Establishing coherence in schoolbook texts

How connectives and layout affect students' text comprehension*

Gerdineke van Silfhout, Jacqueline Evers-Vermeul and Ted J.M. Sanders
Utrecht University

This article focuses on the influence of connectives (*because*, *so*) and layout (continuous placement of sentences versus each sentence beginning on a new line) on the quality of students' mental representations. By using multiple comprehension tasks, we found that cohesive text features have different effects on each facet of deeper text comprehension. On local comprehension tasks (i.e. bridging inference questions), all students performed better after reading history texts containing connectives than after reading texts without these markers. On global comprehension tasks (i.e. sorting tasks), pre-vocational students performed better when coherence relations were marked, regardless of layout, while pre-un iversity students did not need connectives as long as texts were presented in a natural, continuous way. These findings indicate that connectives are an important factor in creating comprehensible texts, in particular for pre-vocational students. Finally, we conclude there is a mismatch between these findings and the current practice in designing optimal educational texts, at least in the Netherlands.

Keywords: coherence, connectives, secondary education, text comprehension, (dis)continuous layout

1. Introduction

Reading and comprehending texts has become very important in our increasingly information-driven society. Printed and digital texts are among the most

* This research was supported by *Stichting Lezen* (Dutch Reading Association), Amsterdam and the Utrecht institute of Linguistics OTS.

important tools to convey new information to others. However, roughly one quarter of the adults in the United States only meet basic reading proficiency levels, and, as a result, do not have the skills and competencies to function effectively in society (National Center for Education Statistics, 2012).

Reading and comprehending texts is also crucial for students, because most learning at school is done by reading texts (Daneman, 1991). Yet, many students struggle to learn to become proficient text comprehenders. For example, 14.3% of the Dutch 15-year-olds are considered low-literate (OECD, 2010), and 25% of the eighth graders in the US are not able to read texts at a basic level (NCES, 2012). Their poor comprehension indicates these students fail to construct a coherent mental representation of the information in the text, which is necessary for successful text comprehension and learning (Gernsbacher, 1990; Graesser, Millis, & Zwaan, 1997; Kintsch, 1998). Therefore, the RAND on Reading for Understanding report emphasizes the pressing need to improve text comprehension in schools (Snow, 2002).

The RAND call for comprehension improvement leads to an important, socially relevant question: how do we improve text comprehension of students in specific subgroups in secondary education, particularly pre-vocational and pre-university education? Previous studies have shown that text comprehension is a product of complex interactions between the characteristics of the reader, the text, and the task (McNamara, Kintsch, Songer, & Kintsch, 1996; Snow, 2002). Taking into account the interactions between characteristics of readers, texts and tasks, we investigate the effects of two text features (coherence marking and layout) on students' comprehension of school texts, comparing pre-vocational and pre-university students' performance on various comprehension tasks that assessed different aspects of text comprehension. Before we discuss this in more detail, we first need to know what text comprehension is, and how it is established. Second, we discuss deficiencies observed in the reading processes of young readers, and third, we hypothesize that coherence marking and layout can help these readers in constructing coherent mental representations of texts.

1.1 Text comprehension

The construction-integration model of text comprehension distinguishes three levels of understanding within the reader's mental representation of the text (Kintsch, 1992; 1998; McNamara et al., 1996). The surface code, the first and the most superficial level, preserves the exact words and syntax of the clauses. At the intermediate level is the textbase, which contains clauses that preserve the meaning rather than the exact words and syntax. There is consensus that the textbase level can be seen as a network in which the propositional representations of the

individual sentences are interrelated by means of referential links, but opinions differ with respect to the question whether coherence relations are also part of the textbase representation (Fletcher, 1994; Mulder & Sanders, 2012). The situation model, the third and deepest level, is the conceptual content of what the text is about and consists of both the information expressed by the text and the information that is inferred from the text on the basis of world knowledge about the same or related topics (Kendeou & Van den Broek, 2005; Kintsch, 1994; 1998; Rapp, Van den Broek, McMaster, Kendeou, & Espin, 2007). Learning from text requires a good situation model, because learning encompasses more than reproducing the content of individual clauses, especially in the case of school texts. Only if readers construct a coherent situation model will they be able to retrieve the text information for a long time after reading the text.

In order to construct a coherent situation model, readers must establish coherence relations, such as cause-consequence, list, contrast, and temporal relations. These coherence relations are regarded as the cornerstone of comprehension (Graesser, McNamara, & Louwerse, 2003; Sanders, Spooren, & Noordman, 1993; Van den Broek, Lynch, Naslund, Levers-Landis, & Verduin, 2003). In other words, basic reading skills such as decoding, fluency, and vocabulary knowledge are not enough to arrive at a coherent textbase and situation model representation. Integration and inference processes and skills are crucial (Cain, Oakhill, & Bryant, 2004; Kendeou, Van den Broek, White, & Lynch, 2009; Zwaan & Singer, 2003). These higher-order processes and skills include the abilities to make inferences, to maintain coherence and integrate information in previous information, the understanding of concepts and ideas conveyed by the text, and the activation of background knowledge to fill in missing details (Bohn-Gettler, Rapp, Van den Broek, Kendeou, & White, 2011; Oakhill, 1994; Oakhill, Cain, & Bryant, 2003; Perfetti, 1985; Rapp et al., 2007). Inadequate higher-order skills seem to be among the major obstacles for students in secondary education when learning new information from study texts (Linderholm et al., 2000; Magliano, Millis, Ozuru, & McNamara, 2007).

1.2 Text and reader characteristics

If we want to improve text comprehension, it seems useful to guide readers in connecting one text idea to another, and to relevant background knowledge (McMaster et al., 2012), and thereby reduce the need for making inferences. To do this, educative publishers and textbook writers can use visual cues, such as a continuous layout which signals that sentences are connected instead of isolated pieces of information.

For many developing readers parsing phrases and clauses seems to be a problem (Fuchs, Fuchs, Hosp, & Jenkins, 2001; LeVasseur, Macaruso, Palumbo, &

Shankweiler, 2006). Consider, for example, the clause *tijdens het baden* ('during bathing') in example (1).

(1) Ze deden dure badolie in het water. *Daardoor* rook het water lekker tijdens het baden.
'They put expensive bath oil in the water. *As a result*, the water smelled nice during bathing.'

The clause should be parsed as one unit, but it is interrupted by a line break. Therefore, reading the clause can only be completed after a return sweep to the next line. Rayner (1998) has shown that moving ones' eyes from the end of one line to the beginning of the next is a complex maneuver. LeVasseur et al. (2006) note that, in particular for inexperienced readers, the interruption in processing necessitated by a return sweep can be disrupting. The authors found that when a line break corresponds to a clause boundary, as in example (2), false starts were reduced and higher fluency ratings were achieved.

(2) They put expensive bath oil in the water.
As a result, the water smelled nice during bathing.

Educational publishers can use coherence markers that show how sentences are connected, such as connectives (*because*, *moreover*, *but* and *also*) and signaling phrases (*a solution for this problem* and *the result is that*). These linguistic indicators that are part of the surface code signal the nature of the relation between events and help readers identify text structure (Degand & Sanders, 2002; Goldman, 1997; Graesser & McNamara, 2011; Halliday & Hasan, 1976; Sanders, Land, & Mulder, 2007; Sanders & Noordman, 2000; Sanders et al., 1993). Coherence markers explicitly show the reader how one part of a text is related to another in at least three ways, which are related to the three levels of representation described above (Cozijn, Noordman, & Vonk, 2011). First, coherence markers inform readers about the structure of the incoming material as they indicate the start of a new syntactic structure. This function of coherence markers is related to the surface code representation. Second, connectives and signaling phrases have an integrative function, because they facilitate the integration of new clauses with previous clauses by specifying the relation between them. Myers, Shinjo and Duffy (1987) consider connectives as relational propositions which in example (2) show that p_2 is caused by p_1. Hence, this integrative function can be related to the textbase (Britton, 1994; Degand & Sanders, 2002; Haberlandt, 1982; McNamara et al., 1996; Sanders & Spooren, 2007). Third, coherence markers trigger readers to infer additional textual information by checking the relations expressed by these markers. During the construction of the situation model, readers integrate text information with background knowledge (Kintsch, 1998; Mulder & Sanders, 2012; Noordman

& Vonk, 1997; Sanders et al., 1993). That's why the inference function can be related to the situation model.

