in terms of comprehension processes. However, only focusing on group means may obscure unobserved heterogeneity. For example, differences between individual participants may average each other out resulting in no differences between text and video at the group level. In both the text and video condition, the standard deviations were high in all five think-aloud categories. This suggests that there are indeed individual differences in how children proceed in understanding texts and videos, which could be quantitative and/or qualitative in nature. We aimed to explore this heterogeneity in the data using LPA. This statistical tool enables tracing back the heterogeneity in the data to more homogeneous subgroups (e.g., Hickendorff et al., 2018) and has been successfully used in previous studies to identify subgroups of children with different profiles of comprehension processes (e.g., Karlsson et al., 2018).

The LPAs were conducted in Latent GOLD version 6.0 (Vermunt & Magidson, 2021). We conducted separate LPAs for the text and video condition. The absolute number of responses (i.e., counts) in the following think-aloud categories were included as indicators: literal responses, inferences, meta-cognitive comments, personal comments, and irrelevant comments. These categories were consistent with the preregistered repeated-measures ANOVA described above. The analyses were carried out with 1,000 random starts to avoid a locally optimal solution. The optimal number of classes was selected based on the Bayesian Information Criterion (BIC) and the Consistent Akaike Information Criterion (CAIC; Nylund et al., 2007) and substantial appeal of the solution. The model fit was evaluated by the entropy values and classification error.

Latent profiles for text. Models with one to eight latent clusters were estimated (Appendix, Table A1). The model with four clusters had the lowest BIC and CAIC values. This model had an R^2 entropy of 0.87, indicating that the profiles described the data well (Nylund-Gibson & Choi, 2018). The

classification error was 0.05, which means that children's cluster membership could be predicted with high certainty.

The four clusters are visualized in Figure 4. The children in the largest cluster (36% of the children) stayed very close to the information as stated in the text. They also generated some inferences, but these children primarily repeated or paraphrased information that was literally stated in the text. We therefore labeled this cluster as *Paraphrasers*. The second largest cluster (32% of the children) made few paraphrases and inferences compared to the other clusters, but they scored relatively high on monitoring comments. As most of their comments were reflections on their own understanding of the text, we labeled this group as *Monitoring readers*. The third cluster (27% of the children) was characterized by a relatively higher number of inferences and personal comments. This indicates that these children went beyond the literal information that was stated in the text and used their own background knowledge and experiences when processing the text. We therefore labeled this cluster as *Elaborators*. Finally, a very small cluster (4% of the children) scored high on almost all categories. Besides a very high number of irrelevant comments, these children also often reflected on their comprehension (i.e., monitoring comments) and generated many inferences. This indicates that they went beyond the literal information stated in the text and used their background knowledge. Due to this combination of comments that represent a higher-level understanding and many irrelevant responses, we labeled this cluster as *Elaborating chatters*.

Next, we tested for differences in comprehension outcomes between the four text comprehension profiles. To do this, children were assigned to the cluster for which they had the highest posterior probabilities. A one-way ANOVA showed no significant differences in comprehension outcomes between *Paraphrasers* ($M = 3.25$, $SD = 0.94$, $n = 30$), *Monitoring readers* ($M = 3.54$, $SD = 1.36$, $n = 28$), *Elaborators* ($M = 3.69$, $SD = 1.20$, $n = 20$), and *Elaborating chatters* ($M = 3.75$, $SD = 0.43$, $n = 3$), $F(3, 77) = 0.70$, $p = .558$, $\eta_p^2 = 0.03$.

Latent profiles for video. Again, models with one to eight latent clusters were estimated (Appendix, Table A2). The model with four clusters had the lowest BIC and CAIC values. The R^2 entropy of this model was 0.91 and the classification error was 0.05.

The four clusters are visualized in Figure 5. The largest cluster (32% of the children) can be characterized as *Paraphrasers*. As in the text condition, these children stayed very close to the information as mentioned in the video. The second largest cluster (29% of the children) was

