Using think-aloud methods to investigate inter- and intra-individual differences in reading processes

Broek, P.W. van den

Routledge
Taylor & Francis Group

Discourse Processes

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

Using think-aloud methods to investigate inter- and intra-individual differences in reading processes

Paul van den Broek

To cite this article: Paul van den Broek (04 Feb 2026): Using think-aloud methods to investigate inter- and intra-individual differences in reading processes, Discourse Processes, DOI: 10.1080/0163853X.2026.2617116

To link to this article: https://doi.org/10.1080/0163853X.2026.2617116

Published online: 04 Feb 2026.

DISCOURSE PROCESSES
https://doi.org/10.1080/0163853X.2026.2617116

Routledge
Taylor & Francis Group

Using think-aloud methods to investigate inter- and intra-individual differences in reading processes

Paul van den Broek

Department of Education and Child Studies, Leiden University, Leiden, The Netherlands

ABSTRACT

To understand how people read, one needs to gain insight into the processes by which they engage with a text and by which they construct meaning. This paper discusses the use of Think-aloud methodologies to investigate such processes as well as differences in processes between and within individuals. Think-aloud methods have considerable potential as they may reveal the actual contents of reading processes but also pose methodological and interpretational challenges. I discuss such potential and challenges, and reflect on considerations for methodological choices that optimize potential and mitigate challenges. I illustrate these issues by reviewing empirical studies on three dimensions of inter- and intra-individual differences in reading behavior (reading skill, processing profile, and reading goals/standards of coherence). The purpose is to give researchers food for thought on how Think-aloud methods can be used best to investigate (differences in) reading behavior and how to make informed choices about the design of their Think-aloud studies.

To understand how people read, it is essential to gain insight into the processes by which they engage with a text and, especially, by which they construct meaning from a text. In this paper I discuss the use of Think-aloud methods to investigate such meaning-construction processes, with particular attention to the study of *inter- and intra-individual differences* in reading behaviors. When it comes to identifying how individual readers may differ in their online processing of the texts they read, Think-aloud methods have considerable potential as they may reveal the actual contents of reading processes – unlike other online processing measures that we[1] and others have used extensively, such as eye-movements, reading times, reaction times and electromagnetic brain recordings. However, using Think-aloud methods also poses considerable challenges with respect to both methodology and interpretation. Some of these challenges are common to all applications of the Think-aloud methodology to study reading comprehension, for example that they only can reveal processes of which the reader is aware (see Van den Broek and Helder (2017) on reader-initiated processes, McNamara et al. (2023) on constructed responses). But other challenges are unique to or amplified in investigations of (intra)individual differences, for example the potential influence of participants' differences in verbal skills on their ability to express their thoughts. This paper discusses the opportunities and the challenges that Think-aloud methods offer in the study of reading processes in general and of differences in such processes between and within individuals, and reflects on considerations for making informed methodological choices and for ways to mitigate the challenges.

These issues will be illustrated and discussed in the context of empirical studies on three dimensions of differences in reading behavior: reading skill, processing profile, and reading goals/standards of coherence. The overall purpose of the paper is to encourage researchers to reflect on

CONTACT Paul van den Broek broekpwvanden@fsw.leidenuniv.nl

how Think-aloud methods can be used to investigate (individual differences in) reading behavior and how to make informed choices in deciding on the research designs of their Think-aloud studies. Consideration of these issues may be useful for researchers who do not have much experience in using Think-aloud methods but would like to start as well as for researchers who already have been using Think-aloud methods but want to reflect on the choices they have been making (implicitly or explicitly). Together with the other articles in this *Discourse Processes* special issue, The Past, Present, and Future of Verbal Protocol Analysis in Text and Discourse, this paper aims to stimulate the creative and careful use of Think-aloud methods in future research on important issues in discourse and language processing.

Think-aloud procedures in reading research

Think-aloud procedures frequently are used in reading research. The common element across studies is that readers verbalize their thoughts, feelings, and strategies while they are reading, thereby providing a glimpse into the processes in which they engage. But the ways in which Think-aloud procedures are implemented vary considerably across studies. They differ, for example, in when and how often readers are asked to give a response, the type of prompts that are provided, whether participants receive training and/or practice, whether they are free to respond whatever they feel like or are instructed to provide certain types of responses, how responses are recorded, and many other ways. All of these choices have a potential impact on what the readers will report and, therefore, on the results in a study.

Besides implementational variations, there are considerable variations in the way that responses are analyzed. Think-aloud procedures can be used to inventory many different types of reader activities (e.g., comprehension as well as non-comprehension activities), and responses can be categorized in many different ways (e.g., at different grain sizes, considering only the first or also subsequent responses to the text input, via different coding schemes). Such inventories and categorizations can be driven by bottom-up decisions: See what kinds of responses occur and then somehow try to derive ad hoc categories. But responses often are too varied and noisy to do so in a transparent and replicable manner. Instead, both the instructions given to the readers and the categorization of responses usually are based on a priori and, preferably, theoretical perspectives. Regardless of how the categorization scheme to analyze responses is developed, once adopted it determines to a large extent which phenomena are (and are not) the focus of the study. These and other decision points will be discussed below.

There are several classic general introductions to the use of Think-aloud procedures to investigate cognitive processes (e.g., Ericsson & Simon, 1993/1984; Chi, 1997) and reading in particular (e.g., Long & Bourg, 1996; Olson et al., 1984; Pressley & Afflerbach, 1995; Whitney & Budd, 1996). There also are several more recent papers with tips and pointers for conducting Think-aloud research (e.g., Charters, 2003; Hu & Gao, 2017; Noushad et al., 2023). These useful tips and pointers are mostly practical in nature (make sure the environment is quiet, prevent distractions, and so on), and focus less on substantive decision points that conceptually influence the results and conclusions, such as those mentioned above. To make these decision points concrete, several studies on inter- and intra-individual differences using Think-aloud procedures from our lab are described next.

Investigating inter/intra-individual differences in reading behavior: example studies

To illustrate the use of Think-aloud procedures and provide a starting point for discussion of decision points, three sets of studies investigating inter- and intra-individual variation in reading processes are summarized-on reading skill, processing profiles, and reading goals/standards ofcoherence, respectively. This section briefly describes the theoretical background and the Think-aloud procedures the studies share, followed by the substantive aims and findings of each set of studies. In the next section, decisions points in the design of these and other studies will be discussed systematically.

The implementation of Think-aloud procedures is similar but not identical across these studies. An important commonality is that they all focus on reading *comprehension*. Conceptually, they draw on the Landscape model (Van den Broek et al., 1999; Yeari & van den Broek, 2011) and other theoretical models of meaning construction from texts (such as the Construction-Integration, Structure-Building, Resonance, Causal Network, and Constructionist models; for reviews see McNamara & Magliano, 2009; Van den Broek & Kendeou, 2022). It is important to point out, however, that the decision points that will be discussed apply equally to the use of Think-aloud procedures from other theoretical perspectives and with research goals other than the study of comprehension.

The sets of studies that are described here use as the Landscape Model of reading comprehension as theoretical framework. The central premise of this model is that comprehension and interpretation of a text revolves around the gradual construction of a mental representation of the text. As readers proceed through a text, they identify meaningful connections between different parts of the text and between the text and their background knowledge and experience (e.g., conceptual knowledge about the text topic and about the world, knowledge about language, texts, and reading strategies, knowledge from personal and affective experience, cultural embedding; e.g., Hattan et al., 2024; Kintsch, 1994; Van den Broek & Kendeou, 2022). Connections are constructed by a combination of passive and active, reader-initiated processes. Passive processes occur without conscious awareness or effort from the reader and include associative processes such as spread-of-activation, whereby incoming text automatically activates concepts that are related in the reader's mental representation of the prior text or in their background knowledge (e.g., Albrecht & O'Brien, 1993; McNamara et al., 2023; Van den Broek & Helder, 2017; Van Moort, Jolles et al., 2020; Van Moort, Koornneef et al., 2020).[2] The connections that result from passive processes are very limited, so readers usually also engage in active, reader-initiated processes to meet their implicit or explicit *standards of coherence*. A reader's standards of coherence reflect the type and depth of coherence the reader aims to achieve (e.g., Van den Broek et al., 2011; Van den Broek & Helder, 2017; Van den Broek et al., 1995. See also Kendeou, 2014; Britt et al., 2017; Elfenbein, 2018, for extensions of the notion of standards of coherence). Reader-initiated processes are effortful actions the reader initiates to meet their standards and to build coherence and create comprehension beyond what the passive processes offer. Such actions may range from rather modest, close-to-the-text processes to involved, far-from-the-text processes. Examples of close-to-the-text processes are rereading the current text input to understand, applying strategies to establish referential and causal coherence, using text structure and headers to fit parts of the text together. Examples of far-from-the-text processes are reflecting on the text, creating a literary critique, evaluating the source and trustworthiness of a text, comparing multiple texts. As the reader proceeds through a text, the connections resulting from the processes triggered by each input cycle lead to a gradual emergence of a mental representation of the text as a whole. It is important to note that Think-aloud methods primarily can be used to investigate processes of which the reader is aware, i.e., those that are reader-initiated.