Adding connectives and other coherence markers seems to be fruitful, because several studies have shown that coherence markers may influence the representation skilled adult readers build while reading a text with such markers. The presence of markers in expository texts results in better recall (Lorch & Lorch, 1986), faster and more accurate response on a prompted recall task (Millis & Just, 1994), faster response on verification tasks (Sanders & Noordman, 2000), as well as higher scores on comprehension tasks (Degand & Sanders, 2002; Kamalski, Sanders, & Lentz, 2008; McNamara et al., 1996). Therefore, these markers may be important for acquiring new information from text.

However, there is a tension between the presence of coherence markers and sentence length. In a corpus-based study, Land (2009) shows that many Dutch textbook writers shorten sentences and remove coherence markers, because sentence length increases when coherence markers are inserted, and because they believe longer sentences tend to place more demands on working memory and should therefore be regarded more difficult (see Graesser, McNamara, Louwerse, & Cai, 2004 for similar practices in the US). As a result of shortening sentences, coherence gaps occur. To fill in or explain these gaps readers have to determine the type of coherence relations between the new clause and previous clause themselves, which results in increased reading times or low scores on comprehension tasks (Best, Rowe, Ozuru, & McNamara, 2005; Graesser & McNamara, 2011; Sanders & Noordman, 2000).

In the above-mentioned studies, however, only undergraduate students participated, who are more experienced, more mature and more skilled in reading. For younger readers, fewer studies have shown the benefit of coherence markers. For children at elementary school, Beck, McKeown, Sinatra, and Loxterman (1991) found that text revisions which reduce the number of inferences had positive effects. Specifically, children's scores increased when they read the revised version of a history text in which explanations were added, connections were made explicit, events and people were more directly linked, and given information generally preceded new information. In addition, Loman & Mayer (1983) found that 15- to 17-year olds benefit more from texts in which the causal structure was more overt, that is inserting connectives, signaling phrases and headings that highlighted main ideas increased recall of relevant idea units and decreased recall of irrelevant idea units. More recently, Land and colleagues (Land, 2009; Sanders et al., 2007) have shown that readers from the pre-vocational level in secondary education benefit from the presence of cohesive signals. In their study, students benefited from texts in which coherence markers were added and sentences were presented in a continuous layout instead of each sentence on a new line.

1.3 Current study

In order to be able to provide evidence-based guidelines for educational publishers, the current study investigates the effects of two text features on the text comprehension of students in secondary education: the presence versus absence of connectives and the presentation of information and a continuous versus a discontinuous layout with each sentence on a new line. The following research questions are central to this study: (1) do the presence of connectives and a continuous layout improve the text comprehension of secondary education students? And (2) do pre-vocational students benefit more from coherence markers than pre-university students?

In earlier research, different coherence markers were manipulated simultaneously, such as adding descriptive elaborations, topic headers, theme sentences and connectives, rearranging words and sentences, and layout. For example, Land (2009) found that students from lower levels in secondary education benefited from integrated texts (explicitly marked and with a continuous layout). However, coherence marking and layout were manipulated simultaneously. In the current study, we will separate the presence of coherence markers and layout and make sure that the effects of these different text features can be singled out.

Furthermore, while investigating the effects of coherence markers we only focus on the effects of connectives. Using connectives is probably the most natural way to make coherence relations explicit (Pander Maat & Sanders, 2006). They are frequent in natural language use in all types of discourse. Our research will be limited to additive and causal connectives. These positive connectives are known to have a facilitative effect during on-line processing (Canestrelli, Mak, & Sanders, 2012; Millis & Just, 1994; Sanders & Noordman, 2000). That's why we prefer positive connectives (*and*, *also*, *because*, *that's why*) over negative (adversative) connectives (*but*, *although*, *however*). Another reason is that adversative connectives often seem to be a prerequisite for constructing negative relations, because without them, the adversative relation cannot be established in the first place (Sanders & Canestrelli, 2012).

Moreover, in several studies a variety of tasks was used to measure text comprehension, such as recall tasks, verification tasks, multiple-choice versus short-answer tasks, and text-based versus inference tasks. In accordance with Magliano et al. (2007), we propose that comprehension measures should be selected on the basis of the text type, the ability levels of the target readers and the comprehension products the measure is designed to observe. We want to measure higher-order processes, such as inferencing and establishing coherence. Furthermore, we investigate whether coherence marking and layout influence performance on text-base and/or situation model tasks. That's why we select multiple comprehension

tasks that each assess other facets of deeper text comprehension: global textbase questions that tap the textbase representation, and inference questions and sorting tasks that tap the situational level of representation.

Finally, we explore whether coherence marking and layout interact with school level, because Dutch textbook writers design textbooks for each school level separately. For example, while the content is more or less the same, pre-university students use other textbooks than pre-vocational students. Moreover, within pre-vocational education, higher-level students use other textbooks than lower-level students.[1] Land (2009) found that textbooks used in the lowest levels of pre-vocational education are the least cohesive ones in which short sentences were combined with a discontinuous layout, with each sentence starting at a new line. Schoolbook writers argue that texts with short sentences place fewer demands on working memory and, therefore, are particularly helpful for less-skilled adults and students from lower educational levels. However, schoolbook writers do not base their claims on evidence from processing or comprehension studies. The results of our study will allow us to give evidence-based advice to educational publishers.

Based on Snow's (2002) framework on text comprehension and findings among adolescents and adults, we expect the presence of coherence markers to be helpful for students in secondary education as well, despite of increased sentence length. More specifically, we expect positive effects of adding connectives on two levels of comprehension: the textbase and the situation model (see also Section 1.2.).

First, we expect that after students will obtain higher comprehension scores on the true-false statements when reading explicit texts. In order to answer these global textbase questions correctly, students have to encode relations between propositions. Because connectives are relational propositions (Myers et al., 1987), they facilitate the integration of propositions, which leads to higher scores.

Second, we expect students to obtain higher scores on inference questions that tap the situational level of representation. Connectives instruct students to establish a specific type of coherence relation, and increase the possibility that they activate and use their background knowledge about the text ideas. In contrast, when there are no connectives, students are dependent on their relatively weak higher-order skills (i.e. inference making and integrating), and we expect that establishing coherence will be more effortful and less successful.

---

1. Dutch pre-vocational education (*vmbo*) is divided into four levels ranging from mainly theoretical to mainly vocational training. In this study, the two highest levels are collapsed, because they mainly concern theoretical education and use similar textbooks (vmbo-high: Dutch *vmbo-tl* and *vmbo-gl*). The two lowest levels are also collapsed, because they mainly concern practical education and use similar textbooks (vmbo-low: *vmbo-bk* and *vmbo-bb*).

Third, we also expect that students' scores on global sorting tasks will improve when students comprehend the coherence relations between adjacent clauses. Koornneef and Sanders (2013) found that participants observe inconsistencies between a pronoun and its antecedent in an earlier stage when there is a causal connective than when there is no connective. Therefore, we expect that students benefit from connectives when filling out sorting tasks in which they have to link traits and events to the correct referents. In addition, we hypothesize that these effects will be larger for students in lower educational levels, because the lower the educational level, the less theoretical education students have, the fewer texts they have to read, and the less reading experience they have (Hacquebord, 2004; Land, 2009; Schram, 2002). Therefore, we expect students from pre-vocational education to benefit more from connectives than the other students.

As regards layout, we conducted our investigation to test two competing hypotheses. On the one hand, we know that sentence fragmentation causes visual breaks. As a result, readers will process the text in a discontinuous lay-out as a list structure instead of a meaningful temporal/causal structure, and previous studies have shown that additive relations are weaker connections than causal relations (cf. Sanders & Noordman, 2000). Confronted with a discontinuous layout, readers will therefore be less encouraged to use a strategy to construct a highly connected representation. As a result, their comprehension scores will decrease both on text-base and situation model tasks.

