Relevance processes in multiple document comprehension

Jean-François Rouet (1) and M. Anne Britt (2)

(1) CNRS and University of Poitiers
(2) Northern Illinois University

Chapter submitted as part of McCrudden, Magliano & Schraw's Edited volume on Relevance-based reading

Revised version

June 2010

Abstract
We introduce the MD-TRACE model (for Multiple-Document Task-based Relevance Assessment and Content Extraction), a descriptive model of the resources and the processes brought to bear in document-based activities. The MD-TRACE model considers the external resources that take part in functional reading activities (beyond text passages), as well as the cognitive resources that these activities call upon. The MD-TRACE model defines 5 core processes: the construction of a task model, the assessment of one's information needs, the selection, processing and integration of document information, the construction of a task product, and the assessment of product quality. These processes are helpful in identifying the many ways in which relevance processes take place during functional reading. We discuss several research questions that are derived from the revised model.

In today's information age, readers are faced with textual information that comes from very diverse sources, both in print and online. Printed and electronic texts also tend to play a role in a wide range of everyday situations. For instance, shopping online requires the consumer to read, comprehend, compare and contrast various written product descriptions. Similarly, people use more textual sources when informing their decisions regarding education, jobs opportunities, investment or health, to take just a few examples. These dramatic changes in the forms and uses of text challenge traditional models of reading comprehension and call for a generalized theory of multiple text processing (Britt, Perfetti, Sandak, & Rouet, 1999; Perfetti, Rouet & Britt, 1999; Rouet, 2006).

At the heart of this new perspective on text processing rests the notion of relevance, which refers to the extent to which information is consistent with the readers' needs and capacities (McCrudden & Schraw, 2007). The goal of this chapter is to examine the processes related to readers' assessment of information's relevance in situations in which readers make use of multiple texts to fulfill complex information needs. After a brief review of the contextual dimensions of text processing, we propose a new framework to describe the resources and processes involved in functional document-based activities. We highlight those phases and steps in the comprehension process in which relevance plays a critical part.

Reading as a contextalized activity

Research on text processing has initially focused on the impact of text structure on reading and comprehension processes (Cirilo & Foss, 1980; Kintsch, Kozminsky, Streby, McKoon & Keenan, 1975; Kintsch & van Dijk, 1978). Reading time and text recall were the two main indicators of processing and comprehension. These two measures were found to be strongly dependent on the contents and organization of the text. For instance, reading time increased with the number of semantic propositions in a passage (Kintsch et al., 1975) and also as a function of the status of propositions in the propositional hierarchy that makes up the literal meaning of the text (Cirilo & Foss, 1980; Guindon & Kintsch, 1984). Other textual features include the content of titles and initial mentions (Kozminsky, 1977; Kieras, 1980) or whether the order of information is consistent with pre- readers' existing schemata (Kintsch & Yarbrough, 1982). Kintsch's Construction-Integration model (Kintsch, 1988; 1998) provided a comprehensive account of the detailed memory mechanisms that account for those effects. Those well-established findings were obtained using tasks in which participants were asked to read for general comprehension purposes (e.g. to be prepared for a memory test, without further instruction as what they would have to do with the information). In authentic reading situations, however, readers often approach the text with more specific needs or purposes in mind, such as answering a question, making a decision, or writing an essay on a topic. Hence, there is a need for a new perspective on reading processes that accounts for different uses of text as a function of contextual variables.

Current discussions about the nature of reading emphasize the fact that contextual dimensions play a critical role in shaping the reader's strategies (Cerdan et al., this volume; McCrudden & Schraw, 2007; Rouet & Vidal-Abarca, 2002; Snow and the RAND reading study group, 2002; Vidal-Abarca, this volume). Snow et al. (2002) provided a detailed account of how individual, text and task dimensions interact to shape the act of reading. They noted that "Reading does not take place in a vacuum. It is done for a purpose, to achieve some end." (p. 15). They labeled "Activity" this pragmatic dimension of reading. Snow et al. (2002) further decomposed the Activity dimension into the purposes, processes, and consequences associated with the act of reading. They suggested that the broader sociocultural context in which reading takes places contributes to shaping each of the dimensions. Even though they focused on the classroom as a context, they acknowledged that context also plays a part in out-of-school reading experiences. The main point of their discussion is that, given a particular reader and a particular text, many different reading behaviors may be observed as a function of when, where and why the reading episode is taking place.

Developing a theory of reading as a contextualized activity requires a new approach to the role and importance of information in shaping the reader's understanding of the text.

McCrudden and Schraw (2007) established a distinction between text-based importance, on the one hand, and relevance (or task-based importance) on the other hand. They defined text-based importance as the degree to which a text segment includes information needed to understand the text (or structural importance), whereas relevance is the extent to which a segment contains information that is necessary to perform a certain task (e.g., to answer a question). They pointed out that in most situations, readers approach the text with more or less specific goals or objectives that may affect the relevance of textual information irrespective of its structural importance. When readers do possess those goals, they may switch from a "default" processing strategy based on textual importance to a relevance-based strategy in which they focus on text segments that are relevant to the task. Indeed, Schraw, Wade and Kardash (1993) demonstrated the impact of relevance instructions on memory for text. In a pilot study, they asked a group of participants to provide importance ratings for each segment of a narrative text describing a house. Two additional groups rated the relevance (or task-based importance) of each text segment from either a homebuyer or burglar perspective. In the main experiment, college students read the text from the perspective of a burglar or homebuyer. The main finding was that readers recalled perspective-relevant segments equally well, irrespective of their textual importance. The memory-based view of reading can account for this finding.

However, some task contexts lead readers to skim, skip, or ignore textual information. Rouet, Vidal-Abarca, Bert-Erboul and Millogo (2001) found that the reading time of paragraphs in a lengthy science text presented on a computer screen was deeply affected by the type of pre-questions college students were assigned prior to reading. Low-level questions prompted a "locate-and-memorize" strategy, where only segments that directly matched the question were considered, whereas higher-level questions prompted a "review-and-integrate" strategy where readers focused more broadly on sections of the text that contained information useful to reflect on the question.

What kind of cognitive processes can those shifts in strategies be attributed to? A memory-based view of text comprehension can account for variations in the saliency of information in a readers' representation of the text (see Lassonde et al., this volume). However, we believe that there is more to goal-based reading than memory activation. The reader's controlled allocation of attention and strategic decisions also play a critical role in shaping patterns of reading time and memory traces. Contextual constraints do not just act as a bias on an otherwise context-free set of procedures and cognitive processes that would be necessary and sufficient to define reading. Instead, expert readers possess a wide repertoire of flexible and differentiated reading procedures that they use depending on both the text they have at hand, and what they want to do with the text. In doing so, readers must be able to gauge the relevance of the information, a complex and multifaceted process that we discuss further in the following sections. We focus on the use of complex and multiple textual documents, which is both one of the most challenging forms of reading and offers a rich context for a close examination of relevance-based mechanisms. We first propose a general framework to represent the relevant resources and processes that take part in multiple document comprehension. We use this framework to reflect relevance-based mechanisms at various stages of the reading process, as well as on individual, text and task characteristics that may affect those processes.

