Task-Oriented Reading in Primary School. The Ability to Discriminate Relevant Content from Multiple Text of Fourth- and Sixth-Graders

Héctor García-Rodicio[1]**, Ruth Villalón**[1]**, Ángeles Melero**[1]**, Belén Izquierdo-Magaldi**[1]

[1] Department of Education, Universidad de Cantabria, Santander

Spain

Correspondence: Héctor García-Rodicio. Facultad de Educación. Avenida de los Castros s/n, código postal 39005, Santander, Cantabria. España. E-mail: hector.garciarodicio@unican.es

Héctor García-Rodicio et al.

Abstract

Introduction. Multiple-text task-oriented reading is using several texts to perform an assigned task. It involves two core processes, namely, comprehending text contents and discriminating the text contents that are relevant for the task. We explored these two processes in primary school students.

Method. We had 150 students from grades fourth (9-10 years old) and sixth (11-12 years old) use three texts to learn about an assigned question. After using the texts, they took a recall test, which required them to recall as much as possible about the question-relevant information, and a verification test, which required them to judge inferences from the text contents, both question-relevant and question-irrelevant, as true or false.

Results. The participants performed worse in recall than in verification. The sixth-graders outperformed the fourth-graders in recall and, more clearly, in verification.

Discussion and Conclusion. The results suggest, first, that students are better able to comprehend text contents than to discriminate task-relevant content and, second, that both the ability to comprehend text content and the ability to discriminate task-relevant content improve throughout the school years but the former may develop faster.

Keywords: Multiple-text; task-oriented reading; text comprehension; relevance processing; memory for text.

Resumen

Introducción. La lectura-orientada-a-metas con textos multiples implica usar varios textos para resolver una tarea que nos han asignado. Engloba dos procesos nucleares: comprender los contenidos del texto y discriminar si los distintos contenidos son o no relevantes para la tarea. Exploramos estos dos procesos en alumnado de educación primaria.

Método. Solicitamos a 150 alumnos de cuarto curso (9-10 años de edad) y sexto curso (11-12 años de edad) que emplearan tres textos para aprender acerca de una cuestión. Tras usar los textos, les pasamos una prueba de recuerdo, en la cual debían recordar todo lo posible sobre la información relevante para la cuestión, y una prueba de verificación, en la cual debían juzgar inferencias generadas a partir de los contenidos del texto, tanto los relevantes como los irrelevantes, como verdaderos o falsas.

Resultados. Los participantes rindieron peor en recuerdo que en verificación. Los de sexto superaron a los de cuarto en recuerdo y, de forma aún más pronunciada, en verificación.

Discusión y Conclusiones. Los resultados sugieren, primero, que los alumnos son más capaces de comprender los contenidos de un texto que de discriminar los que son relevantes para una tarea dada y, segundo, que tanto la capacidad de comprender textos como la de discriminar contenido relevante para una tarea dada mejoran a lo largo de los años escolares, si bien la primera parece evolucionar más rápido que la otra.

Palabras clave: Textos multiples; lectura orientada-a-metas; comprensión de textos; discriminar contenido relevante; recuerdo.

Introduction

Both in academic and personal contexts, we usually read texts to answer questions (Grasser & Lehman, 2011) and, because very often the questions that motivate reading are complex, it is usually necessary to use several texts to generate the answer (McCrudden, 2018). This form of reading can then be called multiple-text task-oriented reading (Anmarkrud, McCrudden, Bråten, & Strømsø, 2013). This form of reading involves two core processes. Firstly, as any activity including text, it requires reading comprehension, which consists of converting words and phrases into mental representations of the situations described in the text (McNamara & Magliano, 2009). Moreover, since there is a task to accomplish, the information in the texts can be either relevant or irrelevant for the task goal; accordingly, an additional core process is discriminating the information that is indeed relevant for the task (McCrudden, 2018). So far, research on multiple-text task-oriented reading in primary school students is scarce (Kiili, Bråten, Kullberg, & Leppanen, 2020) and, moreover, none of the few studies conducted have utilized separate measures for the two core processes involved. Therefore, it is unclear how the ability to perform those key processes develops across grade levels. In the present experiment we examined how well the ability to select task-relevant information from multiple text is developed at fourth- (9-10 years old) and sixth-grade (11-12 years old), as compared with basic text-comprehension skill.