On the other hand, a discontinuous layout with one clause on each line has a better arranged view and the processing of phrases and clauses could be easier because return sweeps only occur after complete clauses. If fragmentation reduces the number of false starts and makes corrective eye-movements unnecessary (cf. LeVasseur et al., 2006), one possibility is that students' processing load will reduce, which can have a positive effect on the texts' meaning representation.

2. Method

2.1 Students

134 students from the eighth grade of two Dutch secondary educational schools participated in this experiment. Students were from different school levels: 67 students from higher educational levels (Dutch havo or vwo) and 67 students from pre-vocational education (Dutch vmbo). Because vmbo is divided into four levels ranging from mainly theoretical (high level, vmbo-tl) to vocational training (low level, vmbo-bb), we selected students from the higher educational levels and students from the lower levels. In sum, we distinguished three experimental groups:

Table 1. Student demographics per school level

	Pre-vocational low (N = 41)	**Pre-vocational high (N = 26)**	**Pre-university (N = 67)**
Age	13 year, 5 months	13 year, 3 months	13 year, 1 month
Gender	31 boys, 10 girls	13 boys, 13 girls	36 boys, 31 girls
Native speakers Dutch	95% (N = 39)	100%	91% (N = 61)
Dyslexia	26.8% (N = 11)	15.4% (N = 4)	13.4% (N = 9)

havo/vwo (67 students), vmbo-high (26 students) and vmbo-low (41 students). Student were 12 and 15 years old (mean 13 years and two months). All but eight students were native speakers of Dutch. About one sixth of the participants was dyslectic. The student demographics per school level are provided in Table 1.

2.2 Materials

2.2.1 *Texts*

The experimental material consisted of four Dutch history texts of approximately 290 words (min. 260/max. 317). We intentionally focused on study texts from history text books used in the vmbo, because these books contain a large amount of textual information and many less-skilled students are unable to read these texts accurately (Bowen, 1999; Land, 2009). History texts are expository in the sense that they convey information, but they have a structure that is narrative in form (Perfetti, Britt, & Georgi, 1995). These history texts have a greater structural complexity, and a greater information density than narrative fiction texts (Ben-Anath, 2005; Coté, Goldman, & Saul, 1998; Spyridakis & Standal, 1987). Topics included hygiene in the Middle Ages, slave trade in the 17th century, wigs in the 18th century, and feminism. We only included topics that had not been taught during the first years of secondary education.

There were two independent variables: the presence or absence of connectives, and layout. To make local coherence explicit, additive and causal connectives were added to signal relations between sentences or clauses. Example (3) shows an additive relation marked by an additive connective, and (4) its implicit counterpart. Example (5) shows a causal relation marked by a causal connective, and (6) its implicit counterpart.

(3) Zulke pruiken waren een stuk goedkoper. *Bovendien* bleven de krullen bij een regenbui goed zitten.
'Such wigs were a lot cheaper. *Moreover*, the curls remained well in place during rain.'

(4) Zulke pruiken waren een stuk goedkoper. De krullen bleven bij een regenbui goed zitten.
'Such wigs were a lot cheaper. The curls remained well in place during rain.'

(5) Arbeiders gingen bijna nooit in bad, *want* een bad was veel te duur.
'Workers rarely took a bath, *because* bathing was way too expensive.'

(6) Arbeiders gingen bijna nooit in bad. Een bad was veel te duur.
'Workers rarely took a bath. Bathing was way too expensive.'

The number of additive and causal connectives in the explicit versions ranged from eight to fourteen. We selected more causal than additive connectives, because several studies have shown that causal relations are a particularly important type of coherence relations (Kendeou et al., 2009). Furthermore, we wanted to include different subtypes of causal connectives: we added markers of cause-consequence (*therefore*, *as a result*) and consequence-cause relations (*because*). Note that both explicit and implicit versions could contain (one to three) temporal (*subsequently*) and contrastive connectives (*but*). Furthermore, we only selected connectives in clause-initial position.

Apart from the presence of connectives, only word order was varied between the implicit and explicit versions. This was dictated by the fact that the word order in Dutch main clauses differs from the one in subordinate clauses. In example (7) for instance, the explicit version contains a subordinate clause with the finite verb *verdiende* ('earned') near the end of the clause, while the implicit version in (8) contains an unmarked main clause with *verdiende* after the subject.

(7) Soms moesten vrouwen ook werken, *omdat* hun man te weinig verdiende.
'Sometimes, women had to work too, *because* their husband earned too little.'

(8) Soms moesten vrouwen ook werken. Hun man verdiende te weinig.
'Sometimes, women had to work too. Their husband earned too little.'

The second variable we manipulated was layout: a continuous vs. discontinuous layout. In the continuous versions, sentences were presented in a continuous layout. In the discontinuous versions, each main clause, coordinate (3) and subordinate clause (7) started at a new line. Example (9) shows the translation of a text passage in a discontinuous layout, while (10) shows the same passage in a continuous layout.

(9) The man was seen as the head of the family.
The woman was subordinate to the man.
That's why the woman had to listen to him.
Men were the principal breadwinners,
because the women were supposed to take care of the families.

(10) The man was seen as the head of the family. The woman was subordinate to the man. That's why the woman had to listen to him. Men were the principal breadwinners, because the women were supposed to take care of the families.

In sum, we constructed four experimental versions for each of the four texts in our study: (1) an explicit text with a continuous layout, (2) an explicit text with each clause starting at a new line, (3) an implicit text with a continuous layout and (4) an implicit text with each clause starting at a new line. A complete version of one of the explicit texts can be found in Appendix A. All students read four texts (one version per text), and text versions and text order were randomly assigned to students.

2.2.2 *Comprehension questions*

For each text, we constructed different types of comprehension questions: true-false statements (N = 6); bridging inference questions (N = 3) and sorting or time ordering tasks (N = 1). These tasks specifically addressed the textbase (true-false statements) or the situation model representation (bridging inference questions and sorting tasks) (Kamalski et al., 2008; Kintsch, 1994; 1998; Land, 2009; Magliano et al., 2007; McNamara, 2001; McNamara et al., 1996). Understanding texts at the situation model level requires active inferencing and adequate prior knowledge (McNamara et al., 1996), while knowledge about time ordering and causality is also crucial (Zwaan & Radvansky, 1998). For example, the participants in our experiment came across the following sentence: *In the Middle Ages diseases spread very fast, because kitchens, basements and cupboards were full of rats.* To construct a situation model representation of this sentence, readers had to integrate these clauses with their knowledge about causality. Specifically, the cause of fast spreading of diseases is the presence of many rats. Also, general prior knowledge about the Middle Ages, diseases, rats and storages is needed. Readers have to be acquainted with the period of the Middle Ages, and they also have to know that the rats might have lived in the kitchens, basements and cupboards because the food was stored there.

First, we constructed true-false statements. The information for answering those questions could be found in separate sentences of the text. Figure 1 shows an example of these true-false statements.

In terms of processing, students had to recall information from the appropriate propositions in the text. Thus, students had to connect multiple propositions in the textbase representation. For example, in order to correctly verify the fifth statement in Figure 1, *The floor was strewn with fragrant herbs*, readers had to know that this clause is about luxurious toilets and not about bathrooms in the fifteenth

In the Middle Ages, it was very expensive to have a bath. For what reason? Check each possible reason and mark with a cross whether this reason is true or false.		
	True	False
1. It was difficult to get hot water.		
2. The towels were very expensive.		
3. They used expensive bath oil.		
4. It was very expensive to make a bathtub.		
5. The floor was strewn with fragrant herbs.		
6. They used expensive linen cloths on the floor.		

Figure 1. Example of a true-false statement task (hygiene in the Middle Ages)

century. For some statements, students also had to construct a paraphrase or summarization of the propositions, and then assess whether the given answer option matched that paraphrase or summarization (Magliano et al., 2007).

Furthermore, we designed open-ended bridging inference questions. In these questions integration of information from at least two clauses is necessary (Best, Floyd, & McNamara, 2008; McNamara, 2001; McNamara et al., 1996; O'Reilly & McNamara, 2007). An example of a bridging inference question is: *In the Middle Ages many people were ill. Diseases spread from person to person. How did that happen? Diseases spread from person to person, because…* The information for answering such questions was stated in separate sentences of the text (see Appendix A). In contrast to true-false questions, open-ended questions require readers to access the memory representation of a text and retrieve and produce the relevant information themselves. Thus, true-false statements can be answered partly on the basis of recognition and reasoning, and provide richer retrieval cues than open-ended questions (Magliano et al., 2007).