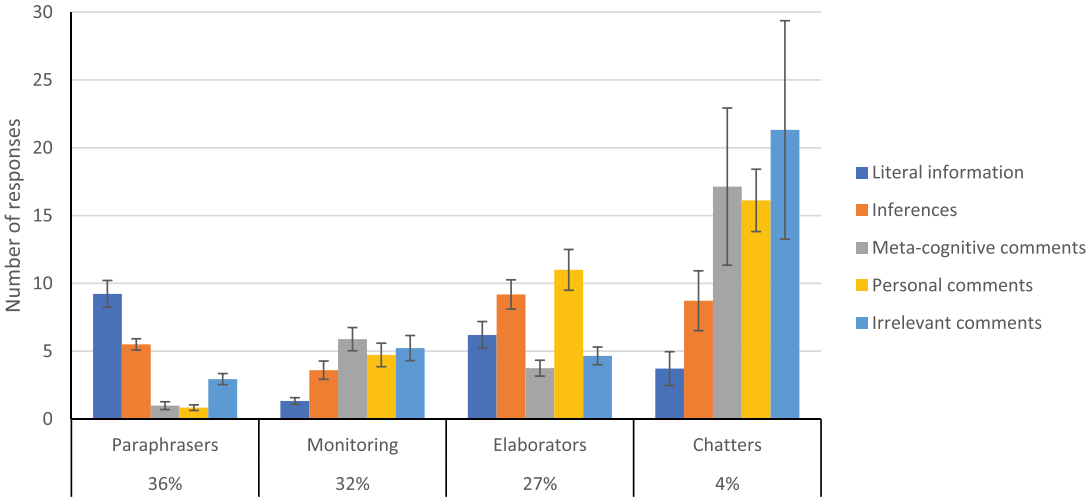

Figure 4. Profiles (with percentages of children) based on absolute numbers of think-aloud responses in the text condition.

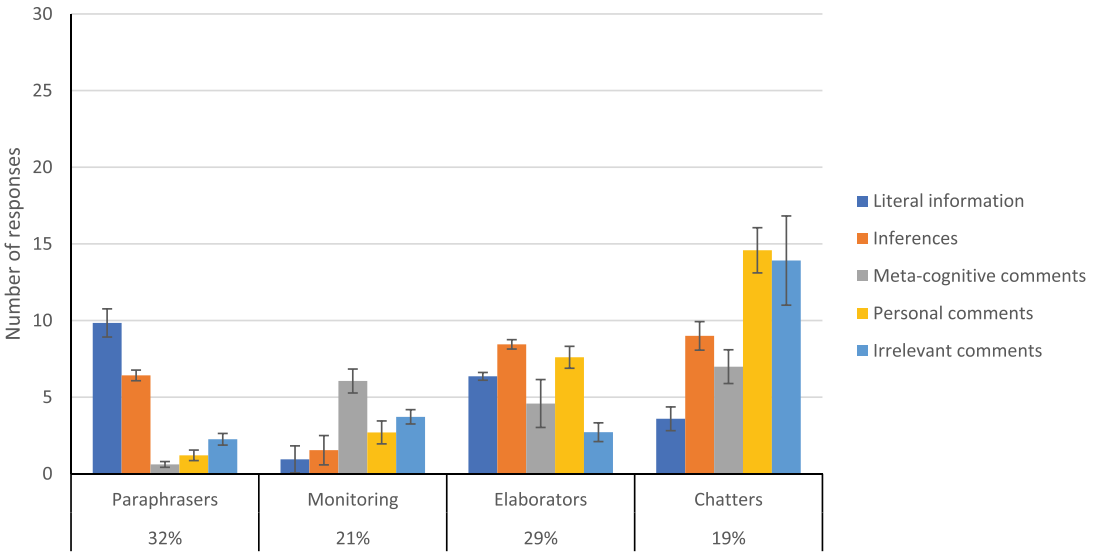

Figure 5. Profiles (with percentages of children) based on absolute numbers of think-aloud responses in the video condition.

labeled as *Elaborators*. This cluster was comparable to the cluster in the text condition, with a focus on inference generation and personal comments. The number of personal comments is somewhat lower compared to the text condition, but still relatively high. The third cluster (21% of the children) was characterized by *Monitoring*. These children focused on reflecting on their own understanding when watching the videos and made few responses in the other categories. The peak in responses in the monitoring category was more extreme compared to the Monitoring readers in the text condition. The smallest cluster (19% of the children) can be characterized as *Elaborating chatters*. This group scored high on most categories, but compared to the text condition, the number of monitoring comments was much lower. Also, the absolute number of personal comments and irrelevant comments was lower.