Investigation of discourse comprehension can be done at various levels of text analysis. The Landscape and related models of comprehension usually take main clauses, major propositions, or sentences as their unit of analysis, but larger units (e.g., paragraphs, segments of multiple sentences) could be used as unit of analysis as well. The example studies use sentences as units.

Given this theoretical framework, the Think-aloud procedures in the summarized studies focus on (a) reader activities that are relevant to comprehension of the text at hand, (b) processing strategies involved in building a coherent mental representation of the text (e.g., constructed responses; see McNamara et al., 2023), (c) both processes that involve information from the text and processes that recruit reader's background knowledge, (d) systematic variation in processing strategies between and within readers as a function of individual differences or of reading goals.

Think-aloud responses are categorized according to a coding scheme based on this theoretical framework and inspired by earlier Think-aloud studies (e.g., Narvaez et al., 1999; Trabasso & Magliano, 1996a, 1996b; Zwaan & Brown, 1996). The main categories for coding the responses in these studies are presented in Table 1.

Table 1. Main categories for coding think-aloud responses in the example studies.

Category	Operational Definition
Text repetitions	Verbatim repetition of all or a large proportion of the current sentence
Paraphrases of the text	Restatement of the current sentence in one's own words
Connecting inferences	Explanation of the current sentence by connecting it to the immediately preceding sentence
Reinstatements	Reactivation of information from earlier parts of the text beyond the immediately preceding sentence
Associations	Activation of the reader's background knowledge that does not contribute to text coherence (e.g., free associations)
Elaborative/explanatory inferences	Activation of background knowledge that helps explain the current sentence; may be *valid* or *invalid* depending on consistency with the preceding text
Predictive inferences	Anticipation of what will occur next; may be *valid* or *invalid* depending on consistency with the preceding text
Affective responses	Emotional reactions elicited by the text
Metacognitive/ comprehension-monitoring	Comments reflecting monitoring of (lack of) understanding (e.g., "mmh, I am not sure I understand that")
Evaluations of the text	Judgments or critiques of the text's quality or clarity

These categories range from *close-to-the-text* responses (e.g., repetitions and paraphrases of the sentence just read), *within-text-integration* responses (e.g., text-connecting processes to the immediately preceding sentence and reinstatements of earlier text information), *elaborations from knowledge* (e.g., recruitment of background knowledge that contributes to understanding the text, that yields embellishments that are irrelevant to coherence and understanding, or that results in forward inferences about upcoming text), to *reflections* (e.g., evaluations and metacognitive comments).

In the studies that are described next, participants were asked to provide their thoughts after every sentence they read. Their responses were collected in individual reading sessions via voice recorder. The transcribed responses were categorized by multiple raters using the above coding scheme, and disagreements were resolved through discussion. All use this general Think-aloud method, adapted for their specific populations and research goals.

Reading skill and text processing

Think-aloud procedures can be used to gain a better understanding of possible differences in reading processes between skilled and less-skilled comprehenders. Here, I summarize the results from a set of studies on comprehension processes by elementary school students using the above Think-aloud method. In these studies, struggling and good readers were identified using a standardized reading-comprehension test (sometimes a combination of tests).[3] In several of the studies, technical reading skills (e.g., word identification, fluency) were also assessed to ensure that any observed comprehension difficulties were not due to a lack of those skills. In individual sessions, the children read age-appropriate texts at their own pace and provided think-aloud responses as described in the preceding section. Their responses were recorded, transcribed and coded according to the above categorization scheme. The studies vary in the grade level of participants, their native language (different countries), the kind of texts the children read (narratives, expository texts or both), and procedural details, thus providing the opportunity for generalizable findings.

Across these studies, good and struggling readers appear to engage in a similar *range* of processes. Both groups of readers made metacognitive comments, evaluations, affective responses, and free associations but neither did so frequently. In contrast, text repetitions, paraphrases, reinstatements and forward/backward inferences made up the bulk of responses for both struggling and good readers. The struggling readers tend to execute them *less effectively*, however (e.g., Karlsson et al., 2018; Kraal et al., 2018; Rapp et al., 2007; Seipel et al., 2016). For example, less-skilled readers engaged in elaborative/explanatory and predictive inferencing pretty much as do skilled readers but more often the inferences they generated were invalid or were inconsistent with the text (Clinton-Lisell et al., 2024; Kraal et al., 2018). Likewise, poor readers' metacognitive comments most frequently pertained to

vocabulary whereas good readers' metacognitive comments most frequently pertained to the content of the text (Seipel et al., 2016).

Interestingly, these findings for elementary-school children are consistent with individual differences in think-aloud results for older readers. For example, less-skilled undergraduate readers tend to focus less on processing causal explanations and more on reprocessing the specific sentence just read than do skilled readers (Magliano & Millis, 2003), suggesting they engage in less coherence building. Similarly, readers in a foreign language strategically adjust their processing to different types of text and reading tasks but do so less efficiently than native readers, especially with more demanding texts (Horiba, 2010).

In summary, these results suggest that struggling and good readers have similar sets of reading-comprehension processes in their toolboxes, and that it is not the case that struggling reader systematically fail to engage in one or more particular types of processes that good readers rely on. However, they do differ in the effectiveness of the execution of processes. For example, struggling readers produce significantly more invalid elaborative and predictive inferences than do good readers The reduced effectiveness of processing among poor comprehenders was also reflected in their lower performance on post-reading outcome measures, such as recall and responses to comprehension questions, across all of the studies reviewed above.

Profiles of processing strategies

A second set of studies investigates systematic differences between readers in their reading-comprehension processes while proceeding through a text from a different vantage point, namely with the goal to determine whether there are differences in the *combinations* or *patterns* of processing strategies in which they engage. To address this goal, individuals read texts while thinking aloud, following the procedures and categorization scheme outlined in the preceding section. In the first of these studies (see McMaster et al., 2014; Rapp et al., 2007), proficient, average and struggling children from elementary, middle and high-schools read narrative texts. Cluster analyses of their think-aloud responses showed that there were two sub-groups of struggling readers with distinct patterns of think-aloud responses. These same subgroups emerged at each of the three school levels.

The subgroups had distinct processing profiles. *Paraphrasers* mostly focused their processing on the text they had just read. They frequently simply repeated or paraphrased the sentence, and made few inferences in which they went beyond the text by drawing on their background knowledge. *Elaborators* frequently went beyond the text and recruited background knowledge to amplify their understanding and interpretation of the text. In fact, the struggling elaborators made as many elaborative inferences as did average and good readers, but their inferences were more likely to be inaccurate or invalid – a point to which I return below.

The finding of these two subgroups of struggling readers is consistent with other findings that some struggling readers fail to generate many inferences and thus have difficulty establishing coherence (e.g., Cain & Oakhill, 2006), whereas other struggling readers often have difficulty constructing coherent representations of text because they use background knowledge or personal viewpoints inappropriately (e.g., Williams, 1993). The results of the cluster analyses of Think-aloud responses indicate that both observations are correct and that, in fact, they reflect distinct subgroups of struggling readers that struggle in their own ways (see also Nation et al., 2002; Perfetti & Helder, 2022; Rapp et al., 2007; Van den Broek et al., 2015). In this respect, it is interesting that the two subgroups of struggling readers did not show differences in their performance on post-reading measures such as recall and comprehension questions – they performed equally poorly. The subgroups only emerged when their *processing* was considered.