However, for constructing a deeper understanding of the text, readers also had to establish connections between currently processed information and information occurring much earlier in the text (Kintsch, 1998). Therefore, we constructed sorting tasks in which students had to categorize a set of sentences from the text into groups (Kamalski, 2007; Kamalski et al., 2008; McNamara et al., 1996; Land, 2009). In contrast to the more locally focused types of comprehension tasks, students had to establish relations between central text ideas in order to construct a more global understanding of the text. Both the labels of the concepts and the sentences were given, in contrast to the sorting tasks used by McNamara et al. (1996) in which students had to put cards with key words into piles according to how they thought the concepts should go together. Figure 2 shows an example of such a sorting task.

For one text (slave trade), we constructed a time ordering task (see Figure 3), because a sorting task with different labels and concepts was less appropriate. Students had to put the numbers of the sentences in the correct boxes of the

Below you see four circles with an important word from the text. You also see nine sentences from the text. Put the number of the sentence in the correct circle. Each number fits in one circle. The first sentence has already been given.

1. At night they used a chamber pot.
2. They were able to get hot water.
3. They did many things to clean their surroundings.
4. They rarely pooled.
5. They had problems with rats in their kitchen, basement and cupboards.
6. They never washed their linen clothes.
7. They used hot water, linen and bath oil.
8. They used herbs to ban rats and lice.
9. They were dying of The Plague.

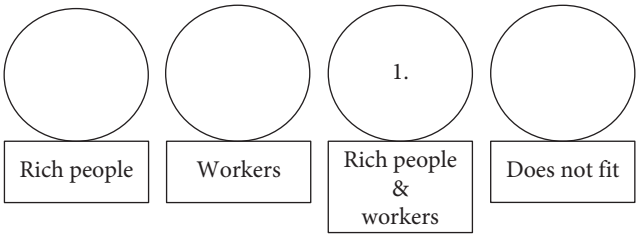

Figure 2. Example of a sorting task (concerning a text about hygiene in the Middle Ages)

Below you see nine sentences. Put the sentences in the correct boxes, that is to say in the correct time order. Each sentence fits in one box. One sentence doesn't fit in any of the eight boxes. Put thissentence in the circle. The sixth sentence has already been entered.

1. They give weapons and ammunition to African tribes.
2. They sell their products for a lot of money.
3. They sail to Africa with weapons and other stuff.
4. They sell the slaves on the market.
5. They bring their products to Europe.
6. They make the slaves work on the plantations.
7. At their slave forts, they get their slaves from African tribes.
8. They create wooden bows and sticks.
9. They transport the slaves across the sea to the slave market

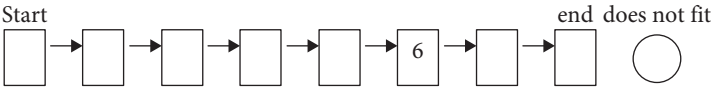

Figure 3. Example of a time ordering task (concerning a text about the Dutch slave trade)

timeline. Both the sentences and the timeline were given. To order the sentences correctly, students had to use their world knowledge as well as the knowledge they obtained from the text about the temporal sequence of central text ideas and events. In the text, the chronological order of text ideas differed from the order in which these ideas were presented in the text.

2.3 Pretest

All texts and comprehension tasks were pretested at three schools. 126 students from the eighth grade participated (non-study participants). Each of the 67 students from the pre-vocational education and the 59 students from the pre-university education tested two of the four texts. The reliability of the three types of situation model questions was not high: for bridging inference questions $.30 < \alpha < .46$, for true-false statements $.19 < \alpha < .47$ and for sorting tasks $.28 < \alpha < .84$. On the basis of these results, difficult words were replaced, such as *feminism* in a heading, and several questions were modified or replaced when they were too easy (more than 90% correct response) or too difficult (less than 25%), or when they led to a decrease in the overall reliability.

2.4 Procedure

All tests were administered at school. Each student was tested individually in a quiet room. Students were instructed that they should read the texts as if they were studying for a test and that, after reading, they would have to answer several comprehension questions without looking back to the text. Furthermore, they were informed about the number of texts, and they were presented with an example of how to complete the sorting tasks. All students read two texts in a session. They received a booklet with some background questions about nationality, age, gender, first (and second) language, school level and dyslexia and read two texts. Texts were cut up and presented on three slides on a PC monitor. When students finished reading a text slide, they pushed a button and the next text slide was presented on the screen.

Then the questions about the two texts followed in the order in which the texts were read. Comprehension questions of the first text were not given immediately after reading this text, because we were not interested in text recognition and retrieval (i.e. the surface code representation). After answering the questions and a five-minute break, the students had to read the next two texts. The whole procedure was then repeated. The duration of one session with two texts was between eighteen and thirty minutes. The experiments were conducted over a period of two weeks.

2.5 Analyses

To examine the answers to the comprehension questions a correction model was developed. A score of one point was assigned to each correct answer to inference questions or true-false statements. Also, a score of one point was assigned to each well-placed sentence in the sorting task, with a total of eight points. In the time ordering task one point was scored for each correct adjacent pair. For each student, we computed the percentage of correct response for each comprehension task separately.

For each comprehension task, we performed separate multilevel model analyses, because each comprehension score is nested both within students, and within texts. Traditional ANOVAs are not fully equipped to handle these random factors at the same time. As a result, the total variance and the standard error will be underestimated, which increases the chance of making a type 1 error. Our multilevel model analyses included eight models. In the second to the fourth model, the percentage of correct response was modeled as a function of the predictors coherence marking, layout and school level. The respective two-way interactions, and the three-way interaction between coherence marking, layout and school level were taken into account in the fifth to eighth model respectively. The model also estimated between-person variance (because one person can comprehend the texts better than another person) and text variance (because one text can be more comprehensible than another text). For each comprehension task, we selected the best fitting model with the smallest number of parameters and the best fit to the data (see Appendix B).

3. Results

The answers were independently examined by two raters: the first author and a research assistant. The interrater reliability was relatively high (Cohen's Kappa: overall .98, bridging inference questions .92, true-false statements .98, and sorting tasks .99). Occasional differences in rating were discussed until agreement was reached. For the multilevel analyses, one final dataset was used.

Per text, the sorting tasks measured the underlying construct in a reliable way ($.49 < \alpha < .89$), while bridging inference questions and true-false statements were less reliable, but sufficient for experimental purposes ($.51 < \alpha < .61$ for bridging inference questions and for true-false statements $.51 < \alpha < .56$). Because all students read four texts (albeit a different version per text), we also investigated whether the comprehension questions measure the same theoretical construct *across* all four texts, and calculated their reliability from the relative proportion of between person variance

as compared to the residual variance (cf. Lord & Novick, 1968; Raudenbush & Bryk, 2002).[2] The reliability for the sorting tasks, inference questions and true-false statements together was low ($\rho = .49$). In accordance with other studies (Land, 2009; McNamara et al., 1996), the reliability score shows that these different comprehension tasks measure different aspects of text comprehension. The reliability of the different comprehension tasks separately appeared acceptable for the sorting tasks ($\rho = .59$), the inference questions ($\rho = .72$) and the true-false statements ($\rho = .61$). For that reason, each comprehension task was analyzed separately.

3.1 True-false statements

For the true-false statements, the model with the fixed parameter 'school level' was the best fitting model. That is, neither the main effects of coherence marking and layout nor the two- or three-way interactions made a significant contribution to the models. The mean comprehension scores on the true-false statements per condition are provided in Table 2.