Next, we compared the four video comprehension profiles on comprehension outcomes using a one-way ANOVA and found a significant difference, $F(3, 81) = 3.51$, $p = .019$, $\eta_p^2 = 0.12$. Post hoc testing using Fisher's Least Significant Difference test showed that *Paraphrasers* ($M = 3.08$, $SD = 0.96$, $n = 27$) had significantly lower comprehension scores than *Elaborators* ($M = 3.95$, $SD = 1.08$, $n = 23$), $p = .006$, and *Elaborating chatters* ($M = 3.98$, $SD = 0.74$, $n = 14$), $p = .014$. Children in the *Monitoring* profile ($M = 3.43$, $SD = 1.45$, $n = 17$) did not significantly differ from the other profiles (all $ps > .05$).

Comparing the clusters for text and video. To examine whether individual children used the same approach for understanding our texts and videos (RQ2b), we compared their cluster membership in the text and video condition using cross-tabulation (Table 2). A chi-square test showed that children's profiles for text and video were significantly related to each other, $\chi^2(9) = 96.53$, $p < .001$. Of all children, 69% fell in the same comprehension profile for text and video. This indicates that most children used a similar approach for understanding texts and videos. There are, however, a few differences in cluster membership that are noteworthy. First, several children who were *Monitoring readers* in the text condition were classified as *Elaborators* when watching videos ($n = 9$). In addition, some children who were *Monitoring readers* ($n = 4$) or *Elaborators* ($n = 7$) when reading texts were in the *Elaborating chatters* cluster when watching videos. These two shifts in profiles suggest that for some children watching a video elicited comprehension processes that went beyond the literal information more so than reading did.

Table 2. Number of children per cluster in the text and video condition.

		Video				
		Paraphrasers	*Monitoring*	*Elaborators*	*Elaborating chatters*	Total
Text	*Paraphrasers*	26	2	2	0	30
	Monitoring	0	15	9	4	28
	Elaborators	1	0	12	7	20
	Elaborating chatters	0	0	0	3	3
	Total	27	17	23	14	81

Discussion

In this study, we investigated how fourth-grade children understand expository texts and videos, focusing on both comprehension processes and comprehension outcomes. Children were instructed to share their thoughts at several points in the texts and videos, providing insight into how they proceed in understanding different media. To measure their comprehension outcomes, children were asked comprehension questions after each text and video. We extended earlier research in two ways, that is, by using an *online* process measure to capture comprehension processes (i.e., a think-aloud protocol) and by using expository rather than narrative texts and videos (see Diergarten & Nieding, 2016; Neuman, 1992). Another important contribution of our research lies in the exploratory analysis of latent profiles to examine whether individual children adopt a similar approach for understanding texts and videos.

Comprehension outcomes and processes

For our first research question, we tested whether children's comprehension outcomes in the text and video condition were comparable by looking at their performance on open-ended comprehension questions (RQ1). It was found that children's comprehension outcomes (i.e., their level of understanding) did not differ for text and video. This aligns with our expectation (Hypothesis 1) and with the Simple View of Reading (Hoover & Gough, 1990), which indicates that once text decoding skills are automated, the level of reading comprehension and listening comprehension (i.e., videos) will equalize. Moreover, our results corroborate findings from previous studies that compared comprehension outcomes across different media using narratives (e.g., Van Zeijts et al., 2023; Verlaan et al., 2017; Wannagat et al., 2017) but also with longer expository texts and videos (Delgado et al., 2022; Tarchi et al., 2021). Importantly, these studies vary in terms of genre and characteristics of texts and videos. However, also some studies have reported an advantage of a text format over other media formats, for example, when looking at comprehension and integration of information across multiple texts and videos (Lee & List, 2019; Salmerón et al., 2020). This suggests that differences between media may emerge depending on the specific comprehension task that is given. Nevertheless, our findings contribute to this body of research by showing that for a *single* expository text and a *single* talking-head video (featuring only a presenter without additional visual elements), children's comprehension outcomes are similar. Together with previous research, this indicates that when the same content is presented using different media formats and when readers possess proficient text decoding skills, the level of comprehension is comparable for text and video.

We extended previous research by not only measuring comprehension *outcomes* but also comprehension *processes* to gain more insight into what happens *during* reading and watching. Hence, for our second research question, we examined whether children's comprehension processes in the text and video condition were comparable by looking at their average number of think-aloud responses (RQ2a). Our results indicated that children were able to paraphrase, generate inferences, reflect on their own understanding, and make personal comments in both the text and video condition. Most importantly, and in line with our expectation (Hypothesis 2a), we found that the average number of responses in the text and video condition

was comparable across all think-aloud categories. This finding is consistent with some of the previous research using retrospective measures to capture the cognitive processes of students when reading expository texts and videos (Delgado et al., 2022; Mason et al., 2022; Tarchi et al., 2021). Both retrospective and online measures can be useful in assessing cognitive processing, but our online measure provides more direct evidence of what happens moment-to-moment.

