Reading comprehension in elementary school children: cognitive studies of the reader, the text, and the task

Karlsson, A.K.J.

Chapter 2

Profiles of Young Readers:

Evidence from Thinking Aloud while Reading Narrative and Expository Texts

This chapter has been published as:
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. doi:10.1016/j.lindif.2018.08.001

Profiles of Young Readers:

Evidence from Thinking Aloud while Reading Narrative and Expository Texts

Abstract

This study aimed to identify reading behavior profiles in nine-to-eleven-year-old children based on their think-aloud responses while reading narrative and expository texts. Three profiles emerged while reading narratives: Literal Readers, who stay close to the literal text by predominantly repeating it; Paraphrasing Readers, who extract meaning from the text by paraphrasing it; and Elaborating Readers, who use background knowledge to explain the text by generating inferences. The three profiles also emerged while reading expository text. Children generally exhibited the same profiles across the two text genres, however, expository texts elicited fewer correct inferences but more invalid inferences than did narratives, suggesting that children are influenced by text demands. Elaborating Readers had better word decoding skills, reading comprehension ability, and non-verbal reasoning ability than readers of the two other profiles, indicating a positive relation between inference generation and language abilities and cognitive resources.

Key words: Reading Profiles, Text Genres, Think Aloud, Developing Readers, Latent Profile Analysis

2.1 Introduction

Readers engage in various reading processes to understand a text. Importantly, readers need to go beyond the literal text and draw upon background knowledge to make inferences to understand the meaning of the text (e.g., Kintsch, 1988; van den Broek, 1990). Young readers differ in their ability to go beyond the text and generate necessary inferences (Cain & Oakhill, 1999; Kendeou, van den Broek, Helder, & Karlsson, 2014; Nation & Snowling, 1997). Such differences may result in children approaching texts in different ways. Studies using think-aloud procedures during reading of narratives revealed contrasting profiles in poor or good comprehending readers; readers in one profile stay close to the literal text, and readers in the other profile generate elaborative inferences that go beyond the text (Carlson, Seipel, & McMaster, 2014; Kraal, Koornneef, Saab, van den Broek, 2017; McMaster et al., 2012; Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007; Seipel, Carlson, & Clinton, 2017). The identification of such reading profiles has led to the development of targeted reading interventions for poor comprehenders (McMaster et al., 2012). However, because reading comprehension is a multidimensional ability, it is important to consider reading profiles in developing readers across the whole range of reading comprehension ability, not only in poor *or* good comprehenders. Furthermore, because narrative and expository texts differ in text demands, it is important to compare readers' profiles for narrative texts to their profiles for expository texts. Considering reading profiles in different text genres may provide useful perspectives for theoretical questions. For example, whether children have a certain stable set of reading abilities with which they process text in a similar way across different situations, or whether they are influenced by conditions such as different text demands. Furthermore, such expansions allow important insights for evidence-based reading instructions across a larger group of developing readers and across different text genres.

2.1.1 Comprehension Processes

A reader can attain different levels of comprehension for a text, ranging from basic to deep understanding. A well-known distinction between such comprehension levels has been proposed in the construction-integration model (van Dijk & Kintsch, 1983; Kintsch, 1988, 1994). In this model three different levels are discussed: the *surface* level, where the reader encodes literal words and phrases, the *textbase,* where the reader understands referential relations within the text, and the *situation model,* where the reader enriches the mental representation of the text by elaborating on it and integrating background knowledge. Although various models on inference generation have been proposed, a consensus has emerged that inferences are important for building a situation model of the text (for a recent overview

see, O'Brien, Cook, & Lorch, 2015). A reader that uses appropriate and global level inferences is more likely to reach beyond the surface level understanding and gains a textbase and situation model understanding of the text (e.g., Goldman, McCarthy, & Burkett, 2015; Graesser, Singer, & Trabasso, 1994; Kintsch, 1994). In developing readers, the ability to make adequate inferences during reading is causally connected to good reading comprehension (Bowyer-Crane & Snowling, 2005; Cain & Oakhill, 1999; Lynch et al., 2008; McGee & Johnson, 2003). Relevant for the current paper, there are three broad types of inferences that contribute to an extended understanding of texts. First, *text-connecting inferences* enable readers to connect a focal event with an event previously mentioned in the text. Text-connecting inferences tend to be routinized in good readers (e.g., McKoon & Ratcliff, 1992; Olson, 1985), and are facilitated by large vocabulary and large working memory (WM) capacity (Singer, Andrusiak, Reisdorf, & Black, 1992). Second, *elaborative inferences* enable readers to connect the text with relevant background knowledge. Elaborative inferences allow for causal connections and are important to create a rich and coherent mental representation of the text (e.g., Graesser et al., 1994; Lynch et al., 2008). Sufficient word reading abilities and world knowledge are some reader characteristics that facilitate the production of valid elaborative inferences (McNamara & Kintsch, 1996; Rapp et al., 2007). Although poor comprehenders also may generate elaborative inferences, these inferences are more often invalid than those of good comprehenders (McMaster et al., 2012). Third, *predictive inferences* are produced when readers predict upcoming events. Predictive inferences are not as routine or critical as the two previously mentioned inferences but rather depend on the text being constraining enough (Cook, Limber, & O'Brien, 2001; Kaakinen & Hyönä, 2005; Klin, Guzmán, & Levine, 1999; van den Broek, 1990). Furthermore, the likelihood of making predictive inferences depends on the interaction of reader characteristics, such as WM capacity, and text characteristics, such as high causality between text parts (e.g., Linderholm, 2002). These three types of inferences may contribute differentially to young readers' ability to process text beyond the literal level and build an enriched mental representation.

Given the positive effects that inference processes have on reading comprehension, it is important to identify whether some children consistently process the text on a basic level whereas others are better able to enrich their mental representation using elaborative inferences. Although much research indicates that good readers generate more inferences than poor readers (e.g., McNamara & Kintsch, 1996), differences have also been found within poor comprehending readers (Rapp et al., 2007). Using a think-aloud procedure, two subgroups of nine-to-ten-year-old poor comprehenders have emerged (Carlson et al., 2014; McMaster et al., 2012; Seipel

et al., 2017). One subgroup of children stayed close to the basic meaning of the text, mainly repeating or paraphrasing the text (Paraphrasers). The other subgroup of children used background knowledge to make inferences, albeit sometimes erroneously (Elaborators). Similar reading profiles have been found in a younger group of Dutch poor comprehending readers, and in their good comprehending peers (Kraal et al., 2017), indicating promising generalizability of reading profiles. Continuing research of reading profiles may help to better understand whether children have a certain approach to process text and whether that approach relates to different levels of text comprehension as described by influential reading models (e.g., Johnson-Laird, 1983; van Dijk & Kintsch, 1983). In the current study we make no a priori distinction between good and poor comprehenders, but aim to identify homogenous subgroups, characterized by their reading behavior, within a heterogeneous population spanning from poor to good comprehension abilities.