Table 2. Mean scores and standard errors on the true-false statements per condition

	Implicit		Explicit	
	Discontinuous	**Continuous**	**Discontinuous**	**Continuous**
Pre-vocational low	64.54 (3.53)	61.32 (3.64)	67.47 (3.76)	63.52 (3.64)
Pre-vocational high	62.42 (4.44)	67.80 (4.35)	63.90 (4.51)	66.14 (4.35)
Pre-university	77.47 (3.01)	75.87 (2.92)	77.94 (2.98)	75.17 (3.02)
Variances				
$S^2_{residual}$	277.93	(21.51)		
$S^2_{persons}$	73.32	(19.49)		
S^2_{text}	12.17	(10.38)		

We found a main effect of the school level ($F(2, 130) = 15.30$, $p < .001$; Cohen's $d = 1.45$):[3] pairwise comparisons show that havo/vwo-students performed better

2. With the parameter estimates obtained from our model, the reliability can be calculated. For each type of questions, inference questions, true-false statements and sorting tasks, the between person variance must be divided by the between person variance added up with the item variance divided by the number of items. This gives a reliability coefficient (ρ) between 0 and 1, indicating the relative proportion of systematic variance as compared to the non-systematic variance (cf. Kamoen, 2012).

3. The effect size of an effect is often classified in relation to the standard deviation (Cohen, 1988). The effect size we report here is based on the between-person standard deviation.

than students from vmbo-high ($p = .001$) and students from vmbo-low ($p < .001$).[4] No differences were observed between students from vmbo-high and from vmbo-low ($p = 1.00$).

3.2 Inference questions

For inference questions, the fixed parameters 'coherence marking' and 'school level' made a significant contribution to the multi-level model. Table 3 presents the mean comprehension scores on the inference questions.

Table 3. Mean scores and standard errors on the inference questions per condition

	Implicit		Explicit	
	Discontinuous	**Continuous**	**Discontinuous**	**Continuous**
Pre-vocational low	36.46 (5.41)	36.88 (5.52)	38.86 (5.64)	35.74 (5.52)
Pre-vocational high	49.20 (6.37)	44.92 (6.28)	49.23 (6.44)	54.89 (6.28)
Pre-university	58.39 (4.89)	61.73 (4.81)	70.26 (4.87)	67.11 (4.91)
Variances				
$S^2_{residual}$	413.08	(31.99)		
$S^2_{persons}$	138.08	(32.52)		
S^2_{text}	58.47	(43.93)		

There was a main effect of the school level ($F(2, 129) = 36.32$, $p < .001$; Cohen's $d = 2.34$), indicating that havo/vwo-students performed better on inference questions than students from vmbo-high ($p = .001$) and vmbo-low ($p < .001$), and that students from vmbo-high performed better than students from vmbo-low ($p = .01$). As expected, there was also a main effect of coherence marking ($F(1, 345) = 8.52$; $p = .004$; Cohen's $d = 0.41$): students' comprehension was better for the explicit texts than for the implicit texts.

3.3 Sorting tasks

For sorting tasks, the fixed parameter 'school level' and the three-way interaction between coherence marking, layout and school level made a significant contribution to the multi-level model. Table 4 shows the mean comprehension scores on the sorting tasks.

4. The pairwise comparisons were corrected by the Bonferroni correction, in order to reduce the chances of obtaining false-positive results (type I errors).

Table 4. Mean scores and standard errors on the sorting tasks per condition

	Implicit		Explicit	
	Discontinuous	**Continuous**	**Discontinuous**	**Continuous**
Pre-vocational low	46.83 (4.29)	44.20 (4.43)	50.85 (4.58)	47.12 (4.43)
Pre-vocational high	54.35 (5.41)	47.77 (5.30)	59.56 (5.50)	59.56 (5.30)
Pre-university	64.05 (3.64)	75.60 (3.53)	72.28 (3.60)	66.61 (3.66)
Variances				
$S^2_{residual}$	445.68	(34.63)		
$S^2_{persons}$	78.48	(26.99)		
S^2_{text}	16.95	(14.79)		

There was a main effect of the school level ($F(2, 127) = 31.43$, $p < .001$; Cohen's $d = 2.53$), showing that havo/vwo-students performed better than students from vmbo-high ($p < .001$) and students from vmbo-low ($p < .001$). No differences were observed between students from vmbo-high and vmbo-low ($p = .10$). Furthermore, the three-way interaction was significant ($F(9, 345) = 1.92$; $p = .048$).

We therefore constructed a multilevel model analysis for havo/vwo-students and a multilevel analysis for students from both vmbo-levels, because no difference in scores were observed between vmbo-high and vmbo-low. The model with the fixed parameter 'coherence marking' was the best fitting model for the scores of the vmbo-students: they performed better when they read explicit versions than when they read implicit versions ($F(1, 162) = 4.12$; $p = .04$; Cohen's $d = 0.67$).

For the scores of havo/vwo-students, the two-way interaction between coherence marking and layout made a significant contribution to the multi-level model ($F(1, 177) = 3.35$; $p = .02$; Cohen's $d = 1.19$). When texts were presented in a continuous layout, havo/vwo-students did not benefit from connectives. In fact, they obtained higher scores when coherence relations were implicit: pairwise comparisons show that students performed better when reading implicit continuous texts compared to explicit continuous texts ($p = .03$). Furthermore, havo/vwo-students benefited more from explicit discontinuous texts and implicit continuous texts than from implicit discontinuous texts (respectively $p = .005$ and $p = .04$). This indicates that these students need at least some form of coherence marking (either presence of connectives or a continuous layout) to properly understand the text.

4. Discussion

The results of this study helped to gain insight into the role of cohesive devices in text comprehension. The most important findings in this study are as follows:

neither the layout nor the presence or absence of connectives had an effect on the scores of true-false questions. For bridging inference questions however, students' comprehension improved when texts contain connectives, regardless of the school level. Students of the pre-vocational level also benefited from connectives when completing sorting tasks regardless of the layout. For pre-university students, the results are more complex. Below, we will discuss these findings in more detail.

4.1 The influence of connectives and layout

Connectives lead to better answers to two of the three comprehension tasks in our study. Therefore, this study predominantly confirms our hypothesis, that is connectives can definitely be said to be an important factor in creating comprehensible texts for readers in educational contexts. However, the effects partially differ per school level.

We have shown that the presence of additive and causal connectives is especially beneficial to students in pre-vocational education. The presence of connectives leads to better answers to both inference questions, which were used to measure coherence relations between consecutive clauses, and sorting tasks, which were used to measure coherence between the main ideas of a text. Similar patterns were found in other studies with undergraduates and children (Beck et al., 1991; Britton & Gülgöz, 1991; Linderholm et al., 2000). In these studies, history texts were revised by making inferences and links among text ideas explicit, in particular causal ones. However, in none of these studies readers from the lowest educational levels in secondary schools participated. Only Sanders et al. (2007) have shown that readers from pre-vocational education benefit from coherence markers. In the above-mentioned studies, various cohesive signals were added simultaneously to help readers establish coherence relations. In our study, we disentangled the influence of the layout and the presence of coherence markers. In the case of textual coherence markers, we used additive and causal connectives.

Pre-university students also benefit from connectives when answering bridging inference questions. However, on sorting tasks, the results are less clear. For continuously presented texts, pre-university students' comprehension only increases when coherence relations are implicit. As long as texts are presented in a natural, continuous way, pre-university students seem to be able to deal with a minimally coherent text that requires coherence-establishing inferences. They do not need connectives to make correct inferences and to construct relations for building a macrostructure of the text, and therefore, for a coherent situation model of the global text meaning. Arguably, these findings are to a certain extent comparable to those in studies in which only readers with low domain knowledge benefit from cohesive texts (Kamalski, 2007; Kamalski et al., 2008; McNamara,

2001; McNamara et al., 1996). When readers with high domain knowledge read texts with coherence markers, they perform less on comprehension tests. The researchers argued that an explicit text causes these readers to process the text passively, because they do not need to make many inferences (cf. Gilabert, Martinez, & Vidal-Abarca, 2005). Linderholm et al. (2000) found that readers only benefit from causal repairs when reading a difficult history text compared to an easy one, and that causal text repairs were particularly beneficial to less-skilled readers.

Havo/vwo-students are less successful in building a macrostructure of implicit discontinuous texts, as we found for vmbo-students. It seems as if a discontinuous presentation of sentences serves as a visual break that prompts the reader to start a new content structure. As a result, students need connectives for a global understanding of discontinuous texts, while continuously presented texts — without connectives — also help improve their performance on sorting tasks.