Further studies using Think-aloud procedures in the USA and in the Netherlands replicated and extended these findings. The same subgroups emerged in an intervention study on struggling 4th-grade students (McMaster et al., 2012). Moreover, the two subgroups benefitted from different interventions, attesting to the reality of the distinction (McMaster et al., 2014;

McMaster et al., 2012). The two profiles also appear to be present at a younger age, 2nd-grade students, and for both narrative and expository texts (Kraal et al., 2018). Similar results were obtained in think-aloud responses to narrative and expository texts by 9–11-year-old children, with the children generally exhibiting the same profiles across the two text genres but expository texts eliciting fewer correct and more invalid inferences than did narratives (Karlsson et al., 2018).

In summary, these results suggest that there are indeed systematic differences in processing patterns between individual readers, that these patterns generalize across texts and across narrative and expository genres, and that they manifest themselves across a considerable age range and in different countries. In combination with the results on good and struggling readers in the previous section, these results also suggest that paraphrasing and elaborating struggling readers struggle in different ways. The paraphrasing readers that struggle may be less efficient in building a coherent mental representation because they focus on the local text information rather than on building explanations. In contrast, the elaborators that struggle do work toward coherence but frequently construct elaborations that are inaccurate or inconsistent with the text. Thus, there are readers who stay too close to the text and readers who recruit background knowledge that is inaccurate or that is not constrained enough by the text. Both are likely to struggle, whereas a balance between constraints provided by the text and accessing (accurate) background knowledge would be most likely to lead to good comprehension (cf., McCarthy & McNamara, 2021; Magliano et al., 1999). This interpretation is supported by the abovementioned finding that the subgroups benefited from different interventions (McCarthy et al., 2021). Further investigation of differences between these (and possibly other) subgroups would be important, both for reading theory and for reading instruction. For example, it is intriguing that Karlsson et al. (2018) observed that elaborating readers had better word decoding skills, reading comprehension ability, and non-verbal reasoning ability than paraphrasing readers, suggesting a positive relation between inference generation, on the one hand, and language abilities and cognitive resources, on the other hand.

In the context of this paper, the results of this section and of the preceding section show that Think-aloud procedures can be a powerful tool to gain insight in individual differences in the processing during reading. Future research using Think-aloud procedures may reveal other subgroups, besides *paraphrasers* and *elaborators* (e.g., Karlsson et al., 2018). In addition, it may give more detailed insight into the differences in processing by struggling readers in the different subgroups. Another promising direction would be to investigate how various processes are connected (e.g., whether they compensate, with occurrence of one reducing the chance of occurrence of another). Finally, the results described here focus on school-aged readers -it would be fruitful to investigate whether profiles exist among adult readers and, if so, whether these profiles are different or similar to those observed in school-aged readers.

Reading goals and standards of coherence

As a final illustration of the use of Think-aloud procedures to study variations in reading processes, I summarize several studies that use Think-aloud methods to investigate whether different reading goals trigger different patterns of processing and whether skilled readers adjust their processing more effectively to changes in goals than do less-skilled readers. Much research on the effects of reading goals – whether with Think-aloud or with other methods – is based on the seminal work by Lorch et al. (1993) who used questionnaires to investigate readers' (in this case, undergraduate students) conditional knowledge of what processes fit with the various goals with which they read. Cluster analyses identified 10 categories of reading situations or goals, with reading for study and reading for entertainment being the most distinct from each other. Studies on the effects of reading goals, including the Think-aloud studies reported here, therefore usually contrast reading for study and reading for entertainment (or some variation thereof).

In these Think-aloud studies, undergraduate students read texts (narratives, expository texts, or both) in individual sessions, one sentence at a time, articulating their thoughts after each sentence. Their reading goal was manipulated via instructions and, sometimes, via props (e.g., sitting on a couch with a coffee table with magazines for the entertainment condition, sitting at a desk with study books for the study condition; e.g., Bohn-Gettler & Kendeou, 2014; Van den Broek et al., 2001).

The results show that readers that read expository texts with a study goal focus more strongly on the text (as shown by repetitions and paraphrases) than do readers with an entertainment goal. They also recruit more background knowledge to create coherence (as shown by more explanatory elaborations and predictive inferences) than do readers with an entertainment goal. Conversely, readers with an entertainment goal produce more free associations and evaluations than do readers with a study goal (Narvaez et al., 1999; Van den Broek et al., 2001). Similar effects of reading goals were present during reading of narratives, but the pattern was not as strong (Narvaez et al., 1999). Thus, reading of expository texts appears more sensitive to the goal for reading.

Reading for study of expository texts is cognitively challenging. This is reflected in the finding of Think-aloud studies that when readers with low working-memory capacity read for study, they emphasized less demanding processes (e.g., text repetitions) over more demanding processes (e.g., comprehension monitoring, coherence-building) to a greater extent than readers with high working-memory capacity did – and recalled less. When reading for entertainment, patterns of processes and recall were similar for high and low working-memory capacity readers, except that readers with low working-memory produced more free associations (Linderholm & van den Broek, 2002). Bohn-Gettler and Kendeou (2014) similarly observed that high working-memory readers more effectively adjust their processing to reading goal than do low working-memory readers. Readers with high working memory engaged in more coherence-building and fewer non-coherence-building processes when reading with a study goal than when reading for entertainment. This differentiation of processes as a function of goals did not occur for readers with low working memory. Together, these think-aloud results indicate that both low and high working-memory capacity readers adjust their processing to fit reading goals but do so in different ways, with the adjustments in cognitive processes and strategies by readers with low working-memory capacity being less effective when reading for study and leading to poorer recall (Bohn-Gettler & Kendeou, 2014; Linderholm & Van den Broek & Kendeou, 2022; cf. similar findings for readers who are processing text in a second language, Horiba, 2010).

In summary, the results of this set of Think-aloud studies indicate that inference generation during reading is partly strategic and is influenced systematically by reading purpose. They support the notion that goals influence readers' *standards of coherence*, which in turn influence the types of inferences they draw and the memory representations that they construct (Van den Broek et al., 2011; Van den Broek & Helder, 2017). Different goals, such as reading for entertainment and reading for study, elicit different patterns of processing. These differences appear to be greater for expository than for narrative texts, and stronger – and more effective – for readers with high working-memory capacity than for readers with low working-memory capacity. Future research could investigate other individual differences (e.g., comprehension skill, topic familiarity, motivation) and other reading goals (e.g., literary critique, identifying author intent).

Concluding remarks on the example studies

These diverse studies illustrate how Think-aloud methods can be used to study variations between individual readers and within readers as a function of reading skill, processing profile, and reading goal, and the kinds of choices that researchers can make. I have focused on these particular examples because I was closely involved in these studies and, therefore, am familiar with the choices, deliberations, and decisions in their design. It is important to point out that the particular choices made in these studies should not be interpreted as the gold standard or as the 'right' way of conducting such research. The choices reflect the goals and research questions of these particular studies and considerations to balance the advantages and limitations posed by each option. Studies with different goals

and research questions may warrant different choices for decision points. In the next section I discuss the decision points more systematically.

Think-aloud procedures in reading research: opportunities, challenges and decision points

The brief overviews of empirical research using Think-aloud methods illustrate how Think-aloud methods can be a fruitful method for capturing the processes and strategies by which readers, ranging from elementary-school children to adults, understand a text. Think-aloud protocols give insight into the actual content of such processes and strategies in ways that are very difficult to attain through indirect empirical paradigms such as those involving reading times, eye movements, electromagnetic recordings, and so on.

The examples focused on processing differences between and within individual readers, but we have used similar approaches to investigate other issues in the field of discourse comprehension. Examples include investigations of the effect of reader-misconceptions and of confronting such misconceptions through refutation texts (e.g., Kendeou & van den Broek, 2005; Van den Broek & Kendeou, 2008), of updating the mental representation of a text (e.g., Blanc et al., 2008) and of the effects of topic interest on reading processes (Clinton & van den Broek, 2012). However, Think-aloud methods are being used creatively and productively by many other researchers as well. In the following sections I therefore also refer to studies by others. It is important to reiterate that the optimal choices in one study may not be optimal in another. Instead, in designing Think-aloud studies it is crucial to consider the research questions and the target population (e.g., elementary-school children, high-school students, adults; struggling or proficient readers), and find a balance that optimally allows answers to these questions.