2.1.2 Text Genres

Expository texts are often more difficult than narratives for developing readers (e.g., Best, Floyd, & McNamara, 2008), and several reasons may explain differences in text demands (e.g. Eason, Goldberg, Young, Geist, & Cutting, 2012). First, topics and hence familiarity of words may differ between the two text genres. Narratives often include everyday language whereas expository texts often introduce new words and terminology (Medina & Pilonieta, 2006). Therefore, expository texts often have a higher information density. Second, compared to narratives, expository texts are often more varied with regard to their structure (e.g. Lorch, 2015). Narratives often follow a more or less similar structure with similar elements and timelines (such as the protagonists initiating goal, actions, reactions, and outcomes; e.g. Mandler & Johnson, 1977). Expository texts come in different formats and require the reader to apply more varied reading strategies (Lorch, 2015). For example, there is not necessarily a timeline to follow but readers need to understand several subordinate ideas in relation to a main idea (Meyer, 1987). Hence, it is important to understand whether children approach the two text genres differently. Young readers are likely to lack in knowledge of both topic (e.g., Samuelstuen & Bråten, 2005) and text structure (e.g., Williams, Hall, & Lauer, 2004), making it difficult to effortlessly comprehend expository texts. For these reasons, expository texts pose difficulties in making inferences using background knowledge, especially for readers who already lag behind in comprehension skills. Indeed, children with poor inferencing skills experience comprehension difficulties when reading expository text (e.g., Best et al., 2008; Eason et al., 2012; Kraal et al., 2017; Schellings, Aarnoutse, & van Leeuwe, 2006). In adolescents, poor readers generate fewer inferences while reading expository compared to narrative

texts (Denton et al., 2015). However, strategically elaborating on expository texts, if anything, *facilitates* in-depth comprehension of expository texts and, therefore, an *increase* in inference making would be desirable (Lorch, 2015; Mayer, 1996). By comparing inference skills in reading profiles of elementary school children across narrative- and expository texts we may examine whether developing readers recognize different text demands, and identify whether children with a certain reading profile could benefit from more practice with inference generation while reading expository texts.

2.1.3 Reader Characteristics

Because text comprehension is a multidimensional ability, different reading profiles may be related to individual differences in other language abilities and cognitive resources. Individual differences predict reading comprehension in both adult and developing readers (e.g., Hannon, 2012; Language and Reading Research Consortium, & Logan, 2016). In particular, and as mentioned above, good word decoding, reading comprehension skills (e.g. Carlson et al., 2014; Olson, 1985; Rapp et al., 2007), large WM capacity, and vocabulary promotes the ability to make different types of inferences while reading (Linderholm, 2002; Linderholm & van den Broek, 2002; Singer et al., 1992). However, some children's inference problems may be caused by a limited vocabulary (Nation & Snowling, 1998, 1999), whereas others struggling with inference generation may possess enough lexical knowledge but not know how to draw on this knowledge (Bowyer-Crane & Snowling, 2005; Cain & Oakhill, 1999), possibly due to an immature reasoning ability (de Leeuw, Segers, & Verhoeven, 2016; Naglieri, 2001). When tracing the heterogeneity in developing readers back to a number of underlying homogeneous reading profiles, it is important to also map out whether these profiles differ in word decoding, reading comprehension skills, vocabulary, non-verbal reasoning skills, and WM capacity to better understand underlying competences.

2.1.4 Distinguishing within Processes Close to Literal Text

So far, we have focused on the importance of inference making beyond the textual information. However, while participating in think-aloud studies readers also show understanding for the literal meaning of text by producing a substantial number of *text repetitions* and *paraphrases*. Prior research on reading profiles has aggregated text repetitions and paraphrases (Carlson et al., 2014; Kraal et al., 2017; McMaster et al., 2012), but there are reasons to examine whether these contribute differently to young readers' literal and in-depth comprehension of text. Literal repetition of text is a superficial reading strategy used when more in-depth reading

strategies are too demanding: readers with a low WM capacity more often resort to text repetitions at the expense of demanding inferences than readers with a high WM capacity (Linderholm & van den Broek, 2002). Paraphrasing a text has been described as a superficial reading strategy for the same reason; instead of making inferences readers paraphrase the text (Magliano & Millis, 2003; Seipel et al., 2017). However, paraphrasing a text has also proven useful to start up higher order comprehension processes, i.e. facilitating inference generation (e.g., Denton et al., 2015; McNamara, 2004). Rephrasing the concepts in a text into one's own words strengthens memory traces for the text (e.g., Bohn-Gettler & Kendeou, 2014) and makes the focal text and its semantic relations more comprehensible to the reader (Coté, Goldman, & Saul, 1998). Hence, distinguishing between text repetitions and paraphrases may contribute to our knowledge about how young readers differ in their ability to start up inferences to enable building a rich representation of the text.

2.1.5 Current Study

In the current study we use a think-aloud procedure to examine how nine-to-eleven-year-old children approach text. We aim to identify subgroups of children characterized by their profile of think-aloud responses while reading narrative and expository texts. Contrary to previous research focusing on poor or good comprehenders, we include readers across a larger range from poor to good reading comprehension ability. To trace back the heterogeneity in these readers to a set of underlying homogeneous subgroups we use a Latent Profile Analysis (LPA). Individual differences in a set of variables, such as categories of think-aloud responses, are complex and often both quantitative and qualitative in nature. Linear models are not always suitable to capture the complexity of such data. LPA is a powerful and flexible tool, able to identify individuals with similar response patterns in such complex sets of variables (e.g., Bergman & Magnusson, 1997; Hickendorff, Edelsbrunner, McMullen, Schneider, & Trezise, 2017).

First, in the current study, we compare reading profiles across text genres. Previous research shows that young readers have more reading difficulties with expository texts compared to narrative texts. Poor comprehenders, in particular, are less able to make inferences while reading expository texts. In the current study, we compare children's think-aloud responses to narrative texts with those to expository texts, by conducting separate LPAs and investigating the similarities in the profiles of reading behavior at the group level, and the interrelation between the profiles of individual children. Identifying reading profiles across different text genres helps to examine whether developing readers recognize and adjust to different text demands. Second, in light of previous research we anticipate that children differ in their ability to make inferences during a think-aloud task. On the one

hand, research investigating reading profiles shows that this difference occurs within poor comprehenders. On the other hand, other studies on inference making show that good comprehenders generally are better able to generate inferences while reading than poor comprehenders. In the current study we are able to put these findings in relation to one another; by conducting MANOVA's we investigate whether the children with different reading profiles based on think-aloud responses differ in their language abilities (word decoding, reading comprehension, and vocabulary) and in their cognitive resources (non-verbal reasoning ability, and working memory). Third, methodologically we expand on previous research by distinguishing not only between far-from-text processes (i.e. inferences) and close-to-text processes, but also within processes close to the literal text: verbatim repeating and paraphrasing the text. The first may entail text decoding on a surface level whereas the latter may be important to start up inference-generation processes. Hence, we may examine differences in inference generation and also whether readers of different profiles differ in their close-to-text processes.

2.2 Methods

2.2.1 Participants

One hundred and seven children (61 girls) between nine and eleven years old (M = 10.3, SD = 0.73) participated in the current study. Children from grades six and seven were recruited from 16 different schools in the south-west of the Netherlands. Parents gave written consent for participation and children gave oral consent. Participants' consent was obtained according to the Declaration of Helsinki, and the study was approved by the Ethical Committee at Leiden University. Inclusion criteria were having a typical development and Dutch as native language. In addition, only children scoring above the 40th percentile on the Dutch national standardized word reading test, the *Three Minute Test* (CITO Drie Minuten Toets, (DMT), Krom, Jongen, Verhelst, Kamphuis & Kleintjes, 2010), were included to ensure they had average to good technical reading skills. To ensure that children with a range of comprehension proficiency were included from both grades, children of all proficiency levels on the Dutch national standardized CITO reading comprehension test (Feenstra, Kleintjes, Kamphuis & Krom, 2010; Weekers, Groenen, Kleintjes & Feenstra, 2011) were included.