However, we cannot explain why we did not find a difference in scores between implicit discontinuous texts and explicit continuous texts. One explanation for this finding is that the effects of text features such as connectives and layout depend on the operationalization of the readers' global situation model. Even though the reliability across texts is high, different tasks were used depending on the text structure. We constructed time ordering tasks as well as sorting tasks. Furthermore, we do not exactly know how reader characteristics interact with text features. The differences between pre-university and pre-vocational students can be explained by the fact that pre-university readers have more skills and strategies to answer the items correctly. As a result, the effects of text features such as connectives and layout are less robust for these readers. Therefore, further research is needed to determine whether time ordering and sorting tasks are appropriate for measuring effects of text features and reader characteristics.

Our findings show sharp contrast with the current practice of educational publishers in The Netherlands. Many publishers select and design school texts on the basis of short words and short, fragmented sentences, because they believe that such texts place fewer processing demands on lower-level cognitive process (Land, 2009). However, our research shows that the lack of connectives places other demands on the reader, such as the need to make inferences about the nature of the coherence links between clauses, a major obstacle for students in secondary education. In addition, we found that a discontinuous layout has no facilitative effect on text comprehension, while many textbook writers design texts in a discontinuous layout, particularly for students from the lowest levels of pre-vocational education.

4.2 Text comprehension measures

In this study we used several comprehension tasks to investigate the effects of connectives on text comprehension. We can conclude that both tasks that tap the situational level of representation, open bridging inference questions and sorting tasks, are sensitive enough to measure the effects of coherence markers (see also Degand & Sanders, 2002; Kamalski, 2007; Kamalski et al., 2008; Land, 2009; McNamara, 2001; McNamara et al., 1996). We found that connectives help students in secondary education to establish coherence between consecutive sentences (bridging inference questions), and in particular readers in pre-vocational education to establish relations between central text ideas in order to construct a more global understanding of the text (sorting tasks).

However, our true-false statements did not show any effects of coherence marking. One of the explanations could be that the true-false format is not an optimal format for measuring the effects of coherence marking, because readers only have to retrieve and recognize text information. More specifically, students only had to verify whether a statement was true or false. In contrast to answer open-ended questions, readers had to generate the answer themselves. Our data support this idea, because the comprehension scores on true-false statements are much higher than the scores on the other tasks, in particular for vmbo-students. Other studies have also shown that more global comprehension measures, such as free recall, are not sensitive enough to find effects of coherence marking (McNamara et al., 1996; Spyridakis & Standal, 1987).

An alternative explanation for our results is that our true-false statements are measurements of readers' textbase. Kintsch and Van Dijk (1978) argue that the most important criterion for constructing a coherent textbase is establishing referential coherence. Readers have to embed one proposition into another on the basis of a shared argument (p. 365). Contrary to our hypotheses, our findings indicate that coherence relations that link these propositions are not part of the textbase. As a result, connectives do not seem to contribute to establishing referential coherence, but only help readers to establish relational coherence. This assumption corresponds to the idea of Mulder and Sanders (2012) that coherence relations, whether or not explicitly realized in the text, are represented at the level of the situation model. Thus, it seems that coherence relations are part of the readers' mental representation of the events described in the text, whereas relations between propositions can be represented at the textbase level, which may explain the lack of effects of connectives on the true-false statements. Our data also support the idea of McNamara et al. (1996) that different comprehension tasks measure different facets of text comprehension: the overall reliability for the

sorting tasks, inference questions and true-false statements together was low, and there was considerable variance in comprehension scores caused by task type.

4.3 Further research

This study still leaves various questions unanswered, which are worth of investigating in future research. First, we found that the effects of connectives and layout on global text comprehension differ for pre-vocational students and pre-university students. However, we do not exactly know why pre-university students have higher comprehension scores in general, and why they do not need connectives as long as texts are presented in a natural, continuous way. It could be the case that pre-university students have better reading skills, and that they are better capable of making inferences. Another possibility is that they have a better memory for the text. A third possibility is that they have more prior knowledge than pre-vocational students. Thus, it is posited that it is important to conduct studies that investigate to what extent prior knowledge and reading skills interact with coherence marking. McNamara et al. (1996), for example, found different effects of coherence marking for skilled and less-skilled adult readers.

Second, it is highly important to conduct additional studies into the dynamics of the cognitive processes of readers in secondary education and the influence of coherence markers and layout on their on-line reading process. Eye-tracking technology is one of the methods to investigate these effects. There is empirical support for the position that coherence markers play a facilitating role during the reading process of skilled adults. Specifically, connectives lead to faster processing of immediately following text segments (Canestrelli et al., 2012; Haberlandt, 1982; Maury & Teisserenc, 2005; Millis & Just, 1994; Sanders & Noordman, 2000).

Third, in this study we focused on the effects of two types of connectives on text comprehension: additive and causal connectives, in contrast to Land (2009), who used different types of coherence markers and lexical signals. At the same time, we do not know whether we can generalize the effects of these type connectives to other types of connectives, such as contrastive connectives. Adversative connectives often seem to be a prerequisite for constructing the contrastive relation, because without them, the relation cannot be established in the first place. Murray (1997) showed that appropriate adversative connectives have a greater impact on on-line processing than additive or causal connectives do. Cain and Nash (2011) also show that different types of connectives have a different impact on children's comprehension of sentence pairs (see Millis, Golding, & Barker, 1995 for similar findings for adults). Hence, further research into the effects of the various markers on text processing and comprehension is imperative.

Fourth, in this study we used history texts. History texts are of the expository type in that they convey new information, but they have a structure that is often narrative in form (Perfetti et al., 1995). Further research is needed to determine the effects of coherence markers in other genres. For example, students in secondary education often read school texts with a narrative form. Narratives usually present reoccurring topics, such as friendship and love, in a specific context involving particular characters, settings, and times (Best et al., 2005). Therefore, readers can draw upon their prior background knowledge to facilitate a coherent interpretation of such texts (Anderson, 2000). Can we generalize the concluded effectiveness of connectives on students' text comprehension to narratives? Or do readers benefit more from the presence of connectives in abstract, informative texts, such as economy, biology and geography texts. Such school texts describe complex components in systems, functions of the components, relationships among components, and logical justifications of claims. Much of this content is abstract and far removed from everyday experiences (Graesser et al., 2003). It is possible that coherence markers play a more facilitative role in these abstract texts.

5. Conclusion

This study shows the clear result that for pre-vocational students connectives lead to better local and global comprehension. Pre-university students also benefit from connectives when establishing local coherence. Even though the effects on global comprehension measures are less clear for pre-university students, it becomes increasingly clear that connectives are an important factor in creating comprehensible texts for readers in educational contexts. While many efforts focus on upgrading reading skills by means of interventions to improve reading achievement in school (cf. Broer, Aarnoutse, Kieviet, & Van Leeuwe, 2002), paying more attention to the design of comprehensible, cohesive texts is crucial as well.

References

Anderson, J.R. (2000). *Cognitive psychology and its implications* (5th ed.). New York: Worth Publishers.

Beck, I.L., McKeown, M.G., Sinatra, G.M., & Loxterman, J.A. (1991). Revising social studies text from a text-processing perspective: Evidence of improved comprehensibility. *Reading Research Quarterly*, 26, 251–275. DOI: 10.2307/747763

Ben-Anath, D. (2005). The role of connectives in text comprehension. *Working Papers in TESOL and Applied Linguistics*, 5(2), 1–27.

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

Best, R.M., Rowe, M., Ozuru, Y., & McNamara, D.S. (2005). Deep-level comprehension of science texts: The role of the reader and the text. *Topics in Language Disorders*, 25(1), 65–83. DOI: 10.1097/00011363-200501000-00007

Bohn-Gettler, C.M., Rapp, D.N., Van den Broek, P.W., Kendeou, P., & White, M.J. (2011). Adults' and children's monitoring of story events in the service of comprehension.*Memory and Cognition*, 39, 992–1011. DOI: 10.3758/s13421-011-0085-0

Bowen, B.A. (1999). Four puzzles in adult literacy: Reflections on the national adult literacy survey. *Journal of Adolescent and Adult Literacy*, 42, 314–323.