When looking at the very limited research that used an *online* process measure (Diergarten & Nieding, 2016; Neuman, 1992), we see that previous findings were mixed. Our results are inconsistent with Diergarten and Nieding (2016), who focused on emotional inferences during reading than during listening, but consistent with Neuman (1992), who measured various types of inferences. Neuman reported that children's inference generation when reading narrative texts and when watching videos is very comparable. We replicated and extended this finding, as we found that the same applies for inference generation with expository texts and videos but also for a range of other comprehension processes, including paraphrasing, comprehension monitoring, and elaborations based on personal experiences.

Individual differences in comprehension processes

To move beyond group-level comparisons between text and video, we exploratively examined children's comprehension processes at an individual level by identifying different comprehension profiles (RQ2b). This allowed us to examine whether individual children display the same pattern of comprehension processes for understanding expository texts and videos. Using LPA, we identified four distinct comprehension profiles, *Paraphrasers, Monitoring, Elaborators*, and *Elaborating chatters*, for both the text and the video condition. The *Paraphrasers* and *Elaborating* readers are consistent with profiles that have been identified frequently in the literature on text comprehension with both narrative and expository texts (e.g., Karlsson et al., 2018; McMaster et al., 2012; Rapp et al., 2007), indicating that these profiles are relatively robust. We identified two additional profiles, namely a group of children that primarily focused on comprehension monitoring and a (relatively smaller) group of children scoring high on most categories. The four-profile solution provided a good description of our data in the text as indicated by R^2 entropy values. Nevertheless, the two additional profiles should be interpreted with caution, as it is uncertain whether they generalize to different participant samples and/or other texts and videos.

We explored whether the four comprehension profiles differed from each other in terms of comprehension outcomes, providing information on the relative success of the different approaches to understanding. Results showed no differences between the profiles in the text condition, but in the video condition we found that children in the *Elaborators* and *Elaborating chatters* clusters had significantly higher comprehension outcomes than children in the *Paraphrasers* cluster. Apparently, the many irrelevant responses of children with an *Elaborating chatters* profile did not interfere with producing inferences and engaging in other high-level comprehension processes. This indicates that with video, actively elaborating on the information that is presented (e.g., by drawing inferences or sharing personal information) is associated with a higher level of understanding. This aligns with the different levels of a mental representation as described by Kintsch (1988), suggesting that the *Elaborators* and *Elaborating chatters* were able to construct a coherent representation at the situation model level.

Interestingly, the profiles that emerged from the separate LPAs for text and video were very similar, and we found that most children (69%) fell in the same profile for text and video. This demonstrates that not only children's comprehension processes are largely comparable for text and video, but also that individual children tend to adopt a consistent approach for understanding information presented in text and video format. This person-centered analysis provides converging evidence for the similarity in children's comprehension processes across media formats. There was, however, also a group of children that did not use the same approach (31% of children), as illustrated by differences in their cluster membership in the text and video condition. These differences in approaches between conditions may be related to factors such as fatigue, attention, or engagement during the task. However,

given that we carefully counterbalanced the order of topics and conditions (text vs. video), we expect the risk of these factors confounding the results to be minimized.

The largest differences in cluster membership were children who were in the *Monitoring readers* cluster in the text condition but were in the *Elaborators* cluster in the video condition. Moreover, there were several children who were in the *Monitoring readers* and *Elaborators* clusters for text but were in the *Elaborating chatters* cluster for video. For these children, watching videos seems to elicit more elaborative responses compared to text. This might be related to the presenter in the video adding a social component to the comprehension process, possibly stimulating some children to talk and elaborate more on the presented information. It is, however, important to note that this finding contradicts with our conclusion based on group averages, which showed no differences between the text and video condition in any of the think-aloud categories. The inconsistency between our findings at the group level and individual level may be related to (small) differences between the clusters that were formed in the text condition and video condition. For example, the *Elaborating chatters* cluster in the text condition was characterized by a higher number of meta-cognitive and irrelevant responses compared to the *Elaborating chatters* cluster in the video condition. Consequently, more children fitted the *Elaborating chatters* cluster in the video condition compared to the text condition. The fact the profiles that were formed for text and video are not entirely identical makes is difficult to determine why some children fell in a different profile. All in all, our main take-away is that the majority of children remain in the same cluster, indicating that they use a similar approach when trying to understand our texts and videos.