2.2.2 Materials

2.2.2.1 Think-aloud Protocol

To assess text comprehension processes during reading we used a think-aloud protocol (e.g., Clinton & Van den Broek, 2012; McMaster et al., 2012; Van den Broek, Lorch, Linderholm, & Gustafson, 2001). Participants read two narrative and two expository texts. We used P-CLIB version 3.0 (Evers, 2008) to calculate basic text characteristics, and T-scan (Pander Maat, Kraf, & Dekker, 2017) to calculate text cohesion, see Table 2.1. Different measures on text length from P-CLIB and the Development Level (D-level) from T-Scan together indicate that syntactical complexity is fairly similar across text genre. (The D-level is a combined measure on syntax complexity which assigns sentences to a level on an 8-level scale, based on how difficult the sentence is from a developmental perspective; Pander Maat, Kraf, & Dekker, 2017). The percentage of frequent words and type-token ratio together indicate a higher information density of expository texts compared to narrative text, whereas the higher number of connectives per clause indicate higher text cohesion of expository text. The two narrative texts followed a typical story structure: in the beginning the protagonist had a task to complete (making or buying a gift for a family member) and made a few attempts before fulfilling the goal (e.g., Mandler & Johnson, 1977). The two expository texts followed a descriptive text structure: a nature phenomenon (earthquakes and snakes) was described with main and subordinate ideas (e.g., Meyer, 1987). English translations of the four texts are included in the supplementary file Translated Test Texts.

Table 2.1

Text Characteristics of the Narrative (N) and Expository (E) Texts based on P-CLIB[(a)] and T-Scan[(b)].

	Mieke picks a present (N)	Holiday shopping (N)	When the earth shakes (E)	Dutch snakes (E)
No. of sentences/ text[a]	17	16	15	16
Average no. of words/ sentence[a]	12	12.4	11.7	11.4
Average no. of letters/ word[a]	4.6	4.5	5.3	4.6
Percentage of frequent words[a]	75.61	77.39	72.57	74.18
Grade level[a]	7	7	8	7
D-level (proportion sentences > level 4)[b]	0.18	0.25	0.27	0.19
Type-Token ratio[b]	0.48	0.56	0.67	0.62
Connectives per clause[b]	0.13	0.11	0.28	0.35

Two undergraduate students (both female) and the first author collected the data, hereafter named test leaders. Before collecting the data, the test leaders set up a test protocol and practiced how to instruct the participants for each task. In the think-aloud task, each sentence was printed on a separate page. Participants were instructed to read each sentence aloud and report what they were thinking before moving on to the next sentence. Instructions were explicit in mentioning that there were no correct or incorrect responses, but that the test leader was just interested in hearing what came to their minds when reading the text. Before participants read the four test texts, they received a narrative practice text. For the first half of this text, the test leader modelled the think-aloud procedure following a script to ensure all participants received the same examples (including paraphrases, comments about the text, and different kinds of inferences), i.e. the test leader read a sentence and then modelled thinking aloud about the sentence. Then the participant practiced with the second half. During reading of the four test texts there was no feedback, only neutral encouragement to think aloud such as "What are you thinking now?" when the participant did not respond to the sentence he or she had read (e.g., Pressley & Afflerbach, 1995).

The recorded think-aloud session for each participant was transcribed and categorized by the undergraduate students while supervised by the authors. For each sentence, the think-aloud responses were parsed into subject-verb clauses (idea units) and assigned to categories based on previous research (e.g., Linderholm &

van den Broek, 2002; McMaster et al., 2012; Trabasso & Magliano, 1996). The inter-rater reliability of the two undergraduate students was good, $K = 0.75$, $p < .01$. The following six categories are relevant to the current study (see Table 2.2): text repetitions, paraphrases, text-connecting inferences, valid elaborative inferences, invalid elaborative inferences, and predictive inferences.

Table 2.2

Labels and Description of Think-aloud Response Categories

Response category	Description	Example She decided to make a necklace for her mom
Text Repetition	Literal repetition of the text	She decided to make a necklace for her mom
Paraphrase	Repeating the text in own words	She had made up her mind, she would put together a pretty necklace to give her mom
Text-connecting Inference	Connecting and reinstating events from prior text	A similar one that her friend made for her mom
Valid Elaborative Inference	Correct use of prior knowledge to explain the text	Because parents like when their children make them things
Invalid Elaborative Inference	Incorrect use of prior knowledge to explain the text	So she had to go to the store to get one
Predictive Inference	Predicting upcoming events in the text	I think her mom will be very happy when she opens the present

Only the first categorized response unit, i.e. the first idea unit the participant responded with, for each sentence was used in the analyses. The initial response is thought to indicate spontaneous thoughts in response to the text itself, whereas subsequent responses may reflect other, interfering thoughts (e.g., Ericsson & Simon, 1980; Hertzum, Hansen, & Andersen, 2009). Across the two narrative texts, the number of initial responses was averaged for each response category (e.g. for text repetitions, for paraphrases and so on) for each participant. Thereafter, a percentage score was calculated for each response category in relation to the total number of responses for this text genre. Similarly, initial response averages and percentages were calculated for each category and child across the two expository

texts. Responses such as meta-cognitive or emotional comments, questions, invalid responses to other categories, and inaudible segments were coded but each made up a small percentage of all responses and were removed from further analyses.

2.2.2.2 Word Decoding

As a standardized measure of word decoding skills, we used the normed ability scores of the CITO DMT (The Three Minute Test developed by the Dutch Central Institute of Test Development; Krom et al., 2010) provided by the school. The reliability of this test is good, $\alpha > .921$ (Krom et al., 2010). The DMT is a word decoding test on which participants have to read aloud words without context as fast and as accurately as possible. The scores of the participants ranged from 80 to 124, reflecting average to good word decoding ability.

2.2.2.3 Reading Comprehension

As a standardized measure of reading comprehension skills, we used the normed ability scores of the Dutch national CITO reading comprehension test (Feenstra et al., 2010; Weekers et al., 2011) provided by the school. The reliability of this test is good, MAcc > .89 (Feenstra et al., 2010; Weekers et al., 2011). This is a paper-and-pencil test where the scores are based on how well the participant answers multiple choice and open-ended questions after having read narrative and expository texts. The questions tap into understanding factual information in the texts, inference ability - both within the text and with prior knowledge, and knowledge about text structure. The multiple-choice questions have four answering possibilities. The scores of the participants ranged from 11 to 98, reflecting the whole range of reading comprehension ability.

2.2.2.4 Vocabulary

To assess receptive vocabulary, we used a Dutch version of the Peabody Picture Vocabulary Test adapted for group administration (PPVT: Schlichting, 2005). The reliability of the Dutch PPVT is good, $\lambda\text{-}2 > .89$ (Schlichting, 2005). Participants received a booklet with one word on each page. Under each word four pictures were displayed. To show conceptual knowledge of the word, participants were asked to circle the picture that matched the word. Each individual's score was the number of correctly circled pictures within 15 minutes. There was one practice item and 60 test items, the highest possible score was 60. The scores of the participants ranged from 23 to 56 points.

2.2.2.5 Reasoning Ability

To assess non-verbal reasoning ability, we used a group-administered version of Raven Standard Progressive Matrices (Raven SPM: Raven, Raven, & Court, 1998). The reliability for Raven SPM is good, > .98 (Raven, 2008). Participants received a booklet with matrices with a missing part and were instructed to "solve as many puzzles" as possible within 30 minutes. The participants' task was to find the missing part out of six or eight options. For each item they wrote the number of the correct answer on an answering sheet. The items continuously increase in difficulty throughout the test. An individual's score was the number of correct answers. There was one practice item and 59 test items, the practice item was included in the score and the highest possible score was 60. The scores of the participants ranged from 13 to 53 points.

2.2.2.6 Working Memory

To assess verbal working memory (WM) we used a Dutch version of the Sentence Span task (Swanson, Cochran, & Ewers, 1989). In this task both manipulation of information (answering a question) and maintenance of information (remembering words) are measured. At the first level, the test leader read two unrelated sentences and thereafter asked an open-ended question about the content of one of the sentences. Participants were instructed to first answer the question and then say aloud the last word of each of the two sentences. For each subsequent level, WM load is increased by adding one sentence until reaching the final level with five sentences. There were three practice trials on the first level and two test trials for each level. If a participant made an error on both trials within a level, the test was discontinued. An individual's score was the number of correctly remembered words for trials where the question was answered correctly. This scoring method has gained a good internal consistency of .79 (Conway et al., 2005). The highest possible score is 28. The scores of the participants ranged from 1 to 22 points.