Multiple-text task-oriented reading in primary school students

Multiple-text task-oriented reading is using two or more texts to complete an assigned task or a self-selected reading goal (Anmarkrud et al., 2013). To our knowledge, only five studies have examined multiple-text task-oriented reading in primary school students (Florit, Cain, & Mason, 2020; Goldman, Lawless, & Manning, 2013; Kiili et al., 2020; Paul, Cerdán, Rouet, & Stadtler, 2018; Salmerón, Sampietro, & Delgado, 2020). Florit et al. (2020) asked fourth-graders to read three texts for the goal of writing an essay on a controversial topic, "is chocolate good for children or not?" and "are videogames useful or not?" Kiili et al. (2020) asked a group of sixth-graders to read four online resources to write an essay on the controversy "should the school purchase an energy drink vending machine?" In a similar way, Paul et al. (2018) had fourth-graders use four documents to write an essay on whether muesli snacks are healthy or not. Another similar study was conducted by Salmerón et al. (2020). They asked fourth-, fifth-, and sixth-graders to decide what form of water is better, bottled water or tap water, by using three resources. Finally, in the study by Goldman et al. (2013),

involving students from grades fifth to seventh, the participants were asked to discover "why the civil rights movement succeeded in the US in 1955-1965 but not before" by using three texts.

In all the cases the students' responses were evaluated globally. That is to say, they were evaluated in terms of how well they responded to the assigned question. Thus, when the assigned question was controversial (Florit et al., 2020; Kiili et al., 2020; Paul et al., 2018; Salmerón et al., 2020), the researchers evaluated the argument structure of the students' essays, that is, the stance they took on the controversy and the reasons justifying it. When the assigned question required providing a complete explanation of a complex phenomenon (Goldman et al., 2013), the researchers evaluated the causal structure of the students' responses and the degree in which they transformed (e.g., repetition, paraphrase, elaboration) the information in the texts. This kind of assessment, therefore, did not explore specific processes separately.

Processes in multiple-text task-oriented reading

Multiple-text task-oriented reading involves two core processes. First of all, as any other activity including text, multiple-text task-oriented reading requires comprehending the different text segments. Comprehending text segments consists of translating words and sentences into propositions, which are mental structures that capture the gist or underlying meaning of the sentences, and, eventually, into mental models representing the situations described in the text (McNamara & Magliano, 2009). This process of translation, in turn, includes several subprocesses, ranging from word recognition to inference-making.

Moreover, because task-oriented reading by definition implies there is a reading goal to accomplish, the different passages and texts can vary on how useful they are for the task, thus requiring readers to identify task-relevant information (McCrudden, 2018). Another core process, then, is to assess the information to determine whether or not is useful for the task at hand. For instance, in the study by Kiili et al. (2020) the participants had to decide whether it is appropriate or not to install an energy drink vending machine in the school. The texts provided arguments either for or against that decision. The texts also provided some factual details, such as the ingredients of energy drinks. The participants had to be able to discriminate that the arguments, but not the factual details, were useful to make their decision and do the task.

Furthermore, depending on the specific reading task, other processes can be distinguished. These processes, then, are not essential but conditional on the specific kind of task. For instance, if the task requires developing an argument on a controversy and the texts provide conflicting views, it is crucial to examine the source of the text, by critically evaluating the intention and the expertise of the source, in order to identify reliable arguments on the controversy (Braasch, Rouet, Vibert, & Britt, 2012).

As already mentioned, prior studies on multiple-text task-oriented reading in primary school students did not measured the two core processes separately. As a consequence, one limitation is that if relevant information was missing from the essays the students generated, it remains unclear if it was due to a failure in discriminating what information was relevant for the task or to a more basic failure in comprehending what the different text passages said in the first place. In other words, the design of the studies did not make it possible to explore whether the process of assigning relevance to a text content builds on comprehending the text content in the first place.

Discriminating task-relevant information from multiple text

Two models explain how exactly readers identify and process relevant content from multiple text. Based on the Goal-focusing model by McCrudden and colleagues (McCrudden, Magliano, & Schraw, 2011; McCrudden & Schraw, 2007), relevance processing takes several steps. The first thing readers have to do is to understand the task instruction to generate the reading goal. Based on the instruction itself and personal intentions, readers generate the reading goal, which provides criteria (or standards of relevance) for determining what information is relevant. Then readers have to judge the different text segments to determine whether they are relevant for the goal. As mentioned, the reading goal provides criteria for discriminating what information is relevant, therefore, determining relevance involves invoking such criteria and evaluating whether the information at hand meets the criteria. After determining relevance, readers allocate more processing to the relevant segments and this results in better learning of the relevant segments (Kaakinen & Hyönä, 2011; McCrudden & Schraw, 2007).

The MD-TRACE model by Rouet and Britt (2011) goes into greater detail about the process of discriminating task-relevant content. In order to assess the relevance of a text segment, the MD-TRACE model assumes that readers follow different sequences, depending on the task. When the task is highly specific, such as locating a particular fact, relevance assess-

ment is straightforward: readers skip most of the information until they reach the passage of interest and, then, they engage in deeper processing. When the task is more complex and not so specific, relevant information is not evident, requiring readers to read and comprehend the different text segments before they can make a decision about their relevance.