Theoretical and educational implications

Following from the results summarized above, our study has implications for theory and practice. First, a theoretical implication of our study is that it provides additional empirical support in favor of the *unitary process view* instead of the *dual process view*. A key contribution of our work is that we asked children to report their thoughts *during* the task, instead of retrospectively (e.g., Lee & List, 2019), which provided insight into children's moment-by-moment comprehension processes. Our findings strengthen the evidence supporting the comparability of cognitive processes across media formats and suggest that theoretical models traditionally developed for text comprehension, such as the Construction-Integration Model, can be extended to other media. We find that comprehension processes are comparable at the group level, but also that most children tend to use a similar approach for understanding texts and videos. This individual-level result supports the idea of a *general comprehension skill* that transfers across different media formats (see Kendeou et al., 2020). Finally, our results suggest that if the content of expository texts and videos is equivalent, children's comprehension processes and outcomes are comparable. This suggests that potential comprehension differences between media formats might be more related to differences in the presented information than to the medium itself. For example, differences in language use (with more complex and formal language in written language compared to spoken communication) or additional visual cues providing context in videos.

Second, our study demonstrates that children are able to actively process the spoken language in videos, similarly to how they process written texts. An often-voiced concern among teachers is that television and videos provoke relatively passive processing compared to reading. Even though our participants were not very used to videos containing expository content in the classroom, we observed that they demonstrated a wide range of relevant comprehension processes, such as inference generation and comprehension monitoring. This is significant for educational practice, as it indicates that the use of language-rich videos can elicit a range of valuable comprehension processes that are also important for reading comprehension. The integration of such videos within reading comprehension instruction may provide students diverse opportunities to practice their general comprehension skills. This warrants further research, particularly through intervention studies, to evaluate whether reading comprehension skills can be effectively practiced using video materials and whether

this transfers to improved reading comprehension skills. If proven effective, this could offer a novel and potentially motivating approach for teaching comprehension strategies and fostering children's reading comprehension skills.

Limitations and suggestions for future research

This study has a few limitations that should be addressed in future research. One limitation relates to the think-aloud protocol that was used to measure children's comprehension processes. We adopted this methodology because this is an online process measure that is (1) feasible in both a reading and video context, and (2) allows us to capture a wide range of comprehension processes, not only inference generation. Children's think-aloud responses provided valuable insights into how they proceed in understanding different media. Nevertheless, it also has several limitations that are important to consider. First, a think-aloud protocol requires children to verbalize their own comprehension process. This asks for insight into one's own understanding and the ability to put this into words differs per individual (Ericsson & Simon, 1993; Schellings et al., 2006). Even though think-aloud has been successfully used with children before (e.g., Schellings et al., 2006), the task of verbalizing cognitive processes may be more challenging for them. A related issue is that think-aloud protocols only allow us to measure conscious processing (Pressley & Afflerbach, 2012). Some comprehension processes are unconscious because they are automated and occur very quickly, such as the generation of local text-based inferences. Children might draw these inferences without even noticing and consequently fail to include it in their verbal reports. Finally, the think-aloud task itself could influence the comprehension process (Caldwell & Leslie, 2010; Rapp et al., 2007), for example, by encouraging children to engage in more active processing. Nevertheless, the validity of think-aloud protocols has been supported by research relating children's verbal responses during reading to their eye movements (Kaakinen & Hyona, 2005; also see Kraal et al., 2018, 2019). Moreover, these limitations are present in both the text and video conditions, and we assumed the influence of these factors to be similar for both media. Note, however, that we were not able to test this assumption, as the present study did not include a condition without think-aloud. Given the discussed limitations, it would be relevant for future research to replicate our text-to-video comparison using a different online process measure. For example, by using an implicit task that is less dependent on children's ability to verbalize their comprehension process (e.g., a reaction time paradigm, see Diergarten & Nieding, 2016), we could also capture unconscious processes.