Think-aloud methods have great potential, but they also pose challenges to methodology and interpretation. Some of these challenges are practical and not very different from those using other empirical paradigms: make sure the environment is quiet, prevent interruptions and distractions, and so on. Other practical tips and pointers are specific to the Think-aloud paradigm: audio/video record responses, transcribe the recordings completely before categorizing the responses, and so on (e.g., see Charters, 2003; Hu & Gao, 2017; Noushad et al., 2023). In the following sections I focus particularly on substantive challenges and choices that directly affect the content of the results and the conclusions the researcher draws. I start with the importance of a research perspective, then turn to aspects of the implementation of a Think-aloud procedure, and conclude with aspects concerning categorization and analysis of the responses. Table 2 lists the decision points that are discussed here.

Research questions and theoretical framework

As with any research, it is important to specify the research questions and theoretical framework of the study beforehand. For Think-aloud studies this is particularly important because of the relatively noisy nature of the data (i.e., the responses). As mentioned above, responses can be

Table 2. Decision points for think-aloud studies discussed in this article.

Research questions and theoretical framework
Frequency and timing of responses
Concurrent versus retrospective think aloud
Instructions, modeling and practice for participants
Criterial task and reading goals
Prompting to encourage responding
Recording and transcribing think-aloud responses
Parsing and coding think-aloud responses
Confounding factors and triangulation of responses
Connecting think-aloud responses to offline reading measures

analyzed in many different ways: Think-aloud procedures are used to inventory many different types of reader activities (e.g., comprehension as well as non-comprehension activities), and responses can be categorized in many different ways (e.g., at different grain sizes, considering only the first response to a text input or also subsequent responses, via different coding schemes). It would be possible to devise such inventories and categorizations in a post hoc, bottom-up fashion: See what kinds of responses occur and then construct ad hoc categories. But responses generally are too varied and noisy to do so in a transparent and replicable manner.[4] Therefore, both the instructions given to the readers and the categorization of responses usually are best based on an a priori and, preferably, theoretical perspective. The research questions and theoretical framework should be concrete and well-articulated so that they can be applied reliably by coders and, in replications, by other researchers. Besides the importance for analysis of the responses, the research questions and theoretical framework also drive decisions about the practical implementation of the Think-aloud procedure, as we will see in points below.

The studies reviewed above build on the Landscape model and other theoretical models of meaning construction from texts (see McNamara & Magliano, 2009; Van den Broek & Kendeou, 2022). As described in the preceding section, the Landscape model provides the framework for the specific categorizations of responses according to different processes postulated by the theories (e.g., text repetitions, paraphrases, valid/invalid explanatory elaborations, associative activations, valid/invalid predictive inferences, metacognitive and affective responses, etc.) and also for grouping them into superordinate categories such as processes that stay close to the text, that draw on readers' background knowledge processes, and that represent reflective 'meta comments.' In addition, the model guided the level of text analysis and the frequency of responses (i.e., after each sentence). In this fashion, the theoretical framework was crucial in providing order in the data and allowing interpretation of the results in the context of comprehension of texts. Of course, other studies will have different research questions and, hence, other theoretical models that guide their decisions. Examples include using Think-aloud methods to investigate the moment-by-moment emotion processes that occur during knowledge revision (e.g., Trevors et al., 2017), to investigate cognitive processes in multiple text comprehension (e.g., Bråten & Strømsø, 2003; Strømsø et al., 2003), to assess reading skill (e.g., Magliano & Millis, 2003), to investigate reader misconceptions and the use of refutation texts to overcome misconceptions (e.g., Kendeou & van den Broek, 2005; Van den Broek & Kendeou, 2008), to investigate updating the mental representation of a text (e.g., Blanc et al., 2008), to study the effects of topic interest (Clinton & van den Broek, 2012) and of text complexity (Dahl et al., 2021) on cognitive processes in reading. The important point here is that studies that use Think-aloud procedures benefit tremendously from a clear, precise and a priori articulation of theoretical perspective and accompanying research goals. I now turn to decision points about the concrete implementation of Think-aloud procedures.

Frequency and timing of responses

The frequency and timing by which the participants in a Think-aloud study are asked to report their thinking can vary. In our studies, participants were asked to provide their thoughts after every sentence they read, but this is not carved in stone. A first decision is whether the researcher determines the think-aloud locations or the participants themselves determine when to think aloud. The experimenter-determined scenario is most widely used in discourse and educational psychology and is particularly suitable for investigating processing strategies associated with mental model construction, whereas the self-nomination scenario may be most suitable for investigations of metacognitive strategies (e.g., Magliano, 1999; Pressley & Afflerbach, 1995).

With respect to experimenter-determined locations, these can be after regular intervals: after every sentence or every nth sentence, but also after meaningful units other than sentences (e.g., paragraphs or longer sections), or conceivably after a smaller unit than a sentence (e.g., clause). Or they can be at

locations that are preselected (and not necessarily at regular intervals), usually based on theory or a discourse analysis (Magliano & Millis, 2003).

Another decision pertains to whether the text remains available to the participant when thinking aloud. In most studies, this is the case but it also is possible to remove the text, thereby requiring participants to rely on their mental model to generate responses (Gilliam et al., 2007).

Several practical and conceptual considerations determine the choices. The more frequently participants report their thoughts, the more data the researcher obtains and, hence, the more stable and reliable observed patterns are. In addition, the responses are closer in time to the actual processes that take place as the reader progresses through the text than when responses are elicited only after larger text sections, such as a paragraph or entire text (Magliano & Millis, 2003). In the latter case the responses are partly retrospective (see below) as they require the participant to recreate the processing that occurred at the beginning of the paragraph or text. Frequent probing also has downsides, however. Aside from the fact that data collection is more time consuming, it also means more frequent interruptions, possibly altering the normal reading process. With respect to the timing of responses, conceptual considerations are important, including considerations of the type of processing one aims to investigate (e.g., representation-building/metacognitive, fine-grained/global, continuous/targeted processes).

There is no perfect solution, the researcher has to balance the advantages and disadvantages of each. Conceptual considerations may be important in deciding. For example, one reason we chose responses after every sentence is that our conceptual questions pertained to comprehension- and coherence-building processes as they unfold as the reader proceeds through the text. If research questions are aimed, for instance, at integrative processes, then responses after each paragraph or section would have been more logical. Thus, in deciding on the timing and frequency of responses in a Think-aloud study one should carefully weigh practical, methodological and conceptual considerations.

Concurrent versus retrospective think aloud

The aim of Think-aloud procedures is to provide insight into the reader's processes during reading, but the 'during reading' part can be operationalized in different ways. Frequently, a distinction is made between concurrent Think-aloud (verbalizing during reading) and retrospective Think-aloud (verbalizing after reading has been completed). In retrospective Think-aloud, participants verbalize their thoughts after completing reading. They reflect on what they did during reading *from memory* or *prompted by a recording* of themselves performing the reading task (e.g., a video/eye-movement record; e.g., Salmerón & Llorens, 2019). Concurrent Think-aloud is considered more appropriate for capturing ongoing thought processes in working memory (e.g., Ericsson & Simon, 1993). It should be noted, however, that in most implementations Think-aloud participants report after each sentence or after a short text segment (e.g., a couple of sentences) rather than truly concurrent with reading. It may be more appropriate to call this 'near concurrent.' Nevertheless, the more simultaneous reporting and actual reading are, the more likely that the responses reflect immediate processes. The decision about concurrent/retrospective overlaps with that on timing and frequency mentioned above.

Instructions, modeling and practice for participants

At the outset of studies on reading comprehension, participants are instructed what to do with text(s) they are about to read. Besides explanations of practical details (e.g., the types of texts, how texts will be presented), they are told what they need to do and how to do that. Instructions for reading can vary in the extent to which they direct the think-aloud responses. They may be neutral (e.g., "tell us everything that comes to mind," in line with the recommendations by Ericsson & Simon, 1993) or they may instruct to provide certain types of responses (e.g., "tell us about the spatial layout of the situation that is described" or "at each point tell us about the emotions the protagonist feels"). In some circumstances, especially for younger children, it may be useful that a researcher or other child models the

procedure. Of course, modeling will steer the responses of participants in the direction that is being modeled. If the intent is to elicit particular types of responses then that may not be a problem, but if the intent is to obtain spontaneous and unfiltered responses then modeling should be as neutral as possible (e.g., by including instances of all response categories). If appropriate, it is important to make clear that there are no correct or incorrect responses. Similar considerations apply to deciding whether the children practice and receive feedback.