Moreover, the MD-TRACE assumes that the process of discriminating relevant content is constrained by two individual (or internal) factors. Firstly, it is constrained by memory skill: in order to decide whether a text segment is relevant for the task, readers have to invoke the reading goal and judge whether what the text segment says is consistent with the goal, which burdens readers' memory capacity. Secondly, the process is constrained by self-regulation skill. Once they have judged the relevance of a segment or a passage, skilled readers flexibly adapt their reading pace shifting from rapid scanning, for irrelevant passages, to slower, in-depth processing, for relevant passages. Self-regulation skill of the reader determines how flexibly he or she can make these adaptations.

An implication from the assumptions of the MD-TRACE model is that the process of discriminating relevant content is a harder, more sophisticated process than comprehending text contents. The implication is based on two assumptions. Firstly, in complex reading tasks, readers have to comprehend text contents before having the opportunity to judge their relevance, so relevance processing is a process that builds on comprehending. Secondly, in order to judge the relevance of a text segment, readers have to confront the segment with the reading goal, which relies on memory skill, and then have to allocate extra processing on the segment if it turns out to be relevant, which relies on self-regulation skill. Accordingly, it seems reasonable to expect that readers acquire and master comprehension before doing so with relevance processing. It is necessary to use separate measures for the two core processes in order to test this possibility, which is what we did here.

The acquisition of multiple-text use skills

Longitudinal studies have examined how students acquire single-text reading comprehension. Based on their findings, there seems to be a steady evolution in reading comprehension and its different underlying subskills, such as word reading accuracy, vocabulary, and verbal working memory. For instance, Oakhill and Cain (2012) followed a sample of students between second- and fifth-grade, observing that they did better in all the subskill-measures when they were in grade fifth than when they were in grades third and second.

Regarding the development of the skills involved in multiple-text task-oriented reading throughout the primary school years, to our knowledge, there are not longitudinal studies to date. Moreover, the studies conducted so far did not use specific measures for comprehension and relevance-processing skills. However, some tentative conclusions can be drawn from the available studies we reviewed above, even though the methods were not identical between the studies and the measures utilized did not assess specific processes. There seems to be a trend showing that sixth-graders perform better than fourth-graders in multiple-text task-oriented reading. For instance, in the study by Kiili et al. (2020), sixth-graders were able to present a clear stance on the controversy and provided two or three reasons on average to support their claims. Differently, a clear majority of the fourth-graders who participated in the study by Florit et al. (2020) failed to even identify the two perspectives of the controversy. In a similar way, the study by Salmerón et al. (2020), in which the participants were fourth-, fifth-, and sixth-graders, revealed a significant effect of grade on the quality of the essays, with the older students performing better than the younger ones.

Furthermore, there is one study on single-text task-oriented reading suggesting that fourth graders may have difficulties in discriminating relevant content given an assigned reading goal. The study utilized a narrative text, which was provided to students from grade fourth and seventh, among others (van den Broek, Tzeng, Risden, Trabasso, & Basche, 2001). In one condition, the text was provided with a reading goal in the form of adjunct questions the students had to answer by using the text; in the control condition, the text was provided alone. Although seventh-graders recalled the question-relevant information better when the questions were present than when they were absent, which means they were able to read the text for an assigned reading goal, the presence of questions for the fourth-graders not only did not benefit recall but also hindered it. In other words, they could not adapt their processing of the text to the reading goal, by identifying task-relevant passages and allocating extra processing on them.

Present study and research questions

Few studies have explored multiple-text task-oriented reading in primary school students and, most importantly, the measures they utilized evaluated students' performance globally but did not consider the ability to comprehend text content and the ability to discriminate task-relevant content separately. Accordingly, we wanted to examine how much the ability to discriminate task-relevant information in multiple text is developed at different

grade levels, as compared with the ability to comprehend text. To do so, we had fourth- (9-10 years old) and sixth-graders (11-12 years old) read three texts about an exotic country for an assigned question and take two tests after. The participants were asked to read the texts for the goal of discovering the problems that the people in the country are facing. The texts presented both question-relevant information (information about the problems in the country, such as the crash in automobile industry as a result of the strong development of the public transport) and question-irrelevant information (information about whatever other aspects of the country, such as the economic growth of the northern region thanks to the establishment of a mine). After reading the texts the participants took a recall test and a verification test. The recall test asked them to recall as much as possible about the question-relevant information, namely, the problems in the exotic country. The verification test required the participants to judge a list of statements about the exotic country as true or false; the statements were inferences from the text contents that had to do with both the question-relevant and the question-irrelevant information. Therefore, the recall test was a measure of the ability to discriminate task-relevant information whereas the verification test was a measure of the ability to comprehend text contents.

The design enabled us to explore the following research questions. The main research question was: *is discriminating relevant content more difficult than comprehending text contents?* If so, then the participants should perform better in the verification test, which required only to understand what the text segments said, than in the recall test, which required to determine whether or not the different text segments were relevant for the goal and to process them accordingly. Based on the MD-TRACE model we expected the participants to perform worse in the recall test, as compared with the verification test.