Another limitation concerns to the ecological validity of the videos used in our study, as these do not represent all types of videos that children watch in their day-to-day life. Our talking-head videos were developed based on expository texts, which were selected from educational practice and did have high ecological validity. We deliberately created videos that closely matched the written text version to ensure equivalence in terms of content, allowing us to make a controlled comparison between the two media (Tversky et al., 2002). This way, potential differences between the text and video condition can be attributed to the medium itself and not to dissimilarities in the information that was presented. As a result, our videos did not contain much visual information, besides a presenter, subheadings, and two images. It is important to note that our findings are likely to depend on characteristics of videos and texts that were used, such as genre, duration, and complexity. For example, the addition of more visual information might affect children's processing of the videos. For future research, it would be interesting to replicate our setup with different types of videos, such as videos containing more visual information, although creating a text version that matches the visual information presented in these videos will be a challenge. This could provide insight into how different features and characteristics of a video influence comprehension processes and outcomes.

Conclusion

This study demonstrates that fourth-grade children use very similar comprehension processes when attempting to understand expository texts and videos. The use of a think-aloud protocol as an *online* process measure contributes to our understanding of how children approach these different media formats. We provide additional empirical support for a *unitary process view*, supporting the assumption often made by researchers that understanding text and video relies on similar comprehension processes. Our findings raise optimism for potential crossover between media, and lay a foundation for intervention studies to explore the potential role of videos in teaching reading comprehension skills.

Notes

1. Our preregistration included a research question about children's attitude toward reading and watching videos. As the present study is focused on comprehension processes and outcomes, we decided to describe the attitude data in a separate article.
2. Our preregistered analysis included literal comments, inferences, meta-cognitive comments, and personal comments as levels. When coding children's think-aloud responses, we added a category for irrelevant comments. We decided to include this in our analyses to examine whether texts or videos would elicit more irrelevant responses.

Acknowledgments

We thank CED-groep for allowing us to use texts from the archive of *Nieuwsbegrip*. We also thank Kirsten Muller and Maya Soeratram for their help with collecting, transcribing, and scoring the data. Finally, we thank Emma-Sophie Ekelmans, Max Pickkers, Michelle Frijters, and Sander van Velze for their role as presenters in our videos.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was part of an Interlinked Research Project funded by NRO (The Netherlands Initiative for Education Research): 40.5.18300.038. The contribution of Marian Hickendorff was supported by a personal grant from the Dutch Research Council (NWO): 016.Veni.195.166/6812.

References

Beege, M., Schroeder, N. L., Heidig, S., Rey, G. D., & Schneider, S. (2023). The instructor presence effect and its moderators in instructional video: A series of meta-analyses. *Educational Research Review*, *100564*. https://doi.org/10.1016/j.edurev.2023.100564

Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology*, *29*(2), 137–164. https://doi.org/10.1080/02702710801963951

Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing*, *11*(5/6), 489–503. https://doi.org/10.1023/A:1008084120205

Caldwell, J., & Leslie, L. (2010). Thinking aloud in expository text: Processes and outcomes. *Journal of Literacy Research*, *42*(3), 308–340. https://doi.org/10.1080/1086296X.2010.504419

Catts, H. W., Herrera, S., Nielsen, D. C., & Bridges, M. S. (2015). Early prediction of reading comprehension within the simple view framework. *Reading and Writing*, *28*(9), 1407–1425. https://doi.org/10.1007/s11145-015-9576-x

Delgado, P., Anmarkrud, Ø., Avila, V., Altamura, L., Chireac, S. M., Pérez, A., & Salmerón, L. (2022). Learning from text and video blogs: Comprehension effects on secondary school students. *Education and Information Technologies*, 1–27. https://doi.org/10.1007/s10639-021-10819-2

Denton, C. A., Enos, M., York, M. J., Francis, D. J., Barnes, M. A., Kulesz, P. A., Fletcher, J. M., & Carter, S. (2015). Text-processing differences in adolescent adequate and poor comprehenders reading accessible and challenging narrative and information text. *Reading Research Quarterly*, *50*(4), 394–416. https://doi.org/10.1002/rrq.105

Diakidoy, I. A. N., Stylianou, P., Karefillidou, C., & Papageorgiou, P. (2005). The relationship between listening and reading comprehension of different types of text at increasing grade levels. *Reading Psychology*, *26*(1), 55–80. https://doi.org/10.1080/02702710590910584

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

Diergarten, A. K., & Nieding, G. (2016). Online emotional inferences in written and auditory texts: A study with children and adults. *Reading and Writing*, *29*(7), 1383–1407. https://doi.org/10.1007/s11145-016-9642-z

Dutch Media Literacy Network. (2021). *Monitor mediagebruik kinderen 7-12 jaar* [Monitor media use among children aged 7-12 years]. Netwerk Mediawijsheid.