In the reviewed studies, participants were instructed to read each sentence (usually on a computer screen but for the younger children on booklets with one sentence per page) and say what they were thinking before moving on to the next sentence. Instructions were neutral (in line with Ericsson & Simon, 1993) and explicitly mentioned that there were no correct or incorrect responses. In some of the studies (with the younger children) the test leader modeled the Think-aloud procedure for the first half of a practice text following a script to ensure all participants received the same examples (e.g., general examples of close-to-the-text, inferential and meta comment responses). Then the participant practiced with the second half of the practice text to ensure they understood the procedure. In other studies, we did not include modeling to obtain converging evidence without influencing the children's responses.

Criterial task and reading goals

The processes that readers will execute are influenced by the criterial task that they receive (e.g., Bohn-Gettler & McCrudden, 2018; Bråten et al., 2018; Higgs et al., 2023; Jenkins, 1979; McCrudden, 2019; McCrudden et al., 2011; Van den Broek et al., 1973; Walsh & Jenkins, 1973). The criterial task gives them the goal with which they read, usually in the form of a task they are to perform after having read the text(s) (e.g., summarize the text(s), recall the text(s), answer comprehension questions, apply the textual information). In many reading situations, particularly outside the lab, readers establish their own goal for reading. Based on the criterial task or goal for reading, participants construct a task model, which directs their processing (e.g., Rouet et al., 2017).

The choice of task or reading goal usually reflects the research questions and theoretical framework. It is worth giving careful thought to the selection of a criterial task, as even with a particular research question and theoretical perspective in mind there usually are multiple options, with subtle effects on reading behavior. For example, for studies aimed at identifying comprehension processes it is common to ask comprehension questions after reading the text(s). Alternatives are to use tasks such as recall, summarization, prepare for giving a presentation, and so on. Importantly, these various instances of comprehension tasks elicit different processes during reading as evidenced, for instance, in eye movements (e.g., Kaakinen & Hyönä, 2005; Yeari et al., 2015).

Consideration should also be given to the points in the procedure at which the criterial task is performed. If the task is performed after each of the multiple texts this will surely influence processing of all texts after the first one. Consider, for example, a design on which comprehension questions are used as the criterial task. Comprehension questions can take various forms: they can be about literal information in the text, about integration of multiple parts of the text, about inferences that draw on background knowledge, and so on. Once posed for one text, the type of comprehension question likely will impact processing of subsequent texts.

The reading goal of participants can be manipulated by explicitly mentioning what their task will be after reading and also by encouraging the reader to keep a particular perspective in mind while reading without necessarily indicating what task they will be asked to perform after reading is completed. A well-known example of the latter consists of studies on the burglar/homebuyer perspectives (originally by Pichert & Anderson, 1977).

The general point is that the task that readers anticipate or the reading goal they hold will have a significant impact on the processes in which they engage (for more extensive discussions of the effects of criterial task, see Higgs et al., 2023; McCrudden et al., 2011; see also the above exemplar studies on the effects of reading goals). In the sets of exemplar studies reviewed earlier, various tasks

were used with the intent of arriving at conclusions that were at least somewhat generalizable across tasks.

Prompting to encourage responding

Researchers need to decide whether to engage in prompting when a reader does not give a Think-aloud response. One possibility is to accept when that occurs as a 'no response' and continue the reading. Another possibility is to provide neutral encouragement to think aloud. Yet another possibility is to provide more explicit prompts (e.g., "can you explain that?"). The latter may help understand what the participant means, but it may also steer the content of the response in a particular direction.

In the reviewed studies we chose for a neutral encouragement (e.g., "what are you thinking now?" or "don't forget to tell us what you think"), in some studies whenever a no-response occurred, in other studies after several no-responses. We chose for this mild form of prompting because children in particular may forget to think aloud or may lack the confidence to express their thoughts if they are not certain they are doing it 'correctly' (despite emphasis that there are no right or wrong responses). This choice reflects a compromise between children's spontaneity and researcher directiveness. Here too, different choices would be defensible as well.

Recording and transcribing think-aloud responses

In most Think-aloud studies, participants are asked to provide their responses by stating them aloud. These responses are audio/video recorded or synchronously written/typed out by a researcher. Of these methods we prefer audio/video recordings. They allow the participants to express their thinking quickly and spontaneously, and without the cognitive load that writing/typo in answers would cause. The latter is of course especially the case with young children or participants with weak language skills. Moreover, audio/video recordings give a complete and infiltered record of the real-time responses that can be reinspected at a later time, if necessary.

Transcribing the responses would seem a straightforward step, but it still involves decisions. It is advisable to transcribe the responses completely, without editing (except for typographical errors in written records, mispronunciations in auditory records, and the like), to avoid having to re-transcribe when categorization needs to be adjusted. One decision concerns whether to include filled/unfilled pauses or interjections (e.g., "mmmh," "what shall I say?") or to include corrected responses (e.g., false starts). In the studies I reviewed these all were initially included in the transcriptions but not in the categorization and analysis that followed.

An alternative to audio/video recording the responses is to have participants write or type their responses (e.g., Muñoz et al., 2006). This approach reduces the need for transcribing (although the written/typed responses still need some clean up, because of typos, unclear handwriting, unfinished sentences, etc.) and allows for large data collection. It puts a heavier demand on participants' expressive (and typing) skills.

Parsing and coding think-aloud responses

Before the responses can be coded according to the categorization scheme, they need to be parsed into units that can be scored. One approach is to parse the responses into idea units (e.g., Trabasso & Magliano, 1996). Subsequent coding of idea units is typically mutually exclusive such that an idea unity can only reflect one coding category. But for practical reasons (e.g., studies with large samples), one can choose to analyze responses at a more general level by scoring each response as a whole along each of the coding categories (e.g., on a 3-point scale. See Magliano et al., 2011; McCarthy et al., 2021).[5]

A further decision needs to be made about what part or parts of each response to enter into the categorization process. This is particularly relevant when responses have incomplete sentences or

contain multiple ideas. One possibility is to have judges determine what the predominant part or essence of each response is. In the case of incomplete sentences, this may involve inferring the missing part of the sentence. It is advisable to have multiple judges, to allow for a reliability check. It may also be helpful to first 'translate' the responses into (implied) subject-verb clauses before determining and categorizing the predominant parts of the responses.

Responses frequently include multiple ideas. For example, a participant may report an initial thought followed by further thoughts – either in a single sentence or in separate sentences – or a participant reports a thought and then corrects it. It is possible to categorize all these ideas or to only categorize the initial response. The initial response is thought to indicate spontaneous thoughts in response to the text itself, whereas subsequent responses may reflect other, more reflective thoughts (e.g., Ericsson & Simon, 1993; Hertzum et al., 2009). An advantage of focusing on the initial response is that this procedure results in equal numbers of responses for all participants, thereby making comparisons between participants possible. If, however, the research questions involve constructed responses (e.g., explanations or answers to specific questions; McNamara et al., 2023) or include hypotheses about 'constructive reading,' the extent to which readers are able to engage in further reflection on the text, then coding all of multiple responses (separately or as an ensemble) is useful (e.g., Pressley & Afflerbach, 1995; Trabasso & Magliano, 1996a, 1996b).

In the example studies, the primary interest was on the immediate processes in which the reader engage. Because of this theoretical focus, think-aloud responses were parsed into single sentences/clauses and only first responses were included in the categorization and analyses. In a few instances, separate analyses were performed on all responses, to doublecheck whether the observed patterns would change. In these instances, results were the same for analyses with only first responses and analyses with all responses.

It is worth noting that computerized coding schemes are being developed to score think-aloud responses for the presence of some processing strategies. A prime example is R-SAT, which makes use of well-developed semantic analysis techniques (e.g., Magliano et al., 2011; Magliano & Millis, 2003; Muñoz et al., 2006). Comparisons between human and computerized think-aloud analyses show considerable overlap, as well as some trade-offs (e.g., computerized analysis is more consistent whereas human analysis provides richer and more detailed results; see Yan et al., 2020). Other computerized methods have been developed with a focus on automated scoring of think-aloud protocols at the level of sentence processing (e.g., Yoo, 2024), on assessing the cohesion of the provided think-aloud responses (Ness-Maddox et al., 2025) or on the kinds of background knowledge a reader recruits (e.g., Cromley et al., 2024).