2.2.3 Procedure

Each child participated in two test sessions, an individual session (lasting approximately one hour) and a group session (lasting approximately 45 minutes). The tasks were administered in the same order to each participant during the individual session, first the think-aloud task and second the WM task. The think-aloud task started with the practice text. Thereafter the narrative and expository test texts were presented in an interleaved fashion; Mieke picks a present, When the earth shakes, Holiday shopping, and, finally, Dutch Snakes. There was a break before the WM task was administered. Group sessions were held with all participants within each school during one occasion. First the Raven SPM was administered

followed by the Peabody Picture Vocabulary Test. After the individual session, the participants were thanked and received a small reward.

2.2.4 Analyses

To identify subgroups of children characterized by a profile of think-aloud responses while reading narrative texts, we performed a Latent Profile Analysis (LPA; Hagenaars & McCutcheon, 2002; Oberski, 2016), carried out in version 5.0 of the statistical program Latent Gold (Vermunt & Magidson, 2013). LPA is a model-based cluster analysis technique, aiming to identify qualitative individual differences (i.e., subgroups or clusters) based on individuals' scores on a set of continuous variables. The following five think-aloud categories were included as continuous variables: the percentage of text repetitions, paraphrases, text-connecting inferences, valid elaborative inferences, and predictive inferences. As there were only few observed values of the variable 'percentage of invalid elaborative inferences', this variable was included as an ordinal (also compatible with the LPA, Vermunt, Tran, & Magidson, 2008) rather than as a continuous variable. To select the optimal number of latent clusters we used statistical information criteria, Bayesian Information Criterion (BIC) and Consistent Akaike Information Criterion (CAIC), which represent a trade-off between model fit (log-likelihood) and model complexity (the number of estimated parameters), combined with the interpretability of the resulting clusters (e.g., Hickendorff et al., 2017). Entropy and classification measures were used to evaluate the absolute fit of the model to the data.

To compare the reading profiles across the text genres (narrative vs. expository text), we performed a second LPA on the think-aloud responses to the expository texts. For this analysis, the percentages of invalid elaborative inferences and of predictive inferences were included as ordinal variables, as these held few observed values. The remaining four think-aloud categories were included as continuous variables. To select the optimal number of latent clusters we used the BIC and CAIC, combined with interpretability of the resulting clusters. To test whether individual children have a similar response profile across narrative and expository texts, we performed a cross tabulation with a chi-square test for independence between the cluster membership based on narratives and the cluster membership based on exposition.

To examine whether the identified profiles are related to differences in language abilities and cognitive resources, we performed two one-way between-groups Multivariate Analysis of Variance (MANOVA). Children's performance scores on word decoding, reading comprehension, vocabulary, non-verbal reasoning ability, and working memory were entered as dependent variables and their LPA cluster membership (based on reading narrative or expository texts) as fixed factor.

Descriptive statistics and correlations of the variables used in the above-mentioned analyses are reported in the supplementary Tables 2.S1 and 2.S2, respectively.

2.3 Results

2.3.1 Profiles Characterized by Think-Aloud Responses to Narrative Text

To identify reading profiles based on narrative texts we conducted LPAs with one to eight clusters. The model with three clusters had the lowest BIC-value and CAIC-value. This model had an R-square entropy of .93 (values range between 0 and 1, and higher values indicate more certainty of classification; Collins & Lanza, 2010) and a classification error of .022, indicating that children's cluster membership could be predicted from their responses very well. The following five response types made significant contributions to the classification: text repetitions, paraphrases, text-connecting inferences, valid elaborative inferences, and predictive inferences (Wald's > 43, all *p*s <.001). Invalid elaborative inferences did not make a significant contribution to the classification (Wald-statistic = .14, $p = .93$). The profiles of the three clusters are visualized in Figure 2.1.

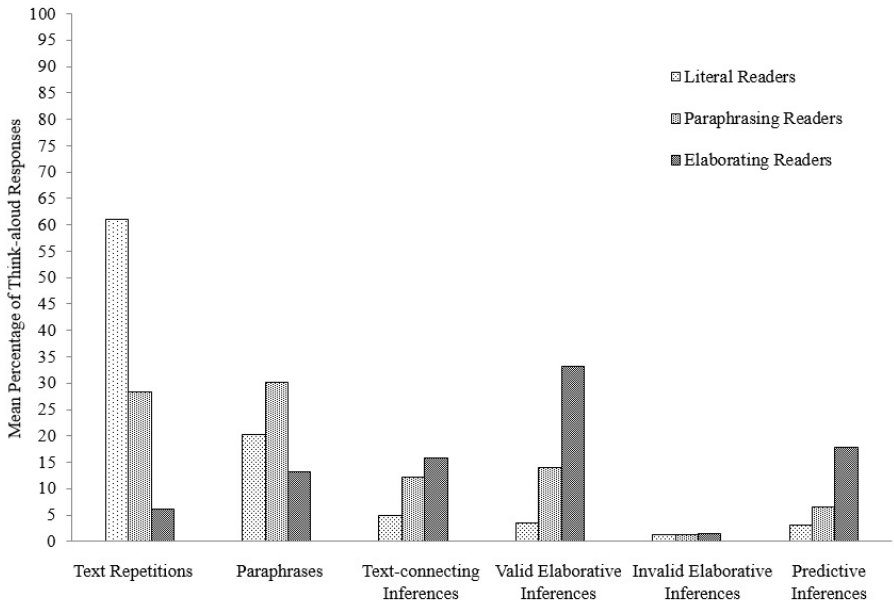

Figure 2.1. Clusters characterized by think-aloud responses while reading narrative texts. The clusters' mean percentage of responses is based on the LPA-estimates and shown on the y-axis. Type of think-aloud category is shown on the x-axis.

The three profiles seem to reflect different approaches to text processing based on the type of responses children in the three clusters were most likely to produce. The largest cluster (43.6% of the children) consisted of *Paraphrasing Readers*. Children belonging to this cluster primarily paraphrased the text, more so than the other two clusters (all *p*s < .008, see Supplementary Table 2.S3). In addition, they often repeated the text literally, and produced a fair number of text-connecting inferences and valid elaborative inferences. The second largest cluster (34.5% of the children) consisted of *Elaborating Readers*. Children belonging to this cluster primarily used valid elaborative inferences to explain the text. In addition, they produced a fair number of text-connecting inferences and predictive inferences. They differentiate from the other two clusters primarily in the number of valid elaborative inferences and predictive inferences (all *p*s < .001, see Supplementary Table 2.S3). The third cluster (21.9% of the children) consisted of *Literal Readers*. Children belonging to this cluster showed a large number of text repetition responses, more so than the other two clusters (all *p*s < .001, see Supplementary Table 2.S3). In addition, they produced a fair number of paraphrases.

2.3.1 Profiles Characterized by Think-Aloud Responses to Expository Text

To examine whether children approach narrative and expository texts similarly we first identified profiles of readers on their responses to expository texts by performing an LPA. From the LPAs with one to eight clusters, the three- and four-cluster models had virtually identical BIC values. Of these two, the three-cluster model had the lowest CAIC value, therefore we report the three-cluster model. This model had an R-square entropy of .87 and a classification error of .057. All response types made significant contributions in the cluster classification (Wald's > 6.8, all *p*s <.05). The profiles of the three clusters are visualized in Figure 2.2.