Britton, B.K. (1994). Understanding expository text. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 641–674). San Diego: Academic Press.

Britton, B.K., & Gülgöz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology*, 83, 329–345. DOI: 10.1037/0022-0663.83.3.329

Broer, N.A., Aarnoutse, C.A.J., Kieviet, F.K., & Van Leeuwe, J.F.J. (2002). The effect of instructing the structural aspect of texts. *Educational Studies*, 28(3), 213–238. DOI: 10.1080/0305569022000003681

Cain, K., & Nash, H.M. (2011). The influence of connectives on young readers' processing and comprehension of text. *Journal of Educational Psychology*, 103(2), 429–441. DOI: 10.1037/a0022824

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability and component skills. *Journal of Educational Psychology*, 96(1), 31–42. DOI: 10.1037/0022-0663.96.1.31

Canestrelli, A., Mak, W.M., & Sanders, T.J.M. (2012). Causal connectives in discourse processing: How differences in subjectivity are reflected in eye-movements. *Language and Cognitive Processes*, 28(9), 1394–1413. DOI: 10.1080/01690965.2012.685885

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum Associates.

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. DOI: 10.1080/01638539809545019

Cozijn, R., Noordman, L.G.M., & Vonk, W. (2011). Propositional integration and world-knowledge inference: Processes in understanding because sentences. *Discourse Processes*, 48(7), 475–500. DOI: 10.1080/0163853X.2011.594421

Daneman, M. (1991). Individual differences in reading skills. In R. Barr, M.L. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research (Vol. 2*, pp. 512–538). White Plains, NY: Longman.

Degand, L., & Sanders, T.J.M. (2002). The impact of relational markers on expository text comprehension in L1 and L2. *Reading and Writing: An Interdisciplinary Journal*, 15, 739–757. DOI: 10.1023/A:1020932715838

Fletcher, C.R. (1994). Levels of representation in memory for discourse. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 589–607). New York, NY: Academic.

Fuchs, L.S., Fuchs, D., Hosp, M.K., & Jenkins, J.R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, 5(3), 239–256. DOI: 239-256, 10.1207/S1532799XSSR0503_3

Gernsbacher, M.A. (1990). *Language comprehension as a structure building*. Hillsdale, NJ: Erlbaum.

Gilabert, R., Martínez, G., & Vidal-Abarca, E. (2005). Some good texts are always better: Text revision to foster inferences of readers with high and low prior background knowledge. *Learning and Instruction*, 15, 45–68. DOI: 10.1016/j.learninstruc.2004.12.003

Goldman, S.R. (1997). Learning from text: Reflections on the past and suggestions for the future. *Discourse Processes*, 23(3), 357–398. DOI: 0.1080/01638539709544997

Graesser, A.C., & McNamara, D.S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 3, 371–398. DOI: 10.1111/j.1756-8765.2010.01081

Graesser, A.C., McNamara, D.S., & Louwerse, M.M. (2003). What do readers need to learn in order to process coherence relations in narrative and expository text? In A.P. Sweet, & C.E. Snow (Eds.), *Rethinking reading comprehension* (pp. 82–98). New York, NY: Guilford Press.

Graesser, A.C., McNamara, D.S., Louwerse, M.M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments and Computers*, 36, 193–202. DOI: 10.3758/BF03195564

Graesser, A.C., Millis, K.K., & Zwaan, R.A. (1997). Discourse comprehension. *Annual Review of Psychology*, 48, 163–189. DOI: 10.1146/annurev.psych.48.1.163

Haberlandt, K. (1982). Reader expectations in text comprehension. In J.F. Le Ny, & W. Kintsch (Eds.), *Language and language comprehension* (pp. 239–249). Amsterdam: North-Holland. DOI: 10.1016/S0166-4115(09)60055-8

Hacquebord, H.I. (2004). Taalproblemen en taalbehoeften in het voortgezet onderwijs. Leerlingen- en docentenvragenlijsten als instrumenten voor taalbeleid [Language problems and language needs in secondary education: Student and teacher questionnaires as instruments for language policy]. *Levende Talen Tijdschrift*, 5(2), 17–28.

Halliday, M.A.K., & Hasan, R. (1976). *Cohesion in English*. London, England: Longman.

Kamalski, J. (2007). *Coherence marking, comprehension and persuasion. On the processing and representation of discourse*. Utrecht: LOT. Retrieved from http://www.lotpublications.nl/publish/articles/002334/bookpart.pdf

Kamalski, J., Sanders, T., & Lentz, L. (2008). Coherence marking, prior knowledge and comprehension of informative and persuasive texts: Sorting things out. *Discourse Processes*, 45, 323–345 DOI: 10.1080/01638530802145486

Kamoen, N. (2012). *Positive versus negative: A cognitive perspective on wording effects for contrastive questions in attitude surveys*. Utrecht: LOT. Retrieved from http://igitur-archive.library.uu.nl/dissertations/2012-0605-200517/kamoen.pdf

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. DOI: 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. DOI: 10.1037/a0015956

Kintsch, W. (1992). A cognitive architecture for comprehension. In H.L. Pick, P. van den Broek, & D.C. Knill (Eds.), *The study of cognition: Conceptual and methodological issues* (pp. 143–164). Washington, DC: American Psychological Association. DOI: 10.1037/10564-006

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, 49, 292–303. DOI: 10.1037/0003-066X.49.4.294

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

Kintsch, W., & Van Dijk, T.A. (1978). Towards a model of text comprehension and production. *Psychological Review*, 85, 363–394. DOI: 10.1037/0033-295X.85.5.363

Koornneef, A.W., & Sanders, T.J.M. (2013). Establishing coherence relations in discourse: The influence of implicit causality and connectives on pronoun resolution. *Language and Cognitive Processes*, 28, 1169–1206. DOI: 10.1080/01690965.2012.699076

Land, J.F.H. (2009). *Zwakke lezers, sterke teksten? Effecten van tekst- en lezerskenmerken op het tekstbegrip en de tekstwaardering van vmbo-leerlingen* [Less-skilled readers, well-built texts? Effects of text and reader characteristics on text comprehension and text appreciation]. Delft: Eburon.

LeVasseur, V.M., Macaruso, P., Palumbo, L.C., & Shankweiler, D. (2006). Syntactically cued text facilitates oral reading fluency in developing readers. *Applied Psycholinguistics*, 27, 423–445. DOI: 10.1017.S0142716406060346

Linderholm, T., Everson, M.G., Van den Broek, P.W., Mischinski, M., Crittenden, A., & Samuels, J. (2000). Effects of causal text revisions on more- and less-skilled readers' comprehension of easy and difficult texts. *Cognition and Instruction*, 18(4), 525–556. DOI: 10.1207/S1532690XCI1804_4

Loman, N.L., & Mayer, R.E. (1983). Signaling techniques that increase the understandability of expository prose. *Journal of Educational Psychology*, 75(3), 402–412. DOI: 10.1037/0022-0663.75.3.402

Lorch, R.F., Jr., & Lorch, E.P. (1986). On-line processing of summary and importance signals in reading. *Discourse Processes*, 9, 489–496. DOI: 10.1080/01638538609544654

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Addison-Wesley: Reading Mass.

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

Maury, P., & Teisserence, A. (2005). The role of connectives in science text comprehension and memory. *Language and Cognitive Processes*, 20(3), 489–512. DOI: 10.1080/01690960444000151

McMaster, K.L., Van den Broek, P., Espin, C.A., White, M.J., Rapp, D.N., Kendeou, P., Bohn-Gettlere, C.M., 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. DOI: 10.1016/j.lindif.2011.11.017

McNamara, D.S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, 55(1), 51–62. DOI: 00008558-200103000-00004

McNamara, D.S., Kintsch, E., Songer, N.B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14(1), 1–43. DOI: 10.1207/s1532690x-ci1401_1

Myers, J.L., Shinjo, M., & Duffy, S. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language*, 26, 453–465. DOI: 10.1016/0749-596X(87)90101-X

Millis, K.K., Golding, J.M., & Barker, G. (1995). Causal connectives increase inference generation. *Discourse Processes*, 20, 29–49. DOI: 10.1080/01638539509544930

Millis, K.K., & Just, M.A. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language*, 33(1), 128–147. DOI: 10.1080/0163853X.2012.692655

Mulder, G., & Sanders, T.J.M. (2012). Causal coherence relations and levels of discourse representation. *Discourse Processes*, 49(6), 501–522. DOI: 10.1080/0163853X.2012.692655

Murray, J.D. (1997). Connectives and narrative text: The role of continuity. *Memory & Cognition*, 25(2), 227–236. DOI: 10.3758/BF03201114

National Center for Education Statistics. (2012). *The nation's Report Card: Reading 2011. National assessment of educational progress at grades 4 and 8*. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2011/2012457.pdf

Noordman, L.G.M., & Vonk, W. (1997). The different functions of a conjunction in constructing a representation of the discourse. In M. Fayol, & J. Costermans (Eds.), *Processing interclausal relationships in production and comprehension texts* (pp. 75–93). Mahwah, NJ: Erlbaum.