Information resources and cognitive resources in multiple document use

In order to fully understand how relevance assessment mechanisms shape the reading experience, one needs a general framework to represent the key steps and processed involved in the contextual use of document information. Early studies of document search (e.g., Guthrie, 1988; Guthrie & Kirsch, 1987; Mosenthal & Kirsch, 1991) emphasized that reading an entire text for the general purpose of understanding its contents differs from other situations of document use. Guthrie (1988) found evidence that goal formation, category selection, content extraction and integration are key steps when searching non-continuous documents (e.g. flight schedules) in order to answer specific questions. Mosenthal (1996) identified three task variables that make document tasks more or less difficult: the type of information to be found (e.g., a concrete detail vs. an implicit rule), the

number of inferences required to match the goal with the documents contents, and the number and plausibility of distracters present in the materials (note that the latter factor is of ever-growing incidence in Web-based engine search).

Based on these and other works, Rouet (2006) proposed a general stepwise model called TRACE (for Task-based Relevance Assessment and Content Extraction) to account for the processes that take place during purposeful interactions with complex documents. Rouet (2006) initially aimed at representing a broad range of document use situations, from the location of a single detail within a text to the integration of information across multiple documents. In the present section, we present an updated version of the TRACE model called MD-TRACE, in order to elucidate the resources and processes involved in the functional use of texts. Compared to the earlier TRACE model, MD-TRACE accounts for a broader range of tasks and materials, including the integration of information across multiple documents.

The MD-TRACE model describes document use as an unfolding cycle of processing steps and decisions (see Figure 1).

insert Figure 1 about here

To the left of Figure 1 are the external resources that may be used during document-based activities: task specifications, the documents themselves along with search devices and organizers, and reader-generated products such as notes or essays. To the right of Figure 1 are the internal, or cognitive resources brought to bear by the reader. We distinguish between relatively permanent resources, such as prior knowledge, reading skills and self-regulation skills, and transitory resources generated as part of the activity, such as a task representation including reader-generated goals associated with the Task Model, and the output of document processing (i.e., a global representation of the various documents and the relationships among them) referred to as the Documents Model (see also Perfetti et al., 1999).

Finally, the center of Figure 1 shows the series of processing steps and decisions that underlie the search, selection, processing and use of document information. As further explained below, relevance decisions may take place during any of these steps.

To further illustrate the functioning of the MD-TRACE model, consider the simple hypothetical scenario described in Table 1.

insert Table 1 about here

In the scenario presented in Table 1, Lisa, a 14-year old 9th grader, is about to study a set of documents as part of a science project. An important question that arises from the scenario is to find out what Lisa's reading activity looks like during the 60 minutes she has to work on her assignment. What resources of her work environment does she make use of and how? What kind of declarative and procedural knowledge does she need in order to make sense of the information? When and how does Lisa apply relevance-based mechanisms? And, finally, what if Lisa was a doctoral student working on a chemistry dissertation on climate change, instead of a 9th grader? How would her reading pattern and outcomes differ from that of a 9th grade student and why? These are some of the questions the MD-TRACE model tries to address. In the subsections below we further develop the external and internal resources and the processing steps involved in the MD-TRACE model, pointing out their implications for a relevance-based theory of reading.

External resources

The external resources involved in using documents include three main types of information presented either prior to the student commencing the task or created and selected during the task completion (see Table 2). The first information resource is the Task specifications, i.e., the set of verbal and non-verbal cues that specify the reading task. Task specifications include the question or assignment itself as well as any instructions or directions for approaching the task. Explicit task specifications are present in many, if not all situations of

document use, and certainly in the vast majority of document reading activities that take place in schools or at the workplace.

In addition to verbal task requirements, most situations of document use include non-verbal constraints and opportunities such as time limitations, quality of the physical environment (e.g., reading in a quiet office with a large desk and a wide screen vs. reading from a mobile phone while standing in a crowded city bus), cost of accessing a resource (think for instance of one's decision to use a dictionary as a function of whether the dictionary is in the same or in a different room), and the availability of other, non-documentary potential sources of information such as peers, teachers or advisors.

Task specifications also include the level of performance expected and the potential consequences of the activity, for instance passing or failing a test, acquiring a better or a poorer product, getting tax or health insurance benefits and so forth.

In the example of Lisa's climate change work (Table 1), the task specifications come mostly from the teacher-generated question, which Lisa needs to read and understand correctly in order to set up appropriate study goals. But other elements of the context are also likely to shape Lisa's engagement with texts. For instance, Lisa is aware she has only one hour; and she knows that she will get input from her teammates before she has to hand out the result of her work. Finally she knows that the essay will be graded and what kind of grading criteria the teacher usually applies to this type of assignment.

insert Table 2 about here

The second and main type of external resource is the information resource per se, that is, the set of documents and access devices (indexes, search engines, text organizers) that are available in the environment. In printed environments, the information system may range from a single text to a full library; in networked electronic environments the information system may be potentially infinite, even though there will often be pragmatic constraints that limit the use of some available resources, such as language, fees, maintenance issues or breakdowns which may make a resource temporarily or permanently unavailable (think of the infamous http 404 error).

Especially in the case of multiple document comprehension, information about the source is an important aspect of the external resources available. Source information may be defined as the set of conditions that surround the production of the text, such as who the author is, when the text was written or updated, what readership the text is intended for and so forth (see Perfetti et al., 1999; Wineburg, 1994). Thus, we consider available documents to include information about the source as well as the content.

In the simple climate change example above, the three web pages found by Lisa come from different types of sources. One is from an online magazine, the second from a high school web site, and the third from a science research institute's Web site. These sources have different agendas and diverse readerships; their information may be more or less complete, recent and accurate; the language they use may be more or less accessible. These and other dimensions of sources are critical in shaping the reader's engagement and processing of the materials.

In many cases, document-based activities involve some form of collaboration with other human agents such as friends, peers, tutors or colleagues. Human agents may provide information that complements the documents, and they may also participate through dialogue, question answering or coaching. Document use may then involve spoken or written communication. Even though reading, collaborative work and help seeking have given rise to rather separate construct and study programs so far, recent theoretical proposals suggest that they may rely in part on a common set of underlying skills and mechanisms (Puustinen & Rouet, 2009). A discussion of relevance mechanisms as they apply to collaborative work or other forms of human interaction, however, is beyond the scope of the present chapter.

The third type of external resource is the set of task products generated by the reader as part of his or her use of the documents. These include highlighting, annotations, drafts, responses or essays. Not all reading experiences involve this type of production but readers

often find it useful to produce information while reading, especially when confronted with lengthy, multiple and/or difficult texts (Kiewra, DuBois, Christian et al., 1991; McGinley,1992; Newell & Winograd, 1989). Reader-generated information may then feedback into the reading process. For example, while reading her documents on climate change Lisa may decide to open a new word processing window, to copy and paste some of the information she has come across, to type in a few additional notes, and to refer back to these materials during the reading of subsequent documents.

In summary, even simple reading assignments may involve the use of multiple information resources. Texts are seldom read in isolation. Instead, they are most often read among other texts, in the context of specific assignments, and augmented with reader-generated notes and other traces.