Aside from practical and conceptual considerations in making decisions on the parsing and coding, there is a methodological consideration as well. Several of the choices (e.g., parsing of units, including all or only the firsts (parts) of the responses, allowing each unit to be assigned to only one or more coding categories) have impact on the kinds of statistical analyses one will be able to perform. It is important to consider the requirements for the anticipated analyses when making decisions about parsing and coding.

Finally, once these choices are made, they need to be implemented in a reliable and consistent manner. Discussion of methods for doing so is beyond the scope of this paper, but involve the development of a clear categorization scheme for the coders, training of coders and building reliability in their coding, refinement of the categorization scheme based on insights gained from the training, and so on. The coding system development, training, and reliability phases are iterative (for an overview see McCarthy et al., 2021)

Confounding factors and triangulation of responses

Think-aloud procedures have inherent limitations, especially because they require participants to construct responses. Limitations include possible interference of the Think-aloud procedure with natural processing, fatigue, dependence on verbal abilities, focus only on processes of which the

participants are aware (e.g., Charters, 2003; Ericsson & Simon, 1993; Hertzum et al., 2009; Hu & Gao, 2017; Noushad et al., 2023). These limitations are relevant for all studies using Think-aloud procedures but become magnified in studies on individual differences. The limitations may interact with the targeted individual difference variables, becoming confounding factors. For example, if the aim is to distinguish processing differences between good and poor readers, then the potential influence of verbal abilities and expressive skills likely will be confounded with the effect of reading skill. Likewise, attentional demands, fatigue, and willingness to express oneself may materialize differently for younger than for older participants.

As with all research methods, it is important to be aware of these limitations, and to try and mitigate them as much as possible. There are no perfect solutions to these challenges, but one approach is to triangulate the results and interpretations of Think-aloud procedures with results from other experimental methods (e.g., three-pronged approach; Magliano, 1999; Magliano & Graesser, 1991). For example, in some of the research on reader profiles, the think-aloud results are replicated and extended via eye-movement recordings (e.g., Kraal et al., 2019; for triangulation with adult readers, see Blanc et al., 2008; Kaakinen & Hyönä, 2005). A weakness of these and other studies is that the triangulation takes place as a between-participants comparison or a within-participant comparison with different texts. As a consequence, conclusions can only be groupwise (e.g., about the existence of reading profiles). Future research that uses concurrent recordings of think-aloud responses and eye movements (or other triangulating measures) would be ideal.

Connecting think-aloud responses to offline reading measures

In the above, I have focused on the use of Think-aloud procedures to inventory the processes in which readers engage while reading. For several reasons it may be useful to try and connect the online Think-aloud response patterns to aspects of the offline representation that the readers have constructed by the time they have completed reading. This may be theoretically important when the research questions include hypotheses about the connection between reading processes and reading products. It also may provide an alternative way to achieve triangulation. A challenge in investigating this possible connection is to develop or select an offline task that is fine-grained enough to be sensitive to effects of processes at the level of detail that Think-aloud procedures provide. For example, in the reviewed studies local comprehension processes were the target of investigation but such processes would likely not significantly alter overall scores on, for instance, a recall task. Indeed, although some think-aloud results show a connection to differences in performance on the post-reading task for children (e.g., reading profiles and reading skill, Kraal et al., 2018; 4th and 6th grade students' reading processes and recall; Coté et al., 1998) and adults (e.g., Magliano et al., 2011; Magliano & Millis, 2003), the connections are not always very clear; Other studies have not observed effects of processing variations on post-reading measures (e.g., Kraal et al., 2019; see Rapp & Mensink, 2011, for a discussion). If part of the research questions concerns the connection between reading processes and reading products, then careful consideration should be given to developing or selecting a fitting measure of the representation after reading is completed.

Using think-aloud procedures in reading research: discussion

The aim of this paper is to provide reading researchers with food for thought on how to use Think-aloud procedures to investigate the processes that readers employ as they are reading text. By using Think-aloud procedures, one can gain insight in the actual content of the processing – unlike other processing measures that we and others have used and that only indirectly point to specific processes, such as eye-movements, reading times, reaction times and electromagnetic brain recordings. But Think-aloud procedures have their limitations that pose challenges for methodology and for interpretation of results (e.g., Charters, 2003; Hertzum et al., 2009; Hu & Gao, 2017; Noushad et al., 2023). Perhaps the most obvious limitation is that responses necessarily are restricted to reading processes of

which the reader is aware, which calls for efforts to triangulate conclusions by replication via other methods that do not rely on conscious awareness. In our studies we sometimes have used eye-tracking for such triangulation but other methods are possible as well.

When designing a Think-aloud procedure, the researcher makes numerous substantive choices that directly affect the content of the results and the conclusions that can be drawn. The impact of these choices often is even larger when developmental, interindividual and intraindividual differences are considered. In this paper I discuss the most important of these choices, touching on the importance of adopting a clear a priori theoretical research perspective, on aspects of the implementation of a Think-aloud procedure, and on aspects concerning categorization and analysis of the responses. The intent is to provide considerations for making informed methodological choices that optimize the potential and mitigate the challenges of Think-aloud procedures. These same considerations can provide important nuances to interpretations of the results. The choices are relevant for reading studies in general, but some are particularly relevant for studies on how online processing of texts may differ between and within individual readers.

To illustrate these decision points, I reviewed Think-aloud studies on several (intra)individual differences in text comprehension. The first set of studies showed that there appear to exist subgroups of readers that differ systematically in their profile of processing: *paraphrasers* that focus their processing on the content of the texts itself (repeating and paraphrasing the text) and *elaborators* that draw much more on their background knowledge to generate explanations of what they read and predictions about what is to follow in the text. These patterns were observed across a wide age range of elementary and secondary school children and for both narrative and expository texts.

The second set of studies showed that individuals with weak comprehension often have a paraphraser profile. There also are individuals with weak comprehension that have an elaborator profile but their problem seems to be that they often make incorrect elaborations. The third set showed that online processes revealed by Think-aloud procedures are influenced systematically by reading goal. They support the notion that goals influence readers' *standards of coherence*, which in turn influence the types of inferences they draw and the memory representations that they construct (e.g., Van den Broek & Helder, 2017).

Several of the decision points discussed in this paper do not have a perfect solution. It is imperative to carefully consider options and ultimately find a good balance in light of the purposes of the study. Here too replication and triangulation may be useful. I hope that this paper will serve as a fruitful starting point for deliberations on methodological and theoretical aspects of the design of Think-aloud studies and lead to creative and productive investigations of reading processes.

Notes

1. I use the plural 'we' when referring to collaborative research with colleagues and/or students.
2. Basic reading processes such as letter- and word-identification also are passive, at least for practiced readers (e.g., Kendeou et al., 2009; McNamara et al., 2023; Perfetti & Helder, 2022), but here I focus on processes specific to discourse comprehension
3. E.g., Gates-McGinitie Reading Comprehension Test (MacGinitie et al., 2002), Curriculum-Based Measurement (Wayman et al., 2007), national CITO test (Feenstra et al., 2010). In these studies, scores were used to establish a dichotomy (i.e., good and struggling comprehenders). With the advent of multilevel analysis methods it now is possible to investigate reading-comprehension skill as a continuum (e.g., Feller et al., 2020, 2024).
4. However, use of think-aloud procedures with a post-hoc categorization schemes may be useful as an exploratory step in a first investigation of an issue in reading. Having an a priori framework is especially useful when the aim is to investigate particular relations or hypotheses.
5. I thank Joe Magliano for raising this issue.

Acknowledgment

Thanks to Joe Magliano and two anonymous reviewers for their thoughtful comments on an earlier version of this paper.

Disclosure statement

No potential conflict of interest was reported by the author.