Another question was: *do the two main component skills of multiple-text task-oriented reading improve throughout the school years?* If so, the participants from sixth grade should outperform fourth-graders in recall and verification. That would mean that sixth-graders have developed relevance processing and comprehension more than fourth-graders. Based on the research on single-text and multiple-text comprehension, we expected sixth-graders to perform better, as compared with fourth-graders.

Finally, another question was: *do the main component skills develop at different rates?* If so, then the gap between grade levels should be different depending on the measure consid-

ered. In other words, if comprehension and relevance processing do not follow parallel developmental rates, then there should be an interaction between test (verification, recall) and grade (fourth, sixth). More accurately, if comprehension improved faster than relevance processing, then sixth-graders should exhibit a greater superiority over fourth-graders in verification than in recall. Alternatively, if relevance processing improved faster once comprehension is sufficiently mastered, then the superiority of sixth-graders over fourth-graders should be greater in recall than in verification. Based on the theoretical premises we have assumed, we expected the first possibility to happen. Specifically, given the difficulty in mastering the ability to process relevance, based on its demands on memory and self-regulation, we expected it to improve slower than the comprehension skill. Accordingly, sixth-graders should exhibit a greater superiority over fourth-graders in verification, as compared with recall.

Method

Participants

One-hundred and fifty students from fourth- and sixth-grade participated in the study. About forty percent of the participants were fourth-graders. Forty-seven percent of the participants were female. All the participants learned Spanish as their first language. The participants were recruited from three primary schools in a northern region of Spain. They all were public schools. The sample was homogeneous with respect to socioeconomic status (i.e., middle class). Their participation required parental and participant consent. All data were treated confidentially. The research ethics committee of our University ensured that the rights of the participants were protected.

Design

The participants were asked to use the texts to solve a question. They were provided with three texts about a fictitious country called Kabuti and the following instructions: "You are going to read some texts about Kabuti, a remote country in the Pacific Ocean. I am sure you never heard about this country but today you will learn how it is. I am going to explain what your task is. As you may know, some countries are widely known for their achievements, such as sports, technology, or tourist resorts; other countries, however, are known for their problems: earthquakes, war, hunger… Your task is to read the texts to discover if Kabuti is one of those countries that face trouble. Specifically, I want you to discover: What are the problems with the new public transport system? What are the problems with the new hiking

trail on the cliffs? What are the problems with the new large surface mine?" The instructions were presented in a sheet of paper, separated from the texts. After reading the texts the participants took two tests, the recall test and the verification test. Performances in the recall and the verification tests were the dependent variables.

In order to ensure that the instructions were understandable we carried out another little pilot study. We recruited 12 fourth-graders and provided them with the texts and with the instructions. After reading the instructions, we asked them to explain orally what their task was. We tested them individually. All the participants were able to identify the task (i.e., "to learn about the problems in Kabuti").

Materials

Texts

The participants were provided with three texts. They all were expository texts introducing Kabuti, a fictitious island country in the Pacific Ocean (see Table 1 for an excerpt). The topic of the texts was fictitious to limit the influence of prior knowledge but based on real-world events. Specifically, it was created by combining characteristics of actual island countries, such as Jamaica, Fiji, or Ireland. One text was entitled "An advanced country" and was 229 words long. Another text was entitled "Movements in the economy" and was 270 words long. The other text was entitled "Wild nature" and was 252 words long. With regard to their difficulty, Szrigriszt-Pazos indexes ranged from 65.8 (Crawford index: 5.1) to 54.4 (Crawford index: 5.9), which corresponded with moderately easy and normal, respectively. To further interpret these values, they correspond to a text complexity measure of about 920L in the Lexile(R) MetaMetrics scale, which is equivalent to a fifth-grade (10-11 years old) text complexity.

In conjunction, the texts described three recent events in Kabuti (the development of public transport, the construction of a hiking trail on the cliffs, the establishment of a large surface mine) and their consequences. The consequences had to do with the economy, the environment, and the quality of life of the Kabuti people. There were good consequences (e.g., reductions in traffic congestion, promotion of nature tourism, economic growth in underdeveloped regions of the island) and bad consequences or problems (e.g., recession in automobile industry, environmental impacts, contamination of soil and water). In total, there were 36 concepts, which were proportionally distributed across the three texts. A third of the

36 concepts had to do with problems in Kabuti and, therefore, were relevant for the reading goal. The other 24 concepts were irrelevant.

Intentionally, all the texts had the same genre: they all had a textbook appearance. By doing this, we eliminated the need to execute sourcing, since all the sources were equally trustworthy.

Table 1. *An Excerpt from one of the Texts*

Since ancient times, the inhabitants of Kabuti have used the *troveló*, a very simple but robust three-wheeled motorcycle, to move from one place to another every day. Two years ago, the president of the country carried out a total reform of the public transport system and the Kabutians no longer use *trovelós* but the new buses, trams and funiculars.