Internal resources

The MD-TRACE model also considers the internal, or cognitive, resources brought to bear in document use. It is admittedly quite difficult to present a complete list of those resources as reading and comprehending text potentially requires most features of the human cognitive system. However, as shown in Table 3, document use rests on a set of key cognitive resources. It is convenient to distinguish permanent resources, that exist prior to the document use episode and presumably survive it, sometimes with some alterations, and non-permanent resources that are created as part of the activity and may vanish soon thereafter.

insert Table 3 about here

Permanent resources include general world knowledge, that is the declarative and procedural knowledge germane to the task at hand, as well as the memory, search, and reading skills needed to perform text search and reading comprehension (Cain, Oakhill, & Bryant, 2004; Cromley & Azevedo, 2007). The document user must possess at least some prior knowledge of the topic in order to be able to make sense of the assignment and the texts; he or she may also have some prior experience with the specific task context, for instance, what kind of information one may find in an index and how to use it (Rouet & Eme, 2002; Rouet & Coutelet, 2008); or how to formulate an effective query in a search engine. Finally, the actual processing of the documents requires memory skills, such as verbal and visuo-spatial capacity and executive control skills. Memory skills are related to students' reading comprehension abilities (e.g., Seigneuric, Ehrlich, Oakhill, & Yuill, 2000). Memory skills also include higher order abilities that play an important role in complex document-based tasks (Greene & Azevedo, 2007; Hacker, 1998; Winne & Hadwin, 1998).

When using multiple, naturalistic documents, readers also need to have some knowledge about the documents as artifacts (Wineburg, 1994). As mentioned earlier, documents are more than content: they include source information that may be critical in its own right but is especially important when interpreting and evaluating content. Thus, students need to be aware of authorship, genre, editorial purposes and time frames. These features of text are most often conveyed in the source of the text. Experienced readers of documents pay a great deal of attention to sources, as source features are sometimes critical in informing relevance decisions. We return to this in the next section.

Transitory resources are the memory representations that are created as part of the activity. These include the individual's representation of the task demands (see "task model" in Figure 1), including one's understanding of the questions or instructions that make up the assignment. The reader may also design an action plan, for instance a list of information sources one may plan to look up in order to fulfill the task. There is evidence that better and poorer document users differ in their ability to set up appropriate action plans. For instance, Raphael (1984) found that more able grade school students were better at understanding the demands of a question (i.e., whether the questions may be answered using one's prior knowledge, or using explicit information, or drawing inferences). Dreher and Guthrie (1990) found that efficient searchers spent relatively more time selecting information categories in a table of contents before engaging with the actual text; and Rouet and Le Bigot (2007) found

that more advanced college students were able to focus on problem-relevant sections of a hierarchical hypertext. Also of particular interest here is the finding by Ozuru, Best, Bell, Witherspoon and McNamara (2007) that college students' prior knowledge of the topic was less predictive of text comprehension performance when students could access the text at the time of answering questions, suggesting that students' ability to retrieve answers from a text is distinct from their ability to construct a memory representation from the text.

As the reader proceeds through the documents, he or she builds up a representation of each passage that is read (at least of those passages that are read in a systematic fashion -- see Figure 1, Step 3b); the "local" representation of a text's meaning corresponds to the notion of a mental model or situation model in text comprehension theories (Kintsch, 1998). However, in naturalistic document-based tasks readers often have to read and comprehend several texts which may contribute to, but do not exactly convey the solution or goal of the task. Therefore, readers may well decide to ignore or reject some of the contents they have read and comprehended. At this stage, the reader's ability to gauge the value of text information against his or her representation of the task specification is critical. As we will further discuss below, it is also a quite demanding process because of its reliance on memory processes and resources.

The MD-TRACE model assumes that, as readers accumulate information from multiple documents, they also build up a global representation of the documents' contents and sources that will serve as a basis for any output, response or solution the individual may have to deliver as part of the task assignment. Even though research into multiple document representation is still scarce, it has been the subject of increasing attention in recent years. Researchers have found that processing multiple documents involves specific integration mechanisms, especially when content information is not entirely consistent across documents. The Documents Model framework (Perfetti et al.,1999) describes some of the specific characteristics of reading and comprehending multiple documents. A fuller presentation of the Documents Model framework is beyond the scope of this article, but readers may refer to Rouet (2006), and to Britt and Rouet (in prep.) for an updated account as well as some instructional implications of the framework.

Processing steps and relevance assessment in multiple document comprehension

The MD-TRACE model defines a series of key steps that participate in document-based activities as shown in Figure 1. It should be pointed out that the decomposition of complex cognitive activities into a series of discrete steps is convenient but inaccurate, as steps may unfold in a less orderly and somewhat parallel way. Such decomposition is helpful, however, in understanding when and how relevance mechanisms may take place. We describe each step and we illustrate them through the Lisa scenario introduced in Table 1. We also consider the status of relevance and we discuss some of the challenges faced by students at each step.

Step 1. Create and update a task model

The TRACE model (Rouet, 2006) assumed that any document-based activity involves the formation of a task model, that is, a representation of the expected outcomes of the reading activity. The task model then drives subsequent processes involved in the search, evaluation and integration of information. Thus, the model is fully consistent with McCrudden and Schraw's (2007) "relevance-based reading" framework. It must be pointed out, however, that translating verbal instructions and other contextual cues into a set of relevance criteria may be a complex task in itself and readers can form different task models even when provided identical task-focusing instructions (e.g., Kobasigawa, 1983). As shown in Figure 1, during Step 1 the reader generates a task representation and a set of goals that will serve as a basis for further relevance-based mechanisms (note that whether the reader will consider the instructions prior to or after reading depends on a number of conditions. See Vidal-Abarca, this volume). During Step 1, the reader examines and interprets the directions against his or her knowledge of the domain and prior experience of similar tasks. The reader also takes into account any other constraints or opportunities that are present in the context, such as time limitations or availability of information resources. The outcome of Step 1 is a set of

goals and also, possibly, plans as to what actions should be taken in order to reach the goals. These goals and plans then become part of the internal resources. The task model is dynamic and can be updated as a result of subsequent stages.

In the climate change example presented in Table 1 above, Lisa has to read and comprehend her teacher's assignment "to find, review and discuss documents on the issue of whether human activity contributes to the global increase in temperatures on Earth". Such a statement typically includes two components that are equally critical in supporting relevance-based reading: a content area and search procedures.

First, the assignment defines a content area to be searched and studied. In this case, the definition of the content area includes a given (i.e., that temperatures are globally increasing on earth) and a query (i.e., whether human activity contributes to the increase). To interpret the statement, Lisa must possess vocabulary knowledge and reading comprehension skills just as she does when comprehending text. She must also possess question-parsing strategies that will let her know what is given and what is asked (Graesser & Franklin, 1990). Young students have often trouble when the task specifications do not readily convey explicit search contents. For instance, Kobasigawa (1983) gave Fourth and Eighth graders an introduction to the issue of food production in China and asked them to propose a topic for a research project on this issue. Only one Fourth-grader out of 20 was able to generate an open-ended question that would require new information to be found (15 out of 20 Eighth graders did so). In another study, Moore (1995) reported that Sixth graders expected to find books whose title matched exactly the terms of their search question.