Oakhill, J.V. (1994). Individual differences in children's text comprehension. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 821–848). New York: Academic Press.

Oakhill, J.V., Cain, K., & Bryant, P.E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes*, 18, 443–468. DOI: 10.1080/01690960344000008

OECD. (2010). *PISA 2009 results: What students know and can do: Student performance in reading, mathematics and science (Vol. 1)*. Retrieved from http://dx.doi.org/10.1787/9789264091450-en

O'Reilly, T., & McNamara, D.S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes*, 43(2), 121–152. DOI: 10.1080/01638530709336895

Pander Maat, H., & Sanders, T. (2006). Connectives in text. In K. Brown (Ed.), *Encyclopedia of language and linguistics*, 3 (33–41). London: Elsevier.

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

Perfetti, C.A., Britt, M.A., & Georgi, M.C. (1995). *Text-based learning and reasoning: Studies in history*. Mahwah, NJ: Lawrence Erlbaum.

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

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park: Sage.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422. DOI: 10.1037/0033-2909.124.3.372

Sanders, T., & Canestrelli, A.R. (2012). The processing of pragmatic information in discourse. In H.J. Schmid (Ed.), *Cognitive pragmatics* (pp. 201–232). Berlin: Mouton de Gruyter.

Sanders, T.J.M., Land, J., & Mulder, G. (2007). Linguistic markers of coherence improve text comprehension in functional contexts. *Information Design Journal*, 15(3), 219–235. DOI: 10.1075/idj.15.3.04san

Sanders, T.J.M., & Noordman, L.G.M. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, 29(1), 37–60. DOI: 10.1207/S15326950dp2901_3

Sanders, T., & Spooren, W. (2007). Discourse and text structure. In D. Geeraerts, & J. Cuykens (Eds.), *Handbook of cognitive linguistics* (pp. 916–941). Oxford: Oxford University Press.

Sanders, T.J.M., Spooren, W., & Noordman, L. (1993). Coherence relations in a cognitive theory of discourse representation. *Cognitive Linguistics*, 4(2), 93–133. DOI: 10.1515/cogl.1993.4.2.93

Schram, D. (2002). Moeilijke tekst en moeilijke lezer? Over het lezen van een verhaal op het vmbo [Difficult text and difficult reader? About reading a story in pre-vocational secondary

education]. In A.M. Raukema, D. Schram, & C. Stalpers (Eds.), *Lezen en leesgedrag van adolescenten en jongvolwassenen* [Reading and reading behavior of adolescents and young adults] (pp. 105–118). Delft: Eburon.

Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.

Spyridakis, J.H., & Standal, T.C. (1987). Signals in expository prose: Effects on reading comprehension. *Reading Research Quarterly*, 22, 285–298. DOI: 10.2307/747969

Van den Broek, P., Lynch, J.S., Naslund, J., Ievers-Landis, C.E., & Verduin, K. (2003). The development of comprehension of main ideas in narratives: Evidence from the selection of titles. *Journal of Educational Psychology*, 95(4), 707–718. DOI: 10.1037/0022-0663.95.4.707

Zwaan, R.A., & Radvansky, G.A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162–185. DOI: 00006823-199803000-00003

Zwaan, R.A., & Singer, M. (2003). Text comprehension. In A.C. Graesser, M.A. Gernsbacher, & S.R. Goldman (Eds.), *Handbook of discourse processes* (pp. 83–121). Mahwah, NJ: Lawrence Erlbaum.

Appendix A. An example of an explicit version of an experimental text with a continuous layout

(This text is a translation of an original Dutch school text. Below, the connectives are underlined. We only included the connectives in the explicit version.)

Hygiene in the Middle Ages

In the Middle Ages, people lived in a very unhealthy way. Nobody cared about garbage and about unpleasant odors. There were no sewers, <u>and</u> there were very few toilets. At night, people used a chamber pot. They emptied this pot on the road. During the day, people defecated in a type of box with a hole. Such a box with a hole stood over a canal or a cesspool. The faeces fell into the cesspool, but there was no water in it. <u>Therefore</u>, a cesspool smelled worse than a canal. Around the fifteenth century, the toilets were more luxurious. The floor was strewn with fragrant herbs, <u>and</u> for toilet paper, people used stripes of linen.

A bath with hot water was only for rich people, <u>because</u> it was difficult to get hot water. They used expensive linen cloths on the floor. They <u>also</u> put expensive bath oil in the water. <u>As a result</u>, the water smelled nice during bathing. Workers hardly ever took a bath, <u>because</u> taking a bath was too expensive. In order to be able to afford taking one bath, a worker had to work a whole week.

In the Middle Ages, diseases spread very fast, <u>because</u> basements, kitchens and cupboards were full of rats. The rats were full of fleas, <u>and</u> the fleas carried terrible diseases. Just imagine the plague. The plague caused a lot of black bumps all over your body. <u>That's why</u> the plague was called the Black Death. The plague was active between 1347 and 1351. In Europe and Asia, about 25 million people died.

Appendix B. Measures of error per model

In each table, the best fitting model is indicated with an asterisk.

Table 1. True-false statements: measure of error per model

Predictor in Model	−2 log-likelihood	Δ χ^2	Δ df	*p*
1. Constant	4014.39		4	
2. Model 1 + Coherence	4014.18	2–1=.21	5	.65
3. Model 1 + Layout	4013.53	3–1=.86	5	.35
*4. Model 1 + Level	3986.66	4–1=27.76	6	< .001
5. Model 4 + Coherence x Layout	3985.45	5–4=1.21	9	.75
6. Model 4 + Coherence x Level	3985.97	6–4=.69	9	.88
7. Model 4 + Layout x Level	3983.15	7–4=3.51	9	.32
8. Model 4 + Coherence x Layout x Level	3982.07	8–4=4.59	15	.87

Table 2. Inference questions: measure of error per model

Predictor in Model	−2 log-likelihood	Δ χ^2	Δ df	*p*
1. Constant	4254.44		4	
2. Model 1 + Coherence	4246.37	2–1=8.07	5	.005
3. Model 2 + Layout	4246.34	3–2=.03	6	.86
*4. Model 2 + Level	4187.70	4–2=58.67	7	< .001
5. Model 4 + Coherence x Layout	4187.23	5–4=.47	9	.79
6. Model 4 + Coherence x Level	4184.37	6–4=3.33	9	.19
7. Model 4 + Layout x Level	4187.50	7–4=.20	10	.98
8. Model 4 + Coherence x Layout x Level	4181.17	8–4=6.53	15	.58

Table 3. Sorting task: measure of error per model

Predictor in Model	−2 log-likelihood	Δ χ^2	Δ df	*p*
1. Constant	4250.10		4	
2. Model 1 + Coherence	4248.82	2–1=1.28	5	.26
3. Model 1 + Layout	4250.08	3–1=.02	5	.89
4. Model 1 + Level	4196.75	4–1=53.35	6	< .001
5. Model 4 + Coherence x Layout	4191.60	5–4=5.15	9	.16
6. Model 4 + Coherence x Level	4192.58	6–4=4.17	9	.24
7. Model 4 + Layout x Level	4194.44	7–4=2.31	9	.51
*8. Model 4 + Coherence x Layout x Level	4178.97	8–4=17.58	15	.04