Now with hardly any trovelós circulating, the country's highways are much less clogged and it is no longer necessary to spend an hour to travel the main avenue of the capital. Besides, the ears of the Kabutians are much healthier, since they no longer have to endure the terrible sound of the *troveló*'s small but forced engine.

We conducted a pilot study to determine the relevance of the question-related concepts. We asked 12 college students to read the texts in order to learn about the problems in Kabuti. We provided them with the texts and the following instructions: "The texts introduce some recent events in Kabuti. Read them so that you understand the problems that those events are causing. What problems are the Kabuti people dealing with?" After they read the texts we provided the students with a list of sentences taken from the texts. The sentences described either the recent events in Kabuti, the good consequences of the events, or the problematic consequences. The students had to rate how relevant they were to the instruction, using a 10-point scale. An ANOVA showed that there were differences between the sentences in relevance, $F(2, 22) = 213.91$, $p < .001$. Posthoc pairwise comparisons showed that the sentences describing the problems received higher ratings than the sentences describing the events ($p < .001$) and those describing the good consequences ($p < .001$).

Moreover, in order to ensure that the topic was credible, we asked the same 12 students to read the texts and to estimate how believable the topic was, using a 10-point scale. The rating mean score was 8.00 ($SD = 0.83$). Therefore, we considered the topic credible.

Recall task

After reading the texts the participants took a recall test. They were provided with a sheet of paper, which included the following instruction: "Now I want you to write down as much as possible about the problems in Kabuti that were presented in the texts you have just read. It is important to write down every bit of this information you can remember." The texts were not available for the students during the test, since we were assessing the students' ability to recall relevant information. This is the usual procedure in studies using focused-reading instructions (e.g., McCrudden, Magliano, & Schraw, 2010). Based on the Goal-focusing model, relevant information is processed more attentively, which then results in better memory for this information (Kaakinen & Hyönä, 2011).

The recall protocols were divided into sentences. The sentences were then classified as relevant or non-relevant. The relevant sentences corresponded to the text passages that described the problems in Kabuti. The non-relevant sentences corresponded to the text passages describing the recent events in Kabuti or the good consequences of the events. Two independent raters scored a sample of 20 protocols (10% of the protocols), obtaining a Cohen's Kappa of 96%. The disagreements were solved through discussion until reaching consensus. This produced some amendments in the scoring procedure. One rater scored the rest of the protocols using the updated guidelines. In order to evaluate the recall protocols, we calculated one index: the number of relevant concepts included in the recall. Each relevant concept correctly recalled was awarded with two points and was awarded one point when was only partially recalled. The maximum score was 24. For instance, a participant who included four question-relevant concepts completely and two question-relevant concepts partially in her recall protocol would obtain a performance of 10 in this measure. It is important to note that written composition quality was not considered in the evaluation: it was the ability to recall the target information, not the writing ability, what mattered here.

Verification task

After taking the recall test the participants were asked to take the verification test. They were provided with a sheet of paper containing 24 statements. Twelve statements had to

do with the problems in Kabuti (e.g., "The sector of manufacturers and sellers of trovelós is growing tremendously") whereas the other 12 concerned the recent events in Kabuti and the good consequences of these events (e.g., "The troveló is a new and modern bus that more and more people use in Kabuti"). The statements were either true or false and the task for the participants was to judge each statement as true or false. The statements were inferences derived from the text contents, so participants had to comprehend the different text contents in order to respond. The texts were not available for the students during the test. Each correct judgment yielded one point, with a maximum score of 24. In order to control the guessing factor, we applied a penalty for wrong answers. Specifically, we calculated the corrected verification score by this formula: number of correct judgements – number of wrong judgements / number of possible answers – 1. For instance, a participant who made 18 correct and 6 wrong judgements would obtain a performance of 12 in this measure. Cronbach´s alpha for this test was .72. This reliability level it is modest indeed, but there are two reasons why it can be considered sufficient. First, the instrument is not intended to measure a general skill, such as intelligence or a personality trait, it is an ad hoc instrument instead. In scale development a minimum of .90 is required whereas in exploratory research .70 is considered appropriate (Lance, Butt, & Michels, 2006). Second, the instrument measured how well the students comprehended the different subtopics covered by the texts (public transport, mining activity, tourism, etc.). Although these subtopics have the same intrinsic complexity and, thus, should all be comprehended with the same ease, a particular participant could experience more difficulty with a particular subtopic for whatever reason. A handful of instances of this phenomenon could have reduced a bit the internal consistency of the measure.

Procedure

We contacted the head teachers of the schools and briefly described them how the task for the students would be and pointed out the ethical issues, such as informed consent, anonymity, and confidentiality. Then we got parental consent. We sent the parents a document describing the experiment and the ethical issues and asked them to sign it if they wanted their children to participate. Parental consent implied that the student wanted to participate in the first place. Approximately 95% of the students/parents agreed to participate. Those students who did not participate in the study did a similar activity, designed by their teachers, during the experiment.