References

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(5), 1061–1070. https://doi.org/10.1037/0278-7393.19.5.1061

Blanc, N., Kendeou, P., van den Broek, P., & Brouillet, D. (2008). Updating situation models during reading of news reports: Evidence from empirical data and simulations. *Discourse Processes*, *45*(2), 103–121. https://doi.org/10.1080/01638530701792784

Bohn-Gettler, C. M., & Kendeou, P. (2014). The interplay of reader goals, working memory, and text structure during reading. *Contemporary Educational Psychology*, *39*(3), 206–219. https://doi.org/10.1016/j.cedpsych.2014.05.003

Bohn-Gettler, K., & McCrudden, M. T. (2018). The effects of task relevance instructions and topic beliefs on reading processes and memory. *Discourse Processes*, *55*(4), 410–431. https://doi.org/10.1080/0163853X.2017.1292824

Bråten, I., McCrudden, M. T., Lund, E. S., Brante, E. W., & Strømsø, H. I. (2018). Task-oriented learning with multiple documents: Effects of topic familiarity, author expertise, and content relevance on document selection, processing, and use. *Reading Research Quarterly*, *53*(3), 345–365. https://doi.org/10.1002/rrq.197

Bråten, I., & Strømsø, H. I. (2003). A longitudinal think-aloud study of spontaneous strategic processing during the reading of multiple expository texts. *Reading and Writing: An Interdisciplinary Journal*, *16*(3), 195–218. https://doi.org/10.1023/A:1022895207490

Britt, M. A., Rouet, J.-F., & Durik, A. (2017). *Literacy beyond text comprehension: A theory of purposeful reading*. Routledge. https://doi.org/10.4324/9781315682860

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

Charters, E. (2003). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education Journal*, *12*(2). https://doi.org/10.26522/brocked.v12i2.38

Chi, M. T. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences*, *6*(3), 271–315. https://doi.org/10.1207/s15327809jls0603_1

Clinton, V., & van den Broek, P. (2012). Interest, inferences, and learning from texts. *Learning and Individual Differences*, *22*(6), 650–663. https://doi.org/10.1016/j.lindif.2012.07.004

Clinton-Lisell, V., Carlson, S. E., Ness-Maddox, H., Dahl, A., Taylor, T., Davison, M. L., & Seipel, B. (2024). Identifying clusters of less-skilled college student readers based on cognitive processes. *Journal of College Reading and Learning*, *54*(2), 125–141. https://doi.org/10.1080/10790195.2024.2359513

Coté, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes*, *25*(1), 1–53. https://doi.org/10.1080/01638539809545019

Cromley, J. G., Mirabelli, J. F., & Kunze, A. J. (2024). Three applications of semantic network analysis to individual student think-aloud data. *Contemporary Educational Psychology*, *79*, 102318. https://doi.org/10.1016/j.cedpsych.2024.102318

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

Elfenbein, A. (2018). *The Gist of Reading*. Stanford University Press.

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

Feenstra, H., Kleintjes, F., Kamphuis, F., & Krom, R. (2010). *Begrijpend Lezen groep 3 t/m 6 LOVS*. Cito B.V.

Feller, D. P., Magliano, J. P., Sabatini, J., O'Reilly, T., & Kopatich, R. D. (2020). Relations between component reading skills, inferences, and comprehension performance in community college readers. *Discourse Processes*, *57*(5–6), 473–490. https://doi.org/10.1080/0163853X.2020.1759175

Feller, D. P., Sabatini, J. P., & Magliano, J. P. (2024). Differentiating less prepared from more prepared college readers. *Discourse Processes*, *61*(4–5), 180–202. https://doi.org/10.1080/0163853X.2024.2319515

Gilliam, S., Magliano, J. P., Millis, K. K., Levinstein, I., & Boonthum, C. (2007). Assessing the format of the presentation of text in developing a Reading Strategy Assessment Tool (R-SAT). *Behavior Research Methods, Instruments, & Computers*, *39*(2), 199–204. https://doi.org/10.3758/BF03193148

Hattan, C., Alexander, P. A., & Lupo, S. M. (2024). Leveraging what students know to make sense of texts: What the research says about prior knowledge activation. *Review of Educational Research*, *94*(1), 73–111. https://doi.org/10.3102/00346543221148478

Hertzum, M., Hansen, K. D., & Andersen, H. H. K. (2009). Scrutinising usability evaluation: Does thinking aloud affect behaviour and mental workload? *Behaviour & Information Technology*, *28*(2), 165–181. https://doi.org/10.1080/01449290701773842

Higgs, K. P., Santuzzi, A. M., Gibson, C., Kopatich, R. D., Feller, D. P., & Magliano, J. P. (2023). Relationships between task awareness, comprehension strategies, and literacy outcomes. *Frontiers in Psychology*, *14*, 1056457. https://doi.org/10.3389/fpsyg.2023.1056457

Horiba, Y. (2010). Reader control in reading: Effects of language competence, text type, and task. *Discourse Processes*, *29* (3), 223–267. https://doi.org/10.1207/S15326950dp2903_3

Hu, J., & Gao, X. (2017). Using think-aloud protocol in self-regulated reading research. *Educational Research Review*, *22*, 181–193. https://doi.org/10.1016/j.edurev.2017.09.004

Jenkins, J. (1979). Four points to remember: A tetrahedral model of memory experiments. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 429–446). Erlbaum.

Kaakinen, J. K., & Hyönä, J. (2005). Perspective Effects on Expository Text Comprehension: Evidence From Think-Aloud Protocols, Eyetracking, and Recall. *Discourse Processes*, *40*(3), 239–257. https://doi.org/10.1207/s15326950dp4003_4

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. (2014). Validation and comprehension: An integrated overview. *Discourse Processes*, *51*(1–2), 189–200. https://doi.org/10.1080/0163853X.2013.855874

Kendeou, P., & van den Broek, P. (2005). The effects of readers' misconceptions on comprehension of scientific text. *Journal of Educational Psychology*, *97*(2), 235–245. https://doi.org/10.1037/0022-0663.97.2.235

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, *101*(4), 765–778.

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

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

Linderholm, T., & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology*, *94*(4), 778–784. https://doi.org/10.1037/0022-0663.94.4.778

Long, D. L., & Bourg, T. (1996). Thinking aloud: Telling a story abouta story. *Discourse Processes*, *21*(3), 329–339. https://doi.org/10.1080/01638539609544961

Lorch, R. F., Lorch, E. P., & Klusewitz, M. A. (1993). College students' conditional knowledge about reading. *Journal of Educational Psychology*, *85*(2), 239–252. https://doi.org/10.1037/0022-0663.85.2.239

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2002). *Gates-macGinitie reading tests fourth edition: echnical report*. Riverside Publishing.

Magliano, J. P. (1999). Revealing inference processes during text comprehension. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 55–75). ErIbaum.

Magliano, J. P., & Graesser, A. C. (1991). A three-pronged method for studying inference generation in literary text. *Poetics*, *20*(3), 193–232. https://doi.org/10.1016/0304-422X(91)90007-C

Magliano, J. P., & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction*, *21*(3), 251–283. https://doi.org/10.1207/S1532690XCI2103_02

Magliano, J. P., Millis, K. K., Levinstein, I., & Boonthum, C. (2011). Assessing comprehension during reading with the Reading Strategy Assessment Tool (RSAT). *Metacognition and Learning*, *6*(2), 131–154. https://doi.org/10.1007/s11409-010-9064-2

Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology*, *91*(4), 615–629. https://doi.org/10.1037/0022-0663.91.4.615

McCarthy, K. S., Magliano, J. P., Snyder, J. O., Kenney, E. A., Newton, N. N., Perret, C. A., Knezevic, M., Allen, L. K., & McNamara, D. S. (2021). Quantified qualitative analysis: Rubric development and inter-rater reliability as iterative design. In E. de Vries, Y. Hod, & J. Ahn (Eds.), *Proceedings of the 15th International Conference of the Learning Sciences - ICLS 2021*. (pp. 139–146). International Society of the Learning Sciences.

McCarthy, K. S., & McNamara, D. S. (2021). The multidimensional knowledge in text comprehension framework. *Educational Psychologist*, *56*(3), 196–214. https://doi.org/10.1080/00461520.2021.1872379

McCrudden, M. T. (2019). The effect of task relevance instructions on memory for text with seductive details. *Applied Cognitive Psychology*, *33*(1), 31–37.