Second, the assignment defines a range of procedures to be conducted in relation to the question, in this case "find, review and discuss documents on the issue of (…)" (Table 1). Understanding this part of the assignment involves understanding what it means to find, review and discuss documents about something. These actions differ from, for instance, "state your opinion on the issue of (…)", or "locate and summarize one document on the issue of (…)". Defining appropriate procedures requires a consideration of pragmatic dimensions of the situation: how much time and what kind of resources are available, the length and quality level of the expected outcome (e.g., first draft vs. polished essay), and the consequences at stake (scores, marks, and so forth). This procedural component of the reading assignment taps into Lisa's prior experiences of document-related tasks and contributes to shaping her action plan. As evidenced in studies of self-regulated reading, planning a sequence of actions is a characteristic of older and better learners. For instance, Wagner and Sternberg (1987, experiment 2) found that only one third of college students used information about task difficulty in order to plan their reading of a series of short texts. Those students who reported explicit planning strategies obtained the best scores on a comprehension test.

If successful, the result of Step 1 is a Task Model which includes a set of goals and procedures for reaching those goals as well as a set of criteria that will help the reader (a) assess the topical relevance of potential documents and passages within documents, (b) set a number of documents to be located and read (in the case of Lisa's assignment, presumably at least two), and (c) envision the kind of output that the teacher is asking for (in this example, a review and discussion of the documents). In the MD-TRACE model this newly created Task Model (i.e., mental representation of the task) becomes part of the student's internal resources (Figure 1, top-right).

Step 2. Assess information needs

Step 2 in the MD-TRACE model is a decision point where the reader has to assess his or her information needs. In order to do that, the reader will compare his or her prior knowledge with the demands of the task. In our climate change scenario, Lisa may realize that she does not have the requisite knowledge to start drafting an essay from scratch. Furthermore, the need to use documents is explicitly stated in the task directions. Therefore, Lisa has at least two good reasons to turn to external resources.

When task specifications are less explicit, however, younger students may sometimes be confused as regards to how best to address the task. Raphael (1984) reported that Fourth to Eighth grade students cannot always tell whether a question may be answered using

one's prior knowledge, by looking at explicit information in the text, or by performing an inference based on text information. Rouet and Coutelet (2008) obtained anecdotal evidence that Fourth graders sometimes tend to answer a search question based on spontaneous reactions to question components. For instance, a Fourth grader who was asked to find out what coyotes and vultures have in common responded that "the vulture is mean". The student had to be further invited to search information in the book (even though this had been previously explained and demonstrated).

Furthermore, as explained below, Step 2 is likely to recur several times during the activity, as the reader acquires information, updates the task product and gets closer to a satisfactory outcome. Information needs may evolve dynamically with respect to both the amount and type of information needed. Such reassessment of information needs after some information has been found can be a challenge for grade school students. For instance, in Rouet and Coutelet (2008) Third, Fifth and Seventh graders had to search an encyclopedia to answer questions, some of which involved the comparison of two animals. Younger students "tended to forget the question en route, to lack an explicit strategy or to ignore content representation devices such as the table of contents, the index or meta-textual signals present on the content pages". Similar difficulties in assessing one's information needs are reported in Kobasigawa (1983) and Moore (1995).

Students may also find Step 2 challenging for the opposite reason. Rather than incorrectly assuming that they can answer a question without additional information, they may persist in searching for information even though they could answer the question without searching for additional information. Although there is no published evidence on this issue (at least to our knowledge), student may have difficulty monitoring the accuracy of their comprehension (see also Vidal-Abarca, this volume), which could complicate the execution of this step.

An implication of Step 2 for a relevance-based theory of reading is that readers do not just have to set up appropriate goals: they also have to sustain their goals throughout the reading process. As evidenced in the examples above, remembering one's goals and reassessing information needs across multiple cycles of document selection processing often represents a challenge in itself for inexperienced readers of multiple documents. The question "do I need information and what kind?" may even get more complicated as students acquire new knowledge and fulfill some, but not all of the task requirements. More generally, one's need for information has to be assessed based on the initial set of task specifications, but also taking into account the evolving knowledge base and task products.

Steps 3a, 3b, and 3c all belong to the complex and interactive process of accessing, processing and integrating information within and across documents. We introduce them as separate steps, but with the view that they may be highly interrelated.

Step 3a. Assess item relevance

Step 3a consists in assessing the relevance of items that make up the external resources as the reader experiences them. A concrete example maybe the three documents represented in Table 1. The concept of relevance has been defined in many ways across diverse fields such as logics, communication, psychology and information science (Saracevic, 2007). In the context of document-based activities, relevance may be defined as a ratio between how much a document contributes to the reader's goal and how much effort it takes for the reader to access and process the document (Wilson & Sperber, 2002). According the MD-TRACE model, relevance assessment must consider topical overlap (i.e., whether the documents deals with the topic of the task), but also task relevance. As shown in Table 4 task relevance includes an evaluation of degree to which one considers factors such as reliability and usefulness of content as well as information accessibility.

Table 4 about here

In addition to assessing topicality (e.g., relevant to a home buyer or thief), readers may assess the relevance of a document based on the reliability or trustworthiness of the content, generally by evaluating a document's source characteristics (e.g., expertise or bias

of author, date of publication and publishing organization) and comparing information consistency across documents (i.e., corroboration). This credibility assessment may play a part in the readers' decision to retain or to reject the information. There is ample evidence that teenage readers' awareness of sources is generally limited and simplistic. For instance, Wineburg (1991) reported that high school students tended to trust their textbook as a most reliable source of information, whereas they disregarded source information in the other documents. Brem, Russell and Weems (2001) reported that high school students tended to consider the author's credentials as an absolute parameter in assessing credibility. Being a doctor or a scientist was often reported as a sign of content credibility, regardless of the context or purpose of the message.

According to the MD-TRACE model, evaluating the usefulness of a document's content in meeting the required goal(s) of the task will be a critical factor in determining whether a document is relevant. For example, different information will be relevant from the same documents if one's task is to find evidence or reasons (e.g., write an argument about the causes of global warming), explain (e.g., describe the process by which humans can increase global warming), or retrieve a fact (e.g., how much has the earth warmed in the past 50 years). Assessing task relevance will be facilitated by providing documents whose author's goals are closely aligned with the student's assigned goal. Facilitating initial processing, however, may not be an optimal condition for deep learning (Bjork, 1994 ; Wiley & Voss, 1999).

There is a cost to accessing and processing information from external sources, and according to the MD-TRACE model, the reader will consider the perceived cost when deciding whether relevant information is available. This type of relevance assessment is most often an iterative process, where readers access and evaluate several items in a row, hence the feedback loop represented in Figure 1 at Step 3a. While this sense of relevance may appear rather practical, we expect that perceived effort and perceived standard of the product will likely play a role in determining relevance.

Assessing how easily information can be retrieved (e.g., click to a pdf vs interlibrary loan or photocopy from library) will be partially determined by factors such as the number of steps required to retrieve the document(s) and the ease and usability of the available search and retrieval devices. There are three common approaches to accessing content in complex information systems available to students. First, the reader may use some document search device, such as an on-line catalog or a search engine, in order to identify one or several documents of potential interest. Second, the reader may examine various characteristics of the document (also referred to source information), such as the title, author, publisher, date and so forth. Third, the reader may use content organizers, such as a table of contents or an index, in order to locate a particular section of interest within a document. Readers can uses these three approaches in conjunction and iteratively to determine whether to retrieve a document.