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

Evers, G. (2008). *Handleiding voor de bepaling van AVI en CLIB met P-CLIB versie 3.0.* [Manual for calculating AVI and CLIB using P-CLIB version 3.0]. Cito.

Fyfield, M., Henderson, M., & Phillips, M. (2022). Improving instructional video design: A systematic review. *Australasian Journal of Educational Technology*, *38*(3), 155–183. https://doi.org/10.14742/ajet.7296

Gernsbacher, M. A., & Robertson, R. R. (1992). Knowledge activation versus sentence mapping when representing fictional characters' emotional states. *Language and Cognitive Processes*, *7*(3–4), 353–371. https://doi.org/10.1080/01690969208409391

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 430–445. https://doi.org/10.1037/0278-7393.16.3.430

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., Wiemer-Hastings, P., & Wiemer-Hastings, K. (2001). Constructing inferences and relations during text comprehension. In T. Sanders, J. Schilperoord, & W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 249–271). John Benjamins.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*(1), 77–89. https://doi.org/10.1080/19312450709336664

Hickendorff, M., Edelsbrunner, P. A., McMullen, J., Schneider, M., & Trezise, K. (2018). Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis. *Learning and Individual Differences*, *66*, 4–15. https://doi.org/10.1016/j.lindif.2017.11.001

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127–160. https://doi.org/10.1007/BF00401799

Kaakinen, J. K., & Hyona, J. (2005). Perspective effects on expository text comprehension: Evidence from think-aloud protocols, eye-tracking, and recall. *Discourse Processes*, *40*(3), 239–257. https://doi.org/10.1207/s15326950dp4003_4

Karlsson, J., van den Broek, P. W., 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., Bohn-Gettler, C., White, M. J., & van den Broek, P. W. (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., 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

Kendeou, P., van den Broek, P. W., 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

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

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*(9), 2017–2040. https://doi.org/10.1007/s11145-017-9789-2

Kraal, A., van den Broek, P. W., Koornneef, A. W., Ganushchak, L. Y., & Saab, N. (2019). Differences in text processing by low-and high-comprehending beginning readers of expository and narrative texts: Evidence from eye movements. *Learning and Individual Differences*, *74*, 101752. https://doi.org/10.1016/j.lindif.2019.101752

Lee, H. Y., & List, A. (2019). Processing of texts and videos: A strategy-focused analysis. *Journal of Computer Assisted Learning*, *35*(2), 268–282. https://doi.org/10.1111/jcal.12328

Linderholm, T., & van den Broek, P. W. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology*, *94*(4), 778. https://doi.org/10.1037/0022-0663.94.4.778

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

List, A., & Ballenger, E. E. (2019). Comprehension across mediums: The case of text and video. *Journal of Computing in Higher Education*, *31*(3), 514–535. https://doi.org/10.1007/s12528-018-09204-9

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., Loschky, L. C., Clinton, J. A., & Larson, A. M. (2013). Is reading the same as viewing. In B. Miller, L. E. Cutting, & P. McCardle (Eds.), *Unraveling the Behavioral, Neurobiological and Genetic Components of Reading comprehension* (pp. 78–90). Brookes Publishing Co.

Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *15*(5), 533–545. https://doi.org/10.1002/acp.724

Mason, L., Tarchi, C., Ronconi, A., Manzione, L., Latini, N., & Bråten, I. (2022). Do medium and context matter when learning from multiple complementary digital texts and videos? *Instructional Science*, *50*(5), 653–679. https://doi.org/10.1007/s11251-022-09591-8

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.004

McMaster, K. L., van den Broek, P. W., Espin, C. A., 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

McNamara, D. S., & Maglino, J. (2009). Toward a comprehensive model of comprehension. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (pp. 297–384). Elsevier. https://doi.org/10.1016/S0079-7421(09)51009-2

Mullis, I. V. S., von Davier, M., Foy, P., Fishbein, B., Reynolds, K. A., & Wry, E. (2023). *PIRLS 2021 International Results in Reading*. https://doi.org/10.6017/lse.tpisc.tr2103.kb5342

National Assessment of Educational Progress. (2022). *The nation's report card*. https://www.nationsreportcard.gov/highlights/reading/2022/

Neuman, S. B. (1992). Is learning from media distinctive? Examining children's inferencing strategies. *American Educational Research Journal*, *29*(1), 119–140. https://doi.org/10.3102/00028312029001119