McCrudden, M. T., Magliano, J., & Schraw, G., Eds. (2011). *Text relevance and learning from text*. Information Age Publishing.

McMaster, K. L., Espin, C. A., & van den Broek, P. (2014). Making connections: Linking cognitive psychology and intervention research to improve comprehension of struggling readers. *Learning Disabilities Research & Practice*, *29* (1), 17–24. https://doi.org/10.1111/ldrp.12026

McMaster, K. L., van den Broek, P., 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., & Magliano, J. (2009). Toward a comprehensive model of comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 297–384). Elsevier Academic Press. https://doi.org/10.1016/S0079-7421(09)51009-2

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*(4–5), 337–362. https://doi.org/10.1080/0163853X.2023.2225757

Muñoz, B., Magliano, J. P., Sheridan, R., & McNamara, D. S. (2006). Typing versus thinking aloud when reading: Implications for computer-based assessment and training tools. *Behavioral Research Methods*, *38*(2), 211–217. https://doi.org/10.3758/bf03192771

Narvaez, D., van den Broek, P., & Ruiz, A. B. (1999). The influence of reading purpose on inference generation and comprehension in reading. *Journal of Educational Psychology*, *91*(3), 488–496.

Nation, K., Clarke, P., & Snowling, M. J. (2002). General cognitive ability in children with reading comprehension difficulties. *British Journal of Educational Psychology*, *72*(4), 549–560. https://doi.org/10.1348/00070990260377604

Ness-Maddox, H., Öncel, P., Hu, S., Luan, L., Magliano, J. P., & Allen, L. K. (2025). Think-aloud verbal protocols and natural language processing to capture coherence building across narrative media. *Discourse Processes*, 1–24. https://doi.org/10.1080/0163853X.2025.2577634

Noushad, B., Van Gerven, P. W. M., & de Bruin, A. B. H. (2023). Twelve tips for applying the think-aloud method to capture cognitive processes. *Medical Teacher*, *46*(7), 892–897.

Olson, G. M., Duffy, S. A., & Mack, R. L. (1984). Thinking-out-loud as a method for studying real time comprehension processes. In D. E. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 253–286). Routledge.

Perfetti, C., & Helder, A. (2022). Progress in reading science: Word identification, comprehension, and universal perspectives. In M. J. Snowling, C. Hulme, & K. Nation (Eds.), *The science of reading: A Handbook* (pp. 5–35). John Wiley.

Pichert, J. W., & Anderson, R. C. (1977). Taking different perspectives on a story. *Journal of Educational Psychology*, *69* (4), 309–315. https://doi.org/10.1037/0022-0663.69.4.309

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Erlbaum.

Rapp, D. N., & Mensink, M. C. (2011). Focusing Effects from online and offline reading tasks. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 141–164). Information Age Publishing.

Rapp, D. N., Van den Broek, P., 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

Rouet, J.-F., Britt, M. A., & Durik, A. M. (2017). RESOLV: Readers' representation of reading contexts and tasks. *Educational Psychologist*, *52*(3), 200–215. https://doi.org/10.1080/00461520.2017.1329015

Salmerón, L., & Llorens, A. (2019). Instruction of digital reading strategies based on eye-movements modeling examples. *Journal of Educational Computing Research*, *57*(2), 343–359. https://doi.org/10.1177/0735633117751605

Seipel, B., Carlson, S. E., & Clinton, V. E. (2016). 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.

Strømsø, H. I., Bråten, I., & Samuelstuen, M. S. (2003). Students' strategic use of multiple sources during expository text reading. *Cognition and Instruction*, *21*(2), 113–147. https://doi.org/10.1207/S1532690XCI2102_01

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

Trabasso, T., & Magliano, J. P. (1996a). How do children understand what they read and what can we do to help them? In M. Graves, P. van den Broek, & B. Taylor (Eds.), *The first R: A right of all children* (pp. 158–181). Columbia University Press.

Trevors, G. J., Kendeou, P., & Butterfuss, R. (2017). Emotion processes in knowledge revision. *Discourse Processes*, *54* (5–6), 406–426. https://doi.org/10.1080/0163853X.2017.1312201

van den Broek, P., Bohn-Gettler, C., Kendeou, P., Carlson, S., & White, M. J. (2011). When a reader meets a text: The role of standards of coherence in reading comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 123–139). Information Age Publishing.

van den Broek, P., Fletcher, C. R., & Risden, K. (1973). Investigations of inferential processes in reading: A theoretical and methodological integration. *Discourse Processes*, *16*(1–2), 169–180. https://doi.org/10.1080/01638539309544835

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., & Kendeou, P. (2008). Cognitive processes in comprehension of science texts: The role of co-activation in confronting misconceptions. *Applied Cognitive Psychology*, *22*(3), 335–351. https://doi.org/10.1002/acp.1418

van den Broek, P., & Kendeou, P. 2022. Discourse comprehension: Inferences and mental representations. In *The science of reading: A handbook (2nd ed.)(pp* M. Snowling, C. Hulme, & K. Nation (Eds.), Wiley 239–260. https://doi.org/10.1002/9781119705116

van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, *29*(8), 1081–1087. https://doi.org/10.3758/BF03206376

van den Broek, P., Mouw, J. M., & Kraal, A. (2015). Individual differences in reading comprehension. In P. Afflerbach (Ed.), *Handbook of individual differences in reading* (pp. 138–150). Routledge.

van den Broek, P., Risden, K., & Husebye-Hartmann, E. (1995). The role of readers' standards for coherence in the generation of inferences during reading. In J. R. F. Lorch & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 353–373). Erlbaum.

van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the online construction of a memory representation. In H. van Oostendorp & S. R. Goldman, (Eds.), *The construction of mental representations during reading* 71–98. Mahwah, NJ: Erlbaum. (Reprinted in R.B. Ruddell & N.J. Unrau (Eds.), Theoretical models and processes of reading (1244-1269). Newark, NJ: International Reading Association)

Van Moort, M. L., Jolles, D. D., Koornneef, A., & van den Broek, P. W. (2020). What you read versus what you know: Neural correlates of accessing context information and background knowledge in constructing a mental representation during reading. *Journal of Experimental Psychology: General*, *149*(11), 2084–2101.

Van Moort, M. L., Koornneef, A., & van den Broek, P. W. (2020). Differentiating text-based and knowledge-based validation processes during reading: Evidence from eye movements. *Discourse Processes*, *58*(1), 22–41. https://doi.org/10.1080/0163853X.2020.1727683

Walsh, D. A., & Jenkins, J. J. (1973). Effects of orienting tasks on free recall in incidental learning: "difficulty," "effort," and "process" explanations. *Journal of Verbal Learning and Verbal Behavior*, *12*(5), 481–488. https://doi.org/10.1016/S0022-5371(73)80028-3

Wayman, M. M., Wallace, T., Wiley, H. I., Tichá, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education*, *41*(2), 85–120. https://doi.org/10.1177/00224669070410020401

Whitney, P., & Budd, D. (1996). Think-aloud protocols and the study of comprehension. *Discourse Processes*, *21*(3), 341–351. https://doi.org/10.1080/01638539609544962

Williams, J. P. (1993). Comprehension of students with and without learning disabilities: Identification of narrative themes and idiosyncratic text representations. *Journal of Educational Psychology*, *85*(4), 631–642. https://doi.org/10.1037/0022-0663.85.4.631

Yan, D., Rupp, A., & Foltz, P. (2020). *Handbook of automated scoring: Theory into practice*. CRC Press/Taylor and Francis.

Yeari, M., & van den Broek, P. (2011). A cognitive account of discourse understanding and discourse interpretation: The landscape model of reading. *Discourse Studies*, *13*(5), 635–643.

Yeari, M., van den Broek, P., & Oudega, M. (2015). Processing and memory of central versus peripheral information as a function of reading goals: Evidence from eye-movements. *Reading and Writing*, *28*(8), 1071–1097. https://doi.org/10.1007/s11145-015-9561-4

Yoo, Y. (2024). Automated think-aloud protocol for identifying students with reading comprehension impairment using sentence embedding. *Applied Sciences*, *14*(2), 858. https://doi.org/10.3390/app14020858

Zwaan, R. A., & Brown, C. M. (1996). The influence of language proficiency and comprehension skill on situation model construction. *Discourse Processes*, *21*(3), 289–327.