These approaches rely on potentially complex language and memory skills, including knowledge of and skill in using document search devices and text organizers (Rouet, 2006, Chapter 2). Students' difficulties in generating appropriate search phrases have been well documented (Borgman, Hirsh, Walter, & Gallagher, 1995; Dinet, Favart, & Passerault, 2004). The production of adequate search phrases -- not too general but not too specific -- relies heavily on one's vocabulary knowledge. Grade school students often reuse the words they have explicitly read or heard as part of the assignment to formulate their search queries. When an initial query fails to return relevant links, they have difficulty reformulating the query. The order of hits in the search list also has a critical and sometimes detrimental impact on the search process. Search engines are supposed to list search hits as a function of their relevance. However, information providers may artificially increase their likelihood to be listed high in the list or to simply purchase the best slots through commercial partnerships. So the "top-is-best" heuristic is not always warranted.

Ready-made indexes and menus may appear relatively simpler to operate than keyword search engines, as they do not involve the production of language. However, the use of these devices is not trivial either. Most grade-school students up to the Fifth grade do

not effectively use the index in a printed book (Dreher & Sammons, 1994; Rouet & Coutelet, 2008), in part because they do not know which word to search for. They also face difficulties when performing search through a table of contents, and they often prefer to browse directly through the content pages. These are important obstacles for the use of relevance-based strategies when selecting documents.

Accessibility can also be defined by how easily the student feels he or she can comprehend the document. Because this will involve an assessment of a documents' readability, required knowledge demands, and available help to interpret the document, we expect that this type of relevance assessment will require active engagement in the task, as well as the monitoring of comprehension processes.

In Lisa's search scenario (Table 1), Lisa may decide to go through the documents in their list order (Britt, Rouet, & Perfetti, 1996). This may have several consequences on her activity. First, she may waste time studying a less important document, and run out of time before she can look at other, more relevant ones. The list order strategy may also affect Lisa's relevance judgments, as the information initially found in text is known to have a special impact on the reader's content representation (Kozminsky, 1977; Kieras, 1980). The initial document read presumably has a special impact on how the reader makes further document selections as well as how they design their response to the task. So far, however, no study of the use of multiple documents as provided any empirical data on this issue.

In addition to decisions about the order of document selection, Lisa may need to make additional relevance-based decisions as she goes through the texts listed in Table 1. There is the decision about which documents to read because not all documents will necessarily contain information useful for her to write her review. Because this decision may not be clear until she begins reading, Lisa may need to read efficiently for making the decision to continue reading. Therefore, the efficient reading of lengthy documents may (and usually does) involve a combination of rapid scanning and slower, in-depth processing of the passages (Cerdan, Vidal-Abarca, Martinez, Gilabert, & Gil, 2009; Rouet et al., 2001) and Lisa may or may not be skilled in exerting such flexible control on her reading rate, pausing, skipping and backtracking or in using signals and organizing devices to locate or infer the contents of the passages. In fact, Rouet and Coutelet (2008, experiment 2) found that these strategies were not typical of students' search behavior until Grade seven.

Acquiring new information may sometimes challenge students' initial representation of the task. For instance, Lisa may initially have understood that her teacher wanted her to compare the various ways in which human activity contributes to global warming. Upon reading D1, however, she may realize that some people believe that non human causes may explain global warming. She may then reinterpret what the phrase "different perspectives" in her task model really means (R1).

In summary, in Step 3a the reader uses various dimensions of relevance to decide whether to read or continue reading a document, in which order to read the documents, and how deeply to process the information at any point in time.

Step 3b. Process text contents

In Step 3b, the reader engages with the document in order to extract and integrate content information. In the best-case scenario, the document is globally relevant and the reader may use a relevance-based comprehension strategy (McCrudden & Schraw, 2007) to build up a content representation. It is then likely that readers will alternate between the rapid scanning of passages and the deeper, more systematic processing of contents, based on how relevant the information seems to be. At that stage, the specificity of the task plays an important part. For highly specific tasks such as locating a particular fact in the document, skilled readers may simply skip most of the information and engage in deeper processing only when they reach the passage of interest (McCrudden, Magliano, & Schraw, 2010; Rouet et al., 2001). In those cases, readers may not even form a mental representation of the text as they would for comprehension. Instead, they merely use the text as a resource for question answering.

Whether or not readers will adjust their reading speed and depth of processing to the relevance of text passages depend on a number of individual characteristics. Just like deep comprehension, the scanning of text for relevant information draws on the reader's prior

knowledge. Symons and Pressley (1993) found that college undergraduates' ability to locate the answer to search questions in a textbook increased after the students had received some instruction on the topic. Rouet and Le Bigot (2007) reported that students with more college education spent more reading time on the relevant portion of a hierarchical hypertext and provide more specific answers to a complex open-ended question (See Vidal-Abarca, this volume, for other sources of individual differences in goal-based reading; see also McCrudden, Magliano & Schraw, 2010).

In our simple scenario, Lisa's ability to read and understand the contents of the documents is largely a function of the prior knowledge she possesses about the topic. In multiple document environments, however, readers have the option to discard difficult texts and to turn to more accessible ones. For instance, the scientific report uses difficult language and explanations, and therefore this document is inappropriate for a reader like Lisa even though it is topically relevant. The criteria that students like Lisa actually use in order to assess documents' adequacy is open to further investigation. Text comprehension research, however, has found ample evidence that readers do not always accurately monitor their comprehension (Dunlosky, & Lipko, 2007; Maki, 1998; Thiede, Anderson, & Therriault, 2003; Thiede, Griffin, Wiley, & Redford, 2009). Therefore, it is likely that grade school students' perception of difficulty level may only be a partially operational factor in relevance-based decisions.

Step 3c. Create/update a Documents Model

Step 3c is typical of multiple document processing. Once information has been acquired from a particular document, the reader has to combine it with information found in other documents. The combination processes may take several forms, as discussed in Britt et al. (1999). Documents may add to each other, support each other, but also oppose or contradict each other. Thus, in the case of multiple document comprehension, Step 3c is far from a trivial accumulation process. Instead, the reader may use a variety of rhetorical predicates to relate information across documents. More importantly, the reader may need to attach content information to its respective source, so as to be able to later remember where the information was found.

Remembering source-to-content links and connecting documents through rhetorical predicates amounts to forming what Perfetti et al. (1999) called a Documents model. Documents models are especially useful when working on complex reading assignments, where a relevance-based strategy does not consist in just identifying what information is germane to the task, but also how the information contributes to informing the issue. For instance, experienced historians working on a controversial topic such as the U.S.'s 1903 military intervention in Panama do not seek to find out whether or not the intervention was right or wrong, but who made claims on either side of the issue, and based on what type of evidence. Therefore, they form memory representations that include the sources, the contents and the relations across texts (Rouet, Favart, Britt, & Perfetti, 1997; see also Wineburg, 1991).

In the case of Lisa's assignment (Table 1), forming a documents model would amount to contrasting documents D1 and D2 as supporting opposed conclusions on the issue of human activity and human warming. Lisa would also link D2 and D3 because D2 makes use of D3 as evidence to support the conclusion that humans are at least partly responsible for global warming.