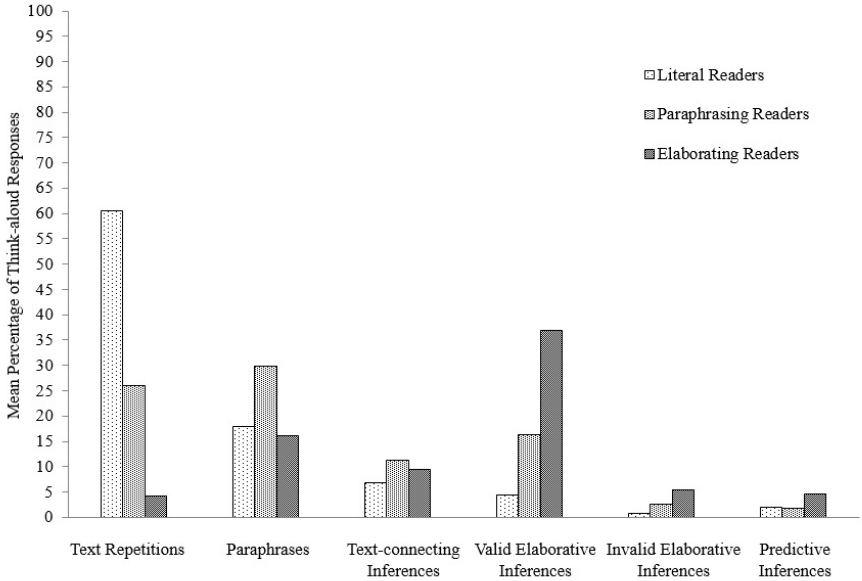

Figure 2.2. Clusters characterized by think-aloud responses while reading expository texts. The clusters' mean percentage of responses is based on the LPA-estimates and shown on the y-axis. Type of think-aloud category is shown on the x-axis.

The three profiles resemble the profiles identified in the LPA on narrative texts and reflect similar approaches to text based on their think-aloud responses. Paraphrasing Readers (47.5% of the children) primarily paraphrased the text or repeated it literally; in addition, they produced a fair number of text-connecting inferences and valid elaborative inferences. Literal Readers (26.9% of the children) produced a large number of text repetitions; in addition, they produced a fair number of paraphrases. Elaborating Readers (25.6 % of the children) primarily used valid elaborative inferences to explain the text; in addition, they produced a fair number of paraphrases and text-connecting inferences. Inspection of Figures 2.1 and 2.2 shows that the response patterns are very similar across narrative and expository texts for all profiles, but the number of inferences differ between the two text genres. Expository texts elicited fewer predictive inferences in both Paraphrasing Readers and Elaborating Readers than narrative texts. Furthermore, both Paraphrasing Readers and Elaborating Readers made more invalid elaborative inferences while reading expository than while reading narrative texts, with the largest difference within the Elaborating Readers. Elaborating Readers also produced fewer text-connecting inferences when reading the expository texts compared to narrative texts.

2.3.3 Comparison between Profiles Characterized by Narrative and Expository Texts

To examine whether individual children showed a similar reading profile across narratives and exposition we performed a cross tabulation of the classification based on the narrative and the expository text. First, all participants were assigned to a profile based on reading the narrative texts, thereafter, based on reading the expository texts. To do so a modal assignment procedure (Vermunt & Magidson, 2013) was used by which children were assigned to the profiles for which they had the highest posterior probability on narrative and expository text, respectively. Across all children, the average classification error was low, .022. The Kendall's tau-b showed a significant relation between profiles on narrative and expository texts, χ^2 (4, N = 107) = 80.43, p < .001, Kendall's tau-b = .71. The majority of Literal Readers (79.2 %), Paraphrasing Readers (71.7 %), and Elaborating Readers (64.9 %) had the same approach to both narrative and expository texts (see Figure 2.3). Across all profiles, 24.6% of the children had a different profile in which they used *fewer* inferences and more paraphrases and repetitions while reading exposition, compared to narratives. Only 8.5% of the children had a different profile in which they used *more* inferences while reading exposition, compared to narratives. The results suggest an ordinal relation between different profiles, which is supported by the high Kendall's tau-b. If an individual approached expository texts differently than narrative texts, there was indeed an ordinal manner in which such change in text approach occurred. For example, some individuals assigned to the profile Elaborating Readers while reading narratives became assigned to the profile Paraphrasing Readers while reading exposition, but none changed into Literal Readers. To summarize, children approach narrative and expository texts similarly and, if they use a different approach, they are likely to stay closer to the literal text in expository texts compared to narrative texts.

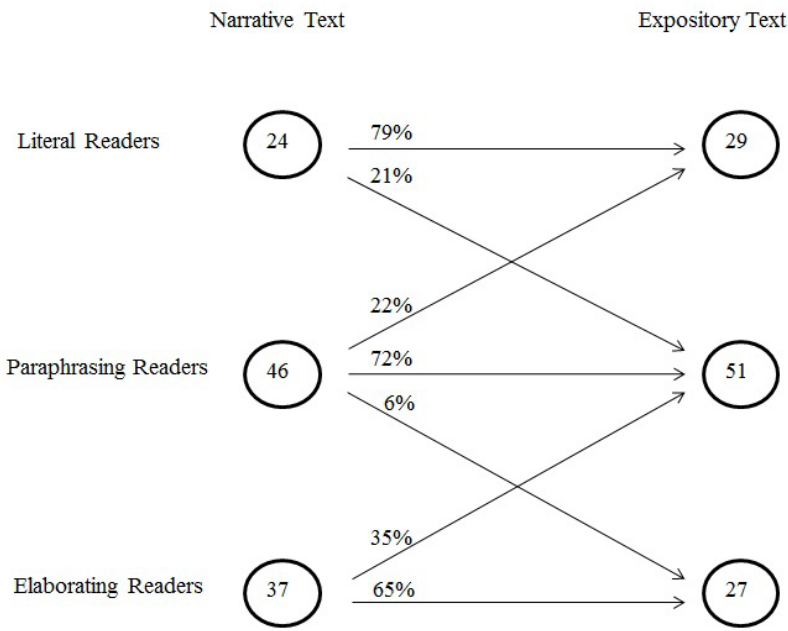

Figure 2.3. Children's profile membership for narrative and expository texts. In the circles, the total number of children in each cluster type is reported for each type of text. The arrows show the change in think-aloud response profiles; the percentages of children keeping the same profile or changing to a different profile are indicated above each arrow.

2.3.4 Profiles Related to Language Abilities and Cognitive Resources

Two MANOVAs were performed to examine whether the identified profiles are related to differences in word decoding, reading comprehension, vocabulary, reasoning ability, and working memory. Due to missing data in the dependent variables for three children the total number of children is 104 in these analyses. For the first MANOVA, all participants were assigned to their profile based on reading the narrative texts (a modal assignment procedure; Vermunt & Magidson, 2013). The data did not show any violations of assumption of homogeneity of variance-covariance matrices ($p > .001$). The equality of variance was not violated for any of the dependent variables (all ps $> .05$), except for word decoding, $p = .040$. As word decoding did not show equality of variance, we used the Pillai's Trace significance level for a more robust test. There was a significant difference between the three profiles on the combined dependent variables, F (10, 196) = 2.88, p = .002, $\eta_p^2 = .13$. Means and standard errors for each profile are reported in Table

2.3. Considering the dependent variables separately in five univariate ANOVAs with Bonferroni-corrected alpha levels of .01, cluster membership had a significant effect on word decoding, F (2, 101) = 6.79, p = .002, η_p^2 = .12, reading comprehension, F (2, 101) = 8.45, p < .001, η_p^2 = .14, and non-verbal reasoning ability, F (2, 101) = 7.36, p = .001, η_p^2 = .13. Neither vocabulary, F (2, 101) = 1.79, p = .171, η_p^2 = .03, nor working memory, F (2, 101) = 2.89, p = .060, η_p^2 = .05, differed significantly between the three profiles. Post-hoc analyses comparing mean scores between the profiles on the three dependent variables with a significant effect were conducted with Tukey HSD adjustment. The Elaborating Readers performed significantly better than Literal Readers on word decoding (p = .003), reading comprehension (p < .001), and reasoning abilities (p = .001). The Elaborating Readers also performed significantly better than the Paraphrasing Readers on word decoding (p = .012), reading comprehension (p = .039), and reasoning abilities (p = .019). Although Paraphrasing Readers numerically performed better than Literal Readers on all three measures, none of these differences were significant (all ps > .103).