The participants were tested in groups in their classrooms. First, the participants received a general instruction, indicating that they were going to do some work with text. Second, they were provided with the materials. The materials consisted of four sheets of paper. The first page contained the instructions for the task. The texts were presented separately in the remaining three pages. Third, the participants read the texts. Fourth, the participants took the recall test. Fifth, they took the verification test. The texts were not available during these tests for the reasons we explained above. The entire experiment was completed in approximately 40 minutes.

Results

We conducted a mixed-design ANOVA with test (verification, recall) as the within-subjects factor and grade (fourth, sixth) as the between-subjects factor. Alpha was set at .05. We checked that the assumptions for ANOVA were met. Means and standard deviations for the dependent variables appear in Table 2.

Table 2. *Means and Standard Deviations in the Dependent Measures*

	Fourth grade $N = 64$			Sixth grade $N = 86$			Total $N = 150$		
	M	*SD*	95%CI	*M*	*SD*	95%CI	*M*	*SD*	95%CI
Recall	21.75	18.76	[19.40, 24.10]	29.36	20.84	27.11, 31.61]	26.11	20.26	[24.46, 27.76]
Verification	52.25	27.45	[48.82, 55.68]	64.28	24.69	[61.62, 66.94]	59.15	26.49	[56.99, 61.31]

Note. Scores are shown as percentages.

Is discriminating relevant content more difficult than comprehending text contents?

The effect of test was significant, $F(1, 148) = 629.82$, $p < .001$. The participants performed better in the verification test than in the recall test.

Do the component skills of multiple-text task-oriented reading improve in sixth-grade, as compared with fourth-grade?

The effect of grade was significant, $F(1, 148) = 8.74$, $p < .005$, with the sixth-graders outperforming the fourth-graders. Independent-sample *t*-tests confirmed this pattern in verification, $t(148) = 2.81$, $p < .01$, and in recall, $t(148) = 2.31$, $p < .05$.

Do the main component skills develop at different rates?

There was a significant interaction between test and grade, $F(1, 148) = 5.22$, $p < .05$. More specifically, the superiority of sixth-graders over fourth-graders was greater in verification than in recall (see Figure 1). In fact, Cohen's *d* was medium in verification ($d = 0.54$) and small in recall ($d = 0.39$).
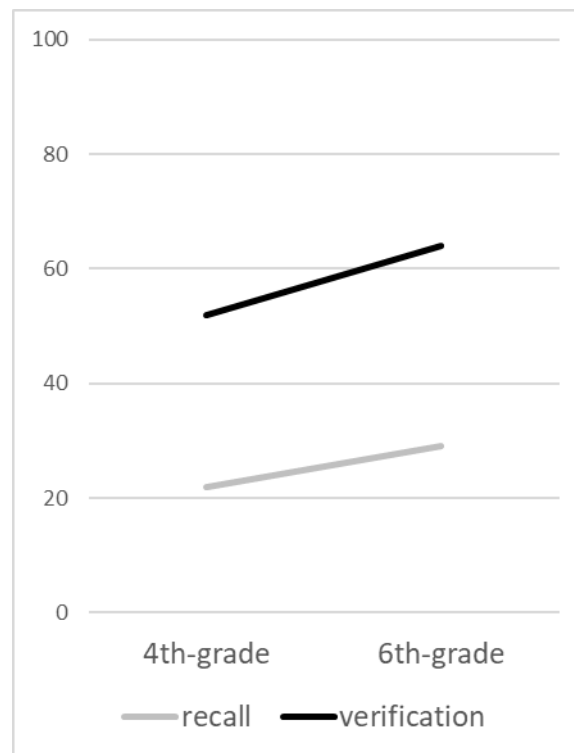

Figure 1. *Interaction between test (recall, verification) and grade (fourth, sixth)*

Discussion

Multiple-text task-oriented reading is using several texts for an assigned task (Anmarkrud et al., 2013). It involves two main processes, namely, comprehending text contents and discriminating the contents that are relevant for the reading goal. Based on the MD-

TRACE model (Rouet & Britt, 2011), it can be assumed that the process of discriminating relevant content is harder than the process of comprehending text content because the former builds on the latter and because the former requires memory skill (readers have to invoke the reading goal and decide whether the text segment at hand is or not consistent with that goal) and self-regulation skill (readers have to flexibly adapt their reading pace depending on the relevance of the passage, shifting from rapid scanning, when the passage is not relevant, to slower, in-depth processing, when the passage is relevant). Since multiple-text task-oriented reading has been scarcely studied in primary school students and using global instead of specific measures, it is unclear so far how well developed are comprehension and relevance-processing skills at different grade levels. In the present study we examined these two skills at grades fourth (9-10 years old) and sixth (11-12 years old). We had the students study a question by using three texts, which contained question-relevant and question-irrelevant information, and take a recall test (students recalled as much as possible about the question-relevant information) and a verification test (students judged sentences about the information in the texts, including question-relevant and question-irrelevant information, as true or false) after reading. The recall test enabled us to explore the ability of the students to discriminate question-relevant information from the texts; the verification test enabled us to explore the ability of the students to comprehend what the texts say.