However, creating integrated models from multiple documents is in large part beyond the average middle school students' ability. Wineburg (1991) observed that even Advance Placement (i.e., above average) high school students tended to focus on the textbook as a necessary and sufficient resource and disregarded other possible sources of knowledge about a controversial event. Rouet et al. (1997) even found differences between graduate students specializing in history vs. psychology in their ability to integrate information from a set of historical documents. Interestingly, the difference could not be attributed to students' prior knowledge of the contents. Instead, Rouet et al. suggested that history students' greater

ability to deal with the documents came from their repeated exposure to similar tasks and documents types, a form of expertise they labeled "discipline expertise".

At least some of the skills needed to organize multiple documents into coherent models can nevertheless be taught at the high school and college level, as evidenced by studies of Britt and Angliskas (2002) and others (Stadtler & Bromme, 2007 ; Wiley, Goldman, Graesser, Sanchez. Ash, & Hemmerich, 2009 ; see also Britt & Rouet, in prep.)

Step 4. Create/update a task product

In Step 4, the reader makes use of document information to construct a response to the task. At this stage, the reader needs to take into account the initial task specifications, the information acquired from the documents, and the type of response or product that is expected. The amount of transforming required to turn document information into a satisfactory response depends both on the type of task (Wiley & Voss, 1999) and on the type of match between the document and the task (Mosenthal, 1996). Furthermore, in most real-life situations such as document-based essay writing, readers will start producing notes and drafts as they go through the available information sources. This interplay of text comprehension and production activities, though important, is not further discussed here (See McGinley, 1992; Voss & Wiley, 1997; Wiley & Voss, 1999 for further discussions).

Even in simple question-answering task contexts, the shaping of a task product may prove difficult. Moore (1995) reported that, among other difficulties, Sixth graders did not always understand that they had to actually build up an answer from the information found in books. As part of an earlier study (Rouet & Coutelet, 2008), the first author of this chapter observed that Third and Fifth grade students who were searching an encyclopedia in order to answer factual questions would sometimes read and re-read a target passage they had correctly located, but remained unable to turn the information into an actual answer.

In more complex document-based tasks, such as using document information to write an argument, Step 4 represents a greater challenge to students. Only a minority of students are skilled in producing written arguments by the end of 12th grade (National Assessment of Educational Progress, 1996, 1998). High school students and college undergraduates often write argumentative essays that fail to state opposing views or respond to opposition (Knudson, 1992, 1994; Kuhn, 1991; Wolfe, Britt, & Butler, 2009). Their essays often failed to take an explicit position on the assigned controversy and often simply listed reason without sufficient elaboration or support (Knudson, 1992, 1994; Wolfe et al., 2009).

We believe that each of these problems can be partially traced to an inappropriate task model. For instance, Wolfe and Britt (2008) found that students perceived other side information to be irrelevant to the task of searching for or selecting information for inclusion in their argumentative essay. While most of the undergraduates did select to read documents presenting an other-side position, half of the students did not include any other side information in their essays. Of those that did not include other side information, some had failed even to take any notes on this information (43%) suggesting that they did not consider this information important for their task, and most (57%) did include information supporting the opposing claim in their notes, suggesting that they considered this important to understanding the controversy but not important to include in their argumentative essay. Further support for the deficient task model hypothesis is that instructions to conduct a balanced search significantly reduced the number of myside essays (Wolfe & Britt, 2008) as did instructions to include other side information (Wolfe et al., 2009).

Wolfe and Britt (2008) also found many essays that either did not take an explicit stance on the assigned controversy or that allowed their stance to shift or change throughout the essay. Such problems may be due, in part, to an inappropriate task model since instructions were helpful (Wolfe et al., 2009) but may also be due to a failure to continually monitor alignment of the task goals with the information learned or the product created throughout the writing process.

Finally, the problem of listing unelaborated reasons could be partially due to a poor task model that is not realizing that elaborative support is required. If they do not believe elaboration is important, they will either end the acquisition phase prematurely (i.e., decision

at Step 2 to stop Steps 3a-c) or end the production phase prematurely (i.e., decision at Step 5 to stop Step 4). In fact, simply telling students that they should provide enough elaboration of each reason to convince the reader did decrease the amount of simple reason listing in college students' argumentative essays (Wolfe, et al., 2009).

Although some of the difficulties students have in producing arguments can be the traced to an insufficient task model and can be remedied by task-interpretation instructions, other difficulties are likely the result of a deeper problem. Arguably three of the most challenging aspects of argumentation tasks, writing complete, audience-appropriate arguments, evaluating the theory-evidence connection, and counterfactual reasoning, all stem from the underlying demand to take and use different perspectives. The first aspect, writing complete, audience-appropriate arguments, requires that students mention other side information, such as counter arguments, and select reasons, counter arguments and responses to counters for which the audience will be most receptive. Each of these argument elements will be effective to the extent that the reader can understand the prospective audience's position. For example, employers and students would not both be equally persuaded to endorse the assertion that comprehensive exams are important by arguing that such exams will make the students apply more effort during studying so they learn more deeply. For students to write arguments that can be effective to different audiences, they have to be able to think about the proposed situation through the filter of the prospective audience's beliefs and knowledge.

The second aspect, evaluating the relationship between theory and evidence, also requires students take another perspective. In this case, they have to work within the constraints of the theory. For example, asking the student to examine a situation and make predictions based on how Freud, Piaget, and Skinner would explain the subject's behavior would require the student to simulate each theory and test for belief consistency. The same challenge occurs on a more simple level in thinking about the effects of isolated physical, social or economic mechanisms in which the student has to think from the perspective of a mechanism functioning in the world. Similarly, the audience may implicitly or explicitly require that the student set a criterion for evidence that is scientifically-based rather than an argument supported by pseudo-evidence, anecdotes, or simply opinion. To the extent that their own perspective is not aligned with such a perspective epistemically, this will also require the student to think outside his or her own belief system.

The final aspect, counter-factual reasoning, involves making inferences or predictions from premises or facts that are different than the way the world actually exists. This type of thinking or reasoning requires the challenge of simulating the perspective of a potential world to project an outcome e.g., what would have happened if the White House had had a better understanding of the Vietnamese people in 1964 or if they had not escalated the war in response to the activity in the Gulf of Tonkin?

Each of these aspects of the production task requires the student to take another person's perspective (e.g., person, group, theory, or situation) in a much deeper way than thinking about what objects may be important to a thief. According to the MD-TRACE model, these tasks are demanding for several reasons. In terms of internal resources, the student has to have the knowledge to construct such a perspective (e.g., knowing Skinner's theory in detail) as well as the skill in switching perspectives and the knowledge of the importance of doing so. While reading and experience can help a student build up the knowledge required for creating these various perspectives, we expect there are developmental constraints that will be too demanding for younger students.

Allowing various perspectives to guide the creation of a task model, a document model, and the task product can also be very challenging. Since this perspective is not in line with the student's own perspective, we expect that it will require effort and attention to maintain the activation of this perspective and monitor internal consistency. Thus, there will likely be developmental constraints on students' ability to selectively attend and control their own memory and search processes.

In Lisa's climate change scenario, the expected outcome includes not just selecting key facts, but also turning source information into a new, coherent piece of text. Thus, in

addition to her comprehension ability, Lisa's text production skills will also be required and most probably challenged. The three source documents described in Table 1 may use diverse language standards that may be difficult to adapt. Furthermore, the conclusions of the documents are discrepant, which prohibits the writing of a simple, main idea type of summary. In fact, it may be surprising and confusing for Lisa to find such discrepancies and she may not know how to proceed. For Lisa to adequately integrate conflicting information, she would have to use sources and rhetorical predicates such as "disagrees" and "supports" in order to accurately report on the topic, which taps on her ability to create a new rhetorical structure.