Table 2.3

Means and Standard Errors from the MANOVA for Each Dependent Variable and Each Reading Profile Based on the Narrative Texts.

	Literal Readers (N = 23)	Paraphrasing Readers (N = 45)	Elaborating Readers (N = 36)
Word Decoding	93.35 (1.94)**	95.62 (1.40)*	101.69 (1.55)
Reading Comprehension	30.78 (3.59)***	39.89 (2.57)*	49.44 (2.87)
Reasoning Ability	34.70 (1.58)***	37.42 (1.13)*	42.08 (1.26)
Vocabulary	40.17 (1.36)	41.10 (0.97)	43.22 (1.09)
Working Memory	6.74 (0.96)	8.24 (0.68)	9.67 (0.77)

Note. Asterisks indicate where the Literal and Paraphrasing Readers differed from the Elaborating Readers at * p < .05, ** p < .01, and *** p < .001. There were no significant differences between the Literal and the Paraphrasing Readers.

For the second MANOVA, all participants were assigned to a profile based on reading the expository texts (Vermunt & Magidson, 2013). The data did not show any violations of assumption of homogeneity of variance-covariance matrices (p > .001) or of assumption of equality of variance for any of the dependent variables (all ps > .05). We kept using the Pillai's Trace significance level for a robust test. There was a significant difference between the three profiles on the combined dependent variables, F (10, 196) = 1.90, p = .047, η_p^2 = .09. Means and standard

errors are reported in Table 2.4. Considering the dependent variables separately in five univariate ANOVAs with Bonferroni-corrected alpha levels of .01, cluster membership had a significant effect on non-verbal reasoning ability, F (2, 101) = 6.07, p = .003, η_p^2 = .11, and a marginally significant effect on reading comprehension, F (2, 101) = 4.76, p = .011, η_p^2 = .09. There were no effects on word decoding, F (2, 101) = 3.80, p = .026, η_p^2 = .07, vocabulary, F (2, 101) = 2.25, p = .111, η_p^2 = .04, or working memory, F (2, 101) = 1.29, p = .280, η_p^2 = .02. Post-hoc analyses showed that the Elaborating Readers performed significantly better than Literal Readers (p = .002), and Paraphrasing Readers performed marginally better than Literal Readers (p = .058) on reasoning abilities. There was no difference between Elaborating Readers and Paraphrasing Readers (p = .239) on reasoning abilities. In addition, the Elaborating Readers performed significantly better than Literal Readers (p = .008) on reading comprehension, no other differences were significant (all ps > .121).

Table 2.4

Means and Standard Errors from the MANOVA for Each Dependent Variable and Each Reading Profile Based on the Expository Texts.

	Literal Readers (N = 27)	Paraphrasing Readers (N = 50)	Elaborating Readers (N = 27)
Word Decoding	93.35 (1.94)	95.62 (1.40)	101.69 (1.55)
Reading Comprehension	34.22 (3.43)**	40.66 (2.52)	49.11 (3.43)
Reasoning Ability	34.52 (1.48)**	38.7642 (1.08)	41.74 (1.48)
Vocabulary	39.48 (1.25)	41.98 (0.92)	43.11 (1.25)
Working Memory	7.41 (0.90)	8.38 (0.66)	9.44 (0.90)

Note. ** indicate where the Literal Readers differed from the Elaborating Readers at p < .01. No other differences were significant.

2.4. Discussion

We examined possible reading profiles using children's think-aloud responses while reading narrative and expository texts and investigated whether young readers approach narrative texts and expository texts differently. Furthermore, we examined whether children with distinct reading profiles differ in language abilities and cognitive resources. In pursuing these aims, we made a distinction within close-to-the-literal-text processes, namely between text repetitions and paraphrases.

Three distinct reading profiles that differ in the number and types of inferences the readers make were revealed by Latent Profile Analyses (LPA). First, *Literal*

Readers stay close to the literal text; they predominantly repeat the text and engage in few inferential processes. Second, *Paraphrasing Readers* extract the meaning from the text; they predominantly paraphrase the text, but also make some text-connecting and elaborative inferences. Third, *Elaborating Readers* use background knowledge and go beyond as well as think ahead of the focal sentence; they predominantly use elaborative inferences, but also a fair number of text-connecting and predictive inferences.

2.4.1 Text Genres

To determine whether children approach various text genres differently, we examined think-aloud responses for both narrative and expository texts. For both types of texts, the profiles of Literal Readers, Paraphrasing Readers, and the Elaborating Readers emerged. There were considerable similarities in how the three subgroups approached narrative and expository texts. However, there also were a few noticeable differences. Children that differed in think-aloud response profiles across text genres tended to go beyond the text and use inferences when reading narratives but stay closer to the literal text and make fewer inferences when reading expository text. This resulted in more children belonging to the Literal and the Paraphrasing Readers in expository text reading than in narrative text reading. Furthermore, expository texts elicited fewer inferences within each profile. For example, there was a decrease in the number of predictive inferences made while reading expository texts compared to narratives. The decrease was particularly noticeable in Elaborating Readers (who produced most predictive inferences while reading narratives), but was also present in Paraphrasing and Literal Readers. Expository and narrative texts make different demands on the reader (e.g. Eason et al., 2012; Lorch, 2015). For example, narratives often include words that are part of everyday vocabulary whereas expository text may include more non-frequent technical terms (Medina & Pilonieta, 2006). In the current texts, a slight lower word frequency and a higher number of unique words were observed for expository texts compared to narratives. In addition, narratives activate familiar script structures (with initiating goals, actions, and outcomes) similar to everyday experiences (e.g., Britton & Pelligrini, 1990) that likely enhance predictive inferences (Narvaez, van den Broek, & Barrón Ruiz, 1999), whereas expository text structure may not offer the constraints necessary for making predictions (e.g., Cook et al., 2001; Klin et al., 1999). In the current study, the expository texts contained more connectives for higher text cohesion, however, the narratives were possibly experienced as easier to comprehend because of higher familiarity of word and text

structure, as also argued by Graesser, McNamara, and Louwerse (2003). Our results indicate that differences in text demands elicit differences in children's reading behavior.

Expository texts did not only elicit a decrease in overall inference generation, but also an increase of invalid elaborative inferences. The subgroup producing most valid inferences - Elaborating Readers – also generated most invalid elaborative inferences, but there was also an increase in Paraphrasing Readers. Because invalid inferences are likely to impede comprehension (cf. McMaster et al., 2012), this observation constitutes an important issue for future studies. For example, investigating reading processes that occur after invalid inferences may shed light on the impact invalid inferences have on local and global text comprehension. Questions pertaining to whether and how the positive influence of making valid inferences outweighs the negative effect of making invalid inferences and, to the interplay of valid and invalid inferences as they occur while a text unfolds are highly relevant.