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569. https://doi.org/10.1080/10705510701575396

Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, *4*(4), 440–461. https://doi.org/10.1037/tps0000176

Pressley, M., & Afflerbach, P. (2012). *Verbal protocols of reading: The nature of constructively responsive reading*. Routledge. https://doi.org/10.4324/9780203052938

Rapp, D. N., van den Broek, P. W., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading*, *11*(4), 289–312. https://doi.org/10.1080/10888430701530417

Salmerón, L., Sampietro, A., & Delgado, P. (2020). Using internet videos to learn about controversies: Evaluation and integration of multiple and multimodal documents by primary school students. *Computers and Education*, *148*, 103796. https://doi.org/10.1016/j.compedu.2019.103796

Schellings, G., Aarnoutse, C., & Van Leeuwe, J. (2006). Third-grader's think-aloud protocols: Types of reading activities in reading an expository text. *Learning and Instruction*, *16*(6), 549–568. https://doi.org/10.1016/j.learninstruc.2006.10.004

Sinatra, G. M. (1990). Convergence of listening and reading processing. *Reading Research Quarterly*, *25*(2), 115–130. https://doi.org/10.2307/747597

Swart, N. M., Gubbels, J., Zandt, M., Wolbers, M. H. J., & Segers, E. (2023). *PIRLS-2021: Trends in leesprestaties, leesattitude en leesgedrag van tienjarigen uit Nederland*. Expertisecentrum Nederlands. https://expertisecentrumnederlands.nl/pirls-2021

Tarchi, C., Zaccoletti, S., & Mason, L. (2021). Learning from text, video, or subtitles: A comparative analysis. *Computers and Education*, *160*, 104034. https://doi.org/10.1016/j.compedu.2020.104034

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* [Scientific justification of reading comprehension tests for fourth grade]. Cito.

Trabasso, T., & van den Broek, P. W. (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

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies*, *57*(4), 247–262. https://doi.org/10.1006/ijhc.2002.1017

van Zeijts, B. E. J., Ganushchak, L. Y., de Koning, B. B., & Tabbers, H. K. (2023). Stimulating inference-making in second grade children when reading and listening to narrative texts. *Reading and Writing*, 1–27. https://doi.org/10.1007/s11145-023-10463-x

Verlaan, W., Pearce, D. L., & Zeng, G. (2017). Revisiting Sticht: The changing nature of the relationship between listening comprehension and reading comprehension among upper elementary and middle school students over the last 50 years. *Literacy Research & Instruction*, *56*(2), 176–197. https://doi.org/10.1080/19388071.2016.1275070

Vermunt, J. K., & Magidson, J. (2021). *Upgrade manual for Latent Gold Basic, Advanced/Syntax, and Choice version 6.0*. Statistical Innovations Inc. https://www.statisticalinnovations.com/wp-content/uploads/LG60manual.pdf

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

Appendix

Table A1. Model fit statistics for 1 to 8 latent classes for the text condition, with lowest values in bold.

Latent Classes	Bayesian Information Criterion	Akaike Information Criterion	Consistent Akaike Information Criterion	Class Error	R^2 Entropy
1	2499.9	2475.9	2509.9	0.00	1.00
2	2372.6	2322.4	2393.6	0.03	0.89
3	2342.3	2265.7	2374.3	0.11	0.77
4	**2328.7**	2225.7	**2371.7**	0.05	0.87
5	2335.8	2206.5	2389.8	0.05	0.90
6	2353.4	2197.8	2418.4	0.05	0.90
7	2371.2	**2189.2**	2447.2	0.04	0.93
8	2398.3	2190.0	2485.3	0.03	0.95

Table A2. Model fit statistics for 1 to 8 latent classes for the video condition, with lowest values in bold.

Latent Classes	Bayesian Information Criterion	Akaike Information Criterion	Consistent Akaike Information Criterion	Class Error	R^2 Entropy
1	2505.3	2481.3	2515.3	0.00	1.00
2	2364.0	2313.8	2385.0	0.05	0.82
3	2309.3	2232.6	2341.3	0.02	0.94
4	**2300.1**	2197.2	**2343.1**	0.04	0.91
5	2305.8	2176.5	2359.8	0.03	0.94
6	2326.0	2170.4	2391.0	0.04	0.93
7	2353.6	2171.6	2429.6	0.05	0.92
8	2371.9	**2163.6**	2458.9	0.05	0.92