Step 5. Assess whether the product meets the task goals.

Once a task product has been generated the reader may need to assess whether the product (4) meets the task goals (1). If the product is deemed satisfactory, the reader will terminate the process. If not, the reader may need to recycle through previous stages. Three main types of recycling may occur. In the simplest case, the reader will complete or revise his or her product in order to improve it (Figure 1, R1). For instance, after re-reading her summary of document D2, Lisa may realize that she needs to state the conclusions that emerge from the magazine article as regards human contribution to global warming.

If the reader can't think of an obvious improvement of the task product, he or she may need to reassess their information needs (Figure 1, R2), and possibly proceed to select a new document or to re-read a previously read document. In our scenario, after she has accounted for D2, Lisa may remember that she had spotted another relevant document and proceed to study it.

Decisions to recycle through earlier problem-solving stages typically belong to the reader's self-regulation skills (Greene & Azevedo, 2007; Winne & Hadwin, 1998). Self-regulation depends in part on higher-order memory processes that develop late during childhood and adolescence. In addition, those skills appear to be related to other dimensions of individual differences among secondary and post-secondary students such as epistemic beliefs (Bråten & Strømsø, 2005). The limitation of teenage students' self-regulating ability is a factor to consider when assigning complex document-based reading activities, whether in print or online (Azevedo, Guthrie & Seibert, 2004).

Summary and conclusions

In this chapter, we have proposed a framework to describe the resources and processes involved in studying multiple documents in relevance-based contexts. The MD-TRACE model defines the resources and processing steps involved in readers' interactions with complex documents and can be useful for specifying the relevance-based mechanisms that pervade the entire document processing cycle. McCrudden and Schraw (2007) refer to this process as a goal-focusing, but do not explicitly detail the complex interactions between these resources and processes or their recursive nature.

Summary of the MD-TRACE model components

The MD-TRACE model includes five main steps: task model construction (Step 1), information need assessment (Step 2), document processing (Steps 3a to 3c), task product creation (Step 4), and task product assessment (Step 5). In most normal circumstances, the five steps are likely to unfold in a highly interactive manner. We summarize the key features of each step and the main relevance-based processes that take place during them.

Relevance assessment initially takes place during the readers' examination of task specifications and pragmatic constraints, i.e. prior to the reader's actual engagement with texts. We have illustrated the potential complexity of the task model construction step (Figure 1, Step 1), by pointing out that reading tasks usually involve a description of the topic or issue to be studied, but also a set of requirements concerning what is to be done with the information and how to do it. In order to make appropriate decisions concerning document information, the reader needs to construct an accurate task model. The reader also needs to sustain the model throughout the entire study period, and potentially to update it as a function of the outcomes and obstacles found en route.

The MD-TRACE model acknowledges the reader's assessment of information needs (Step 2) as one of the core steps in document-based activities, but also potentially in any

functional reading situation. It is important to note that Step 2 is part of a loop that also involves thinking about the task requirements (Step 1), engaging with documents (Steps 3a-3c), and constructing a task product (Step 4). Indeed, most reading situations involve a combination of various cognitive activities, with the actual sustained reading of continuous passages being only one component. One challenge for further research, as explained below, is to find out how one's awareness of information needs drive subsequent steps in the reading process.

Unlike simpler models of text processing, the MD-TRACE model defines three core substeps to describe how readers interact with multiple documents: First, students have to select appropriate or relevant documents or passages within a document (Figure 1, Step 3a). Second, students have to accurately process the selected document (Figure 1, Step 3b), which involves a combination of rapid scanning and slower, more in-depth processing of the materials (with the latter presumably involving situation model construction as described in Kintsch's (1998) theory of text comprehension). According to the MD-TRACE model, the reader's decision to engage in reading a particular passage, to switch from scanning to deeper processing, and to quit reading the passage are mostly driven by relevance evaluations, which are supposedly conducted in parallel to the actual processing of the materials. At that point, our framework overlaps in large parts with McCrudden and Schraw's (2007) relevance-based reading. In addition, the MD-TRACE model maps relevance decisions to specific features of each document: the availability of search devices, organizing devices, and source information. Third, the MD-TRACE model also considers the integration of information across documents as a specific step (Figure 1, Step 3c). This is because, most of the time, readers cannot just aggregate information across texts, as they would do for example across the paragraphs of a single text. Instead, specific integration mechanisms, such as contrast, corroboration or support, are required (Perfetti et al., 1999).

In addition to document processing, the MD-TRACE model includes a production step where the reader makes use of text information to construct or update a "task product" (Step 4). At this point, the procedural component of the task model (Step 1) plays a critical part.

Finally, in Step 5 the reader has to decide whether the task product is good enough given the task specifications and the means available. Based on their own assessment, readers may then recycle through previous steps, either revising their product or reassessing their information needs.

Steps 2, 3a and 5, involve making decisions and selections based on one's awareness of one's prior knowledge, awareness of the task demands and monitoring of progress. These processes heavily draw on the reader's self-regulation skills. For instance, to find out whether more information is needed (Step 2), readers have to contrast their task model with their pre-existing knowledge. In Step 3a, readers again use the task model to gauge the relevance of documents or links. And in Step 5, readers check whether the task product meets the requirements of the task model.

Some research questions that emerge from the MD-TRACE model.

The study of the processes involved in using multiple documents is a burgeoning area of investigation. The MD-TRACE model can be used to identify some issues that are still largely open on the research agenda. More specifically, we suggest four different research issues that may be subjected to empirical investigation:

(1) At each decision point readers must assess progress and determine their next step by a comparison to their task model. As such, the task model is critical in guiding decisions. Therefore, it is important to get a more detailed analysis of the processes involved in creating a task model from task instructions. In other words, we need to ask how students interpret teachers' directions and turn these into an actual plan of actions. This will involve knowledge of and attention to both the predicates (e.g., identify, describe, argue, explain, contrast) as well as the semantic and lexical nouns mentioned in the task specifications (See Eduardo chapter). It is assumed that the creation of task goals will be affected by the actual task instructions (e.g., identifying a single fact, becoming informed to make a decision about a medical treatment) as well as many other more pragmatic factors (e.g., time available,

explicitness of question, instructions for achieving the goal, type and number of documents available, need to search for documents and the tools available for searching, the physical environment, and the expected level of performance expected).

Further research is also needed in better understanding the dynamic updating of one's task representation throughout the process of accessing and processing contents and generating task products. Information needs may evolve as readers acquire information and develop a task product. For instance, Lisa may initially need to understand what is the current evidence about global warming, and then to identify the various causal explanations of the phenomenon.

(2) Given the importance of decision-making steps in the MD-TRACE model (Steps 2, 3a, and 5), more research is needed to better understand how individual differences in internal resources (e.g., prior knowledge, search skill, and working memory) and external resources (e.g., reading or searching through other related or unrelated documents) affect these decisions. Because the model points to the importance of maintaining a task model throughout the processing of the documents, one serious challenge for students is to both maintain the task model and evaluate relevance while reading to create a documents model. Given the limited resources in working memory, one would assume that the task model is cast aside to make processing resources available for incoming information. The model makes it clear, however, that relevance decisions are performed almost continuously before, during and after reading, which suggests that the task model exerts a permanent regulatory function until the completion of the task. How readers manage task specifications while selecting, evaluating and processing document information has yet to be explained. For instance, how does irrelevant information found *en route* impact the availability of the task model? And what instructional conditions are likely to support students' monitoring of their task model throughout the document processing steps?