Considering reading profiles under different conditions provides insight on the stability of readers' approaches to text and their use of specific sets of reading processes. On the one hand, research on reading profiles suggests that children possess a certain set of skills, a reading trait. Longitudinal research adds evidence that such skills are relatively stable and predict future performance (e.g. Kendeou, van den Broek, White, & Lynch, 2009). On the other hand, research investigating different task- and text demands (for a recent review, Kendeou & van den Broek, 2017) suggests that reading processes may fluctuate depending on the situation, and that readers' standards of coherence moderate the processes in which they engage (e.g. van den Broek & Helder, 2017). The current findings show that children produce a very similar set of think-aloud responses across text genres but also adjust to the different situations. Hence, the results suggest a certain stability in children's text approaches with room for situational differences.

2.4.2 Reader Characteristics

Including readers from a large range of poor to good comprehension ability enabled us to investigate whether the children with different reading profiles differ in underlying competences. Based on narrative texts, Elaborators perform significantly better than the other two groups on word decoding, reading comprehension, and non-verbal reasoning. Based on expository texts, Elaborators perform significantly better than Literal Readers on reading comprehension and non-verbal reasoning. Thus, children with good comprehension and non-verbal reasoning abilities also make more frequent use of background knowledge and make more predictions about the upcoming text. This is in line with previous findings that

reading skills (e.g., Cain & Oakhill, 1999; Carlson et al., 2014) and non-verbal reasoning skills (de Leeuw et al., 2016; Naglieri, 2001) are positively related to inference making. The causal relation between inference generation and reading comprehension may take different forms; inference generation may aid text comprehension, but also good word and text comprehension may provide possibilities to connect the text to background knowledge. Interestingly, current results suggest that non-verbal reasoning skills are also important in this equation. These findings encourage future research to investigate whether the relation between inference generation and comprehension is mediated by reasoning abilities.

The profiles did not differ significantly on vocabulary or working memory. However, and perhaps to be noted as a limitation of the study, it is difficult to dismiss these measures as unimportant when investigating individual differences. As can be seen in the supplementary Table 2.S2, all ancillary measures correlate with each other, with the exception of word decoding which did not correlate with vocabulary and reasoning skills. Furthermore, it is worth mentioning that, numerically, Elaborating Readers (based on both narrative and expository reading) performed better than readers with the two other profiles on all five measures of language abilities and cognitive resources. Similarly, Paraphrasing Readers numerically performed better than Literal Readers on all five measures. This consistency suggests a hierarchy between the three profiles in which better language abilities and cognitive resources are related to the ability to make inferences that go beyond the literal text.

2.4.3 Paraphrasing Text Facilitates Inference Generation

There were considerable differences in processing texts between Literal and Paraphrasing Readers. Not only do Paraphrasing Readers make more paraphrases and fewer repetitions than Literal Readers do, they also produce more inferences. In light of previous findings that paraphrasing a text may help readers start up inferences (e.g., Bohn-Gettler & Kendeou, 2014), one may speculate that the Paraphrasing Readers profited from being able to summarize the text in their own words to enable inference generation. Conversely, being less able to extract the meaning beyond the literal level of a text may hamper inference making and reading comprehension in Literal Readers.

2.4.4 Levels of Reading Comprehension

An interesting aspect of the observed three reading profiles is that they resemble ideas about different levels of text comprehension: surface, textbase, and situation-model understanding (e.g., van Dijk & Kintsch, 1983; Kintsch, 1988, 1994). Whereas Literal Readers mainly show a surface level understanding of the text

with many verbatim repetitions, Paraphrasing Readers mainly show a textbase understanding by focusing on the *meaning* of the text input. Finally, Elaborating Readers create an enriched situation model of the text by drawing on background knowledge. Again, this suggests a hierarchy among the reader profiles, with Elaborating Readers being most likely to attain deep comprehension. Empirical support for such hierarchy comes from two of the current findings. First, Elaborating Readers, showed best comprehension and reasoning abilities, followed by the Paraphrasing Readers and, thereafter, the Literal Readers. Second, profile membership across text genre was rather stable, but children that differed in reading profile across text genres seemed to follow a hierarchical pattern: some individuals who were identified as Elaborating Readers while reading narrative texts were identified as Paraphrasing Readers while reading expository texts, but none changed his or her response pattern so drastically as to become a Literal Reader. Together, these findings suggest a hierarchy between the three profiles with reading comprehension ranging from basic to deep understanding. Whether these profiles are stable over time or reflect stages in children's reading development is an empirical question which can only be answered by longitudinal research.

2.4.5 Educational Implications

The central findings of the current research are that readers systematically differ in how they process text and that these differences are associated with inference generation and comprehension abilities. In an educational context, they suggest the importance of assessing the extent to which individual readers are able to enrich their mental representation of text with background knowledge (e.g., Carlson et al., 2014), and of using such assessment for individualized support of deep reading comprehension. For example, a paraphrasing training may benefit Literal Readers and encourage them to move beyond the literal meaning of the text to eventually enable inference generation (Bohn-Gettler & Kendeou, 2014; McNamara, 2004; McNamara et al., 2009). Similarly, questioning techniques have proven useful to encourage readers to generate more inferences (e.g., McMaster et al, 2012), which may be particularly beneficial for Paraphrasing Readers. Elaborating Readers use inferences correctly for narrative texts and displayed best performance on an ancillary comprehension measure. However, they make more invalid inferences for expository than for narrative texts. For these students it may be useful to focus on awareness and acknowledgement of invalid inferences (e.g., Van den Broek & Kendeou, 2008) and teach strategies that may compensate for lack of background knowledge. Teaching strategies for comprehension monitoring and metacognitive processes (e.g. McCrudden & Kendeou, 2014) could increase awareness of whether one's background knowledge matches the topic of the text.

In addition, reading methods such as the Concept Oriented Reading Instruction (CORI; e.g. Guthrie & Alao, 1997) may enhance motivation, increase strategy use, and increase conceptual knowledge during reading. Hence, we encourage tailoring instruction to individual needs, however with the aim of stimulating higher-order thinking and deep comprehension for all types of readers (Mayer, 1996; Oakes, 2008).

2.4.6 Reading Profiles; Additions and Limitations

Research on reading profiles with person-centred analyses (e.g. LPA) is promising as it addresses the fact that not all measures of reading processes are linearly related but are heterogeneous both between and within individuals. For example, the Literal and Elaborating Readers display almost opposite patterns in their think-aloud responses; Literal Readers make many text repetitions and few inferences, and Elaborating Readers make few text repetitions and many inferences. In addition, a reading profile between these two emerges, the Paraphrasing Readers. These differences may not have been picked up on using variable-centred analyses (e.g. linear models) which assume relations to be the same for all individuals. In this respect, current results support and expand previous literature on reading profiles. By distinguishing between close-to-text processes (text repetitions and paraphrases), it appears there are at least three reading profiles rather than the two reported before (e.g. Carlson et al., 2014; McMaster et al., 2012; Seipel et al., 2017); in addition, the profiles apply to a wider population than just poor or just good readers. However, the limited sample size of the current study did not allow to replicate the LPAs with a split-sample method. Hence, we could not test the generalizability of these three reading profiles, neither in a sample with a range of reading proficiency, nor at different levels of reading proficiency. Further research alternating the pool of participants and the reading task is still needed to generalize reading profiles.

Our results were obtained with a think-aloud method and we make interpretations within the possibilities of this method. However, we believe it is important to discuss some limitations of the think-aloud method. First, thinking aloud while reading may focus a reader's attention on processes he or she is not normally aware of and, thereby, alter the reading experience and text comprehension. However, given that in the current study the reading profiles are fairly stable across text genres and that Elaborating Readers systematically score highest on reading comprehension and reasoning abilities, evidence suggest that the obtained profiles are not just an artefact of the think-aloud method. Second, developing readers may not have enough attentional capacity to extensively reflect on all higher-order cognitive processes occurring while reading. If individual differences in reporting far-

from-text processes (i.e. inferences) are due to an inability to formulate thoughts on ones' reading processes, then distinguishing between two kinds of close-to-text processes (i.e. text repetitions and paraphrases) takes on added importance, as it allowed investigation of qualitative differences between young readers' processing of the literal text.