The first research question was if discriminating relevant content from multiple text is indeed harder than comprehending text contents. The MD-TRACE model assumes that when the task is not highly specific, readers have to comprehend text contents in the first place to then having the opportunity to determine whether they are relevant for the goal. This means that discriminating relevant content builds on the process of comprehending text contents. Moreover, the process of discriminating relevant content relies on memory skills and on self-regulation skills. For these reasons, it seems reasonable to expect relevance processing to be more difficult than comprehension and, therefore, readers should acquire and master comprehension skill before doing so with relevance processing. Finally, there was some evidence that fourth-graders experience difficulties when having to read a narrative text for an assigned reading goal (van den Broek et al., 2001). For all these reasons, we expected a significant effect of test: the students would perform better in the verification test, which measured comprehension, than in the recall test, which measured relevance processing. The results confirmed this prediction and, thus, indicate that given its complexity, students are less able to

discriminate relevant content from multiple text than to comprehend text contents, at least at the grade levels studied here.

The second research question we explored was if the two main skills involved in multiple-text task-oriented reading improve in sixth-grade, as compared with fourth-grade. We expected this to happen based on prior findings and on theoretical bases. Firstly, although they did not use specific measures for particular processes, prior studies on multiple-text task-oriented reading found that older students (11-14 years old) perform better than younger students (9-11 years old) in these tasks (Florit et al., 2020; Goldman et al., 2013; Kiili et al., 2020; Paul et al., 2018; Salmerón et al., 2020). Secondly, since performance in multiple-text task-oriented reading relies on the development of several fundamental cognitive skills, such as vocabulary or memory, it could be safely expected that older students have developed comprehension and relevance-processing skills more, as compared with younger students. In fact, this is what we found: there was a significant effect of grade on performance and, more specifically, sixth-graders outperformed fourth-graders both in verification and recall.

Another research question was if relevance processing and comprehension follow different developmental rates. We found that here that there is a superiority of sixth-graders over fourth-graders both in verification and in recall. If the development of the skills underlying those performances followed different rates, then the superiority of sixth-graders should be different depending on the specific measure we are considering. We found an interaction between grade (fourth, sixth) and test (verification, recall) indeed. Specifically, the superiority of sixth-graders over fourth-graders was greater in verification than in recall. The pattern suggests that the ability to comprehend text contents improves faster throughout grade levels than the ability to discriminate task-relevant contents. The result is consistent with the theoretical premises we have assumed. Based on the MD-TRACE model, we have assumed that relevance processing draws on memory skill (i.e., readers have to invoke the task goal to decide whether the text segment at hand is or not relevant for the task and do this with all text segments) and on self-regulation skill (i.e., readers have to adapt their reading depth to the specific segment, skipping the irrelevant segments and slowing the pace on the relevant ones). A possible explanation for the interaction, therefore, is that given the strong dependence of relevance-processing skill on these fundamental skills, namely, memory skill and self-regulation skill, which have a lot of room for improvement even at sixth-grade (Dignath & Büttner,

2008), the ability to discriminate relevant content from multiple text still lacks the necessary basis to fully operate even at sixth-grade.

The present research has several implications. First of all, the present study extends prior research in two ways. We focused on primary school students, an age which has been scarcely studied so far (Kiili et al., 2020). Moreover, we utilized specific measures for the two basic processes involved in multiple-text use, namely, comprehension of text contents and discrimination of task-relevant content, and we considered two grade levels. Prior studies had not utilized separate measures for these processes (Florit et al., 2020; Goldman et al., 2013; Kiili et al., 2020; Paul et al., 2018; Salmerón et al., 2020).

Regarding the theoretical implications, the present study provides additional support for the MD-TRACE model (Rouet & Britt, 2011). The MD-TRACE model assumes that discriminating task-relevant content is a process difficult to carry out. Firstly, it builds on the process of comprehending text contents. Secondly, it relies on memory and self-regulation skill. The present results showed indeed that primary school students are significantly more able to comprehend text contents than to identify the content from multiple text that is relevant for the assigned task and process it accordingly.

A practical implication could be also derived from the present findings. We have seen that in fourth-grade and even in sixth- grade performance on measures requiring the discrimination of task-relevant content is significantly worse than performance on measures requiring the comprehension of text content. In other words, even in sixth-grade, primary school students find it difficult to use texts for an assigned goal. At the same time, goal-oriented reading is a crucial ability both in academic and personal contexts. For these reasons, it seems appropriate to provide students with support and/or instruction in task-oriented reading. There is evidence that even short-term interventions can be effective (e.g., Mason, Junyent, & Tornatora, 2014).