(3) MD-TRACE focuses on situations where students read and relate multiple documents. In this context, it is important to understand how students decide whether to access more information and how they conduct such a search. This may be partially a result of memory-based activation of related information (see Lassonde et al., this volume) or a more strategic search (see Vidal-Abarca et al., this volume). In searching, how does reading one document influence the relevance of other documents and the decision to continue reading? And, finally, to what extent are readers flexible and efficient in determining what is relevant and how it meets the goals in ones' task model?

(4) The MD-TRACE model predicts that in addition to topic relevance, readers will also take into account task relevant factors such as document reliability, accessibility and appropriateness to goal. How do students determine relevance based on these other characteristics? Of course, the use of these relevance criteria will be partially a result of readers' skill and knowledge. There has been little research on how relevance criteria are updated during the processing of complex documents (but see, however, de Vries & de Jong, 1999). Future research should also address individual differences in the internal resources, in particular knowledge (e.g., criteria of document reliability, search tools and document organizers, and standard forms of tasks) as well as efficiency in proceduralizing this knowledge (e.g., search skill, comprehension skill, production skill).

Understanding the cognitive and language processes that underlie readers' ability to interact with complex documents is needed in order to come up with appropriate instructional curricula and procedures. Given the increasing rate of information production and dissemination in both printed and electronic forms, training readers to assess the relevance of documentary materials is a challenge of increasing importance for educational systems, and of vital importance for the cohesion of modern societies.

References

Azevedo, R., Guthrie, J.G., & Seibert, D. (2004). The role of self-regulated learning in fostering students' conceptual understanding of complex systems with hypermedia. *Journal of Educational Computing Research*, *30*, 87-111.

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (Eds.) *Metacognition: Knowing about knowing.* (pp.185-205). Cambridge, MA: MIT Press.

Borgman, C.L., Hirsh, S., Walter, V. & Gallagher, A. (1995). Children's searching behavior on browsing and keyword online catalogs: The science library catalog project. *Journal of the American Society for Information Science*, *46* (9), 663-684.

Bråten, I., & Strømsø, H.I. (2005). The relationship between epistemological beliefs, implicit theories of intelligence, and self-regulated learning among Norwegian post-secondary students. *British Journal of Educational Psychology, 75*, 539-565.

Brem, S.K., Russell, J., & Weems, L. (2001). Science on the Web: Students' evaluation of scientific arguments. *Discourse Processes*, *32*, 191-213.

Britt, M.A., & Angliskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction*, *20*, 485-522.

Britt, M. A., Perfetti, C. A., Sandak, R., & Rouet, J. F. (1999). Content integration and source separation in learning from multiple texts. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.). *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso (pp. 209-233).* Mahwah, NJ: Lawrence Erlbaum Associates.

Britt, M.A. & Rouet, J.-F. (in prep.). Studying multiple documents: Cognitive processes and instructional implications.

Britt, M.A., Rouet, J.-F. & Perfetti, C.A. (1996). Using hypertext to study and reason about historical evidence. In J.-F. Rouet, J.J., Levonen, A.P. Dillon, and R.J. Spiro (Eds.). *Hypertext and Cognition (pp. 43-72).* Mahwah, NJ: Lawrence Erlbaum Associates.

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*, 31-42.

Cerdán, R., Vidal-Abarca, E., Martínez, T., Gilabert, R., & Gil, L. (2009). Impact of question-answering tasks on search processes and reading comprehension. *Learning and Instruction*, *19*, 13-27.

Cerdan et al., To appear in McCrudden, Magliano & Schraw's Edited volume on Relevance-based reading.

Cirilo R.K., Foss D.J. (1980). - Text structure and reading time for sentences. *Journal of Verbal Learning and Verbal Behavior*, *19*, 96-109.

Cromley, J.G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, *99*, 311–325

Dinet, J., Favart, M. & Passerault J.M. (2004). Searching for information in an Online Public Access Catalogue (OPAC) : the impacts of information search expertise on the use of Boolean operators. *Journal of Computer Assisted Learning*, *20*, 336-346.

Dreher, M.J. & Guthrie, J.T. (1990). Cognitive processes in textbook chapter search tasks. *Reading Research Quarterly, 25*, 323-339.

Dreher, M.J. & Sammons, R.B. (1994). Fifth graders' search for information in a textbook. *Journal of Reading Behavior*, *26*, 301-314.

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science*, *16*, 228–232.

Graesser, A.C. & Franklin, S.P. (1990). QUEST: A model of question answering. *Discourse Processes*, *13*, 279-303.

Greene, J.A., & Azevedo, R. A (2007). A theoretical Review of Winne and Hadwin's Model of Self-Regulated Learning: New Perspectives and Directions. *Review of Educational Research*, *77*, 334-372.

Guindon, R. & Kintsch, W. (1984). Priming propositions: evidence for the primacy of macropropositions in the memory for text. *Journal of Verbal Learning and Verbal Behavior*, *23*, 508-518.

Guthrie, J.T. (1988). Locating information in documents: examination of a cognitive model. *Reading Research Quarterly*, *23*, 178-199.

Guthrie, J.T. & Kirsch, I. (1987). Distinctions between reading comprehension and locating information in text. *Journal of Educational Psychology*, *79*, 210-228.

Hacker, D.J. (1998). Self-regulated comprehension during normal reading. In D. Hacker, J. Dunlowsky & A. Graesser (Eds.) *Metacognition in educational theory and practice (pp. 165-191)*. Mahwah, NJ: Erlbaum.

Kieras, D.E. (1980). Initial mention as a signal to thematic content in technical passages. *Memory and Cognition*, *8*, 345-353.

Kiewra, K.A., DuBois, N.F., Christian, D., McShane, A., Meyerhoffer, M. & Roskelley, D. (1991). Note-taking functions and techniques. *Journal of Educational Psychology*, *83*, 240-245.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*, *95*(2), 163-182.

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

Kintsch, W., Kozminsky, E., Streby, W.J., McKoon, G. & Keenan, J.M. (1975). Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior*, *14*, 196-214.

Kintsch, W. & van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363-394.

Kintsch, W. & Yarbrough, J.C. (1982). Role of rhetorical structure in text comprehension. *Journal of Educational Psychology,* 74, 828 834.

Knudson, R.E. (1992). The development of written argumentation: An analysis and comparison of argumentative writing at four grade levels. *Child Study Journal*, *22*, 167-184.

Knudson, R.E. (1994). An analysis of persuasive discourse: Learning how to take a stand. *Discourse Processes*, *18*, 211-30.

Kobasigawa, A. (1983). Children's retrieval skills for school learning. *The Alberta Journal for Educational Research*, 29(4), 259-271.

Kozminsky, E. (1977). Altering comprehension: the effect of biasing title on text comprehension. *Memory and Cognition*, *5*, 482-490.

Kuhn, D. (1991). *