2.5. Conclusions

We identified three reading profiles -Literal, Paraphrasing, and Elaborating Readers- that differ in the number and types of inferences they make. The reading profiles were remarkably constant across narrative and expository texts. However, for expository texts children tend to fall back on reading processes closer to the literal text. Furthermore, differences between these profiles reveal a positive relation between the ability to elaborate on the text at hand and comprehension ability as well as reasoning ability.

The identification of profiles has implications for reading theory as well as for educational practice. On the theoretical side, it deepens our understanding of individual differences in inference generation in particular, and of the complex interaction of processes involved in reading comprehension in general. On the educational side, it provides a basis for conceptualizing individualized and adaptive approaches to reading instructions with the aim to promote in-depth comprehension.

Supplementary
Think-Aloud Texts Translated from Dutch to English

Mieke Picks a Present

One day, Mieke remembered that her mother's birthday was coming soon. She wanted to give her mother a nice present. Mieke went to the store and found a nice pair of earrings. She bought them for her mother, went home, and wrapped them in bright paper. When the present was nicely wrapped, she hid it in her closet.

The next day, Mieke saw her friend Sandra make her own jewellery. When she was little, she had made a bracelet and she remembered that her mother liked hand-made jewellery. She decided to make a nice necklace for her mother. She chose the pearls she wanted to use and followed Sandra's instructions. Finally, Mieke had made a beautiful and long necklace. Mieke made a knot and tied a hook to the string. She stored it in a beautiful red box in her closet until the big day. Mieke was very happy with her decision to make a necklace and returned the earrings to the store.

At last, her mother's birthday arrived. Mieke took the red box from her closet. Her mother opened the box and was very happy when she saw the necklace. She thanked Mieke for the beautifully hand-made gift.

When the Earth Shakes

Sometimes earthquakes happen in South of Europe. An earthquake occurs when the ground suddenly begins shaking. About 100 kilometres inside the earth are layers of rock that move and crack. It is the moving and cracking of these rock layers that causes earthquakes. Each year, many small earthquakes affect the South of Europe. Even though most people normally can't even feel small earthquakes, large earthquakes can cause major damage. They can cause buildings, bridges, and homes to fall down.

Because earthquakes are so common, buildings need to be built in special ways. Although this helps protect buildings from earthquakes, damage is still possible. Recently, large earthquakes have shaken several cities in South of Europe. For example, one woman recollected a memory for us about a large earthquake that happened in Rome in 2009. First, she heard a loud crack. Then, the dining area of her second-floor apartment moved downward. "I felt like I was falling," she recalled. "Until I was outside of the building, I was unsure of what had happened. The earthquake had completely crushed the first floor."

Holiday Shopping

Lisa thought it was time to go shopping for Holiday presents. She disliked shopping when the stores were busy. When she got to the mall it was crowded, the parking lot was full.

She was determined to buy her three-year-old daughter an adorable doll. She found one, paid for it, and left to continue shopping for other things.

Lisa had no idea what to get her husband. She went to a men's clothing shop and looked at expensive ties, suits, and cologne. Then she went to a tech store and examined laptops, MP3-players, and televisions. Nothing at these places seemed appropriate. She walked by a pet store and saw an adorable puppy. She knew instantly that this was the perfect gift. She left the mall and decided to postpone more shopping until the next day.

Lisa got in her car and drove home. She realized she couldn't bring the puppy home. She went to a friend's place and asked if she could leave the puppy there. Her friend said yes and invited her inside for some thee.

Dutch Snakes

In the Netherlands the temperature is most often agreeable for humans. This is not the case for snakes. They don't think it is hot enough. Because of this there are only three types of snakes in the Netherlands, but they are rarely sighted.

A snake that we do see more often is the collar snake, this snake is active during the day. Collar snakes can be recognized by the yellow collar behind their head. Their food consists of frogs, lizards, and fish, and they can swim very well.

The viper is the only poisonous snake in the Netherlands. You may encounter this snake at the Veluwe or in Drenthe. The poison is deadly for their normal pray: field mice and lizards. Vipers live all over Europe, even in North of Russia. Because of the cold there it takes two years for the eggs to hatch.

In the Netherlands there is also a small constrictor, this snake is not poisonous. They can become up to 75 centimetres long and eat preliminary lizards and mice. You can recognize constrictors by the black and brown stripe on both sides of their heads. They mostly live in dry areas such as moors, grass fields and at the outskirts of forests.

Supplementary Descriptive Analysis

Table 2.S1

Means, Range, Standard Deviation, Skewness and Kurtosis for All Measures are Reported here for All Children. The (N) and the (E) Signals Descriptive Statistics for the Narrative and the Expository Texts, Respectively.

	Mean	Range	Standard Deviation	Skewness	Kurtosis
Word Decoding	97.22	44.00	9.83	.86	.60
Reading Comprehension	41.18	87.00	18.44	.83	.18
Reasoning Ability	38.43	40.00	8.04	-.71	1.18
Vocabulary	41.63	33.00	6.58	-.16	-.26
Working Memory	8.40	21.00	4.68	.48	-.16
Text Repetition (N)	27.80	88.57	22.36	.64	-.42
Paraphrase (N)	22.08	51.43	12.04	.29	-.50
Text-connecting Inference (N)	11.80	42.86	8.37	1.10	1.43
Valid Elaborative Inference (N)	18.29	65.71	14.26	.85	.23
Invalid Elaborative Inference (N)	1.33	14.28	2.52	2.53	7.98
Predictive Inference (N)	9.67	57.14	11.47	1.88	4.04
Text Repetition (E)	29.71	84.85	23.46	.58	-.63
Paraphrase (E)	23.11	60.61	11.86	.25	.20
Text-connecting Inference (E)	9.63	33.33	6.30	.88	.97
Valid Elaborative Inference (E)	18.41	57.57	14.20	.71	-.23
Invalid Elaborative Inference (E)	2.80	15.15	3.50	1.61	2.63
Predictive Inference (E)	2.55	21.21	4.06	2.12	5.15

Supplementary Correlational Analysis

Table 2.S2

Correlations among Word Decoding, Reading Comprehension, Reasoning Ability, Vocabulary, and Working Memory are Reported here for All Children (N=104).

	1	2	3	4	5
1. Word Decoding	1	-	-	-	-
2. Reading Comprehension	.358**	1	-	-	-
3. Reasoning Ability	.108	.530**	1	-	-
4. Vocabulary	.179	.618**	.459**	1	-
5. Working Memory	.194*	.436**	.340**	.272**	1

Note. Asterisks indicate significant levels at * $p < .05$, ** $p < .01$.

Supplementary Post Hoc ANOVAs

Table 2.S3

Posthoc ANOVA comparisons between the subgroups. In addition to the LPA's visualized in figure 2.1 and 2.2, we here show the p-value of the Posthoc ANOVAs to display whether the subgroups differ significantly between each other on each think-aloud response category. LR. = Literal Readers, PR = Paraphrasing Readers, and ER = Elaborating Readers. The (N) and the (E) Signals Statistics for the subgroups based on Narrative and Expository Texts, Respectively.

	LR vs. PR (N)	LR vs. ER (N)	PR vs. ER (N)	LR vs. PR (E)	LR vs. ER (E)	PR vs. ER (E)
Text Repetition	.000	.000	.000	.000	.000	.000
Paraphrase	.000	.008	.000	.000	.781	.000
Text-connecting Inference	.001	.000	.052	.010	.301	.432
Valid Elaborative Inference	.000	.000	.000	.000	.000	.000
Invalid Elaborative Inference	-	-	-	.043	.000	.000
Predictive Inference	.368	.000	.000	.940	.063	.013