The present study has some limitations, which hopefully can suggest directions for future research. One limitation is that we did not take measures of memory capacity or self-regulation skill. The MD-TRACE model, which served as the theoretical basis for the study, assumes that these two fundamental skills constrain the ability of students to identify and process relevant content from multiple text. Given the gap we have found between the perfor-

mance in the comprehension measure and the performance in the relevance-processing measure, we have inferred this is due to such dependence. However, it would be good to further confirm this relationship by measuring memory and self-regulation skills directly. Another limitation is that we did not take online measures to see how exactly the participants processed the different text segments. As they were able to recall relevant content, it can be inferred that they processed those passages more in-depth (Kaakinen & Hyönä, 2011; McCrudden & Schraw, 2007). However, it would be good to further confirm this assumption by measuring segment-by-segment processing directly. Finally, it would have been good to get a higher level of internal consistency in the verification task, although the current .72 can be considered sufficient for the reasons we explained before. Future studies could address these issues. The present study could be considered, then, a first step in these directions.

References

Anmarkrud, Ø, McCrudden, M. T., Bråten, I., & Strømsø, H. I. (2013). Task-oriented reading of multiple documents: Online comprehension processes and offline products. *Instructional Science*, *41*, 873- 894. https://doi.org/10.1007/s11251-013-9263-8

Braasch, J. L. G., Rouet, J.-F., Vibert, N., & Britt, M. A. (2012). Readers' use of source information in text comprehension. *Memory & Cognition*, *40*(3), 450-465. https://doi.org/10.3758/s13421-011-0160-6

Florit, E., Cain, K., & Mason, L. (2019). Going beyond children's single-text comprehension: The role of fundamental and higher-level skills in 4th graders' multiple-document comprehension. *British Journal of Educational Psychology*, *90*, 449-472. https://doi.org/10.1111/bjep.12288

Goldman, S. R., Lawless, K., & Manning, F. (2013). Research and development of multiple source comprehension assessment. In M. A. Britt, S. R. Goldman, & J. F. Rouet (Eds.), *Reading from words to multiple texts* (pp. 160-179). New York: Routledge.

Graesser, A.C., & Lehman, B.A. (2011). Questions drive comprehension of text and multimedia. In M.T. McCrudden, J.P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 53-74). Information Age Publishing.

Kaakinen, J. K., & Hyönä, J. (2011). Online processing of and memory for perspective relevant and irrelevant text information. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 223-242). Information Age Publishing.

Kiili, C., Bråten, I., Kullberg, N., & Leppänen, P. H. T. (2020). Investigating elementary school students' textbased argumentation with multiple online information resources. *Computers and Education*, *147*, 103785. https://doi.org/10.1016/j.compedu.2019.103785

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). What did they really say? *Organizational Research Methods*, *9*, 202–220. ttps://doi.org/10.1177/1094428105284919

Mason, L., Junyent, AA., & Tornatora, MC. (2014). Epistemic evaluation and comprehension of web-source information on controversial science-related topics: Effects of a short-term instructional intervention. *Computers & Education*, *76*, 143-157. ttps://doi.org/10.1016/j.compedu.2014.03.016

McCrudden, M. T. (2018). Text relevance and multiple source use. In J. L. G. Braasch, I. Bråten, y M. T. McCrudden (Eds.), *Handbook of multiple source use* (160-179). Routledge.

McCrudden, M. T., Magliano, J.P., & Schraw, G. (2011). Relevance in text comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 1-17). Information Age Publishing.

McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review*, *19*(2), 113-139. ttps://doi.org/10.1007/s10648-006-9010-7

McNamara, D. S., & Magliano, J. P. (2009). Towards a comprehensive model of comprehension. In B. Ross (Ed), *The psychology of learning and motivation*, vol. 51, (pp. 297-284). Elsevier Science.

Oakhill, J.V., & Cain, K. (2011). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, *16*, 91-121. https://doi.org/10.1080/10888438.2010.529219

Paul, J., Cerdán, R., Rouet, J. F., & Stadtler, M. (2018). Exploring fourth graders' sourcing skills. *Journal for the Study of Education and Development*, *41*(3), 536-580. ttps://doi.org/10.1080/02103702.2018.1480458

Rouet, J., & Britt, M. A. (2012). Relevance processes in multiple document comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 19–52). Information Age.

Salmerón, L., Sampietro, A., & Delgado, P. (2020). Using Internet videos to learn about controversies: Evaluation and integration of multiple and multimodal documents by pri-

mary school students. *Computers & Education*, *148*, 103796. ttps://doi.org/10.1016/j.compedu.2019.103796

Van den Broek, P., Tzeng, Y., Risden, K., Trabasso, T., & Basche, P. (2001). Inferen-tial questioning: Effects on comprehension of narrative texts as a function of grade and timing. *Journal of Educational Psychology*, *93*, 521–529. https://doi.org/10.1037//0022-0663.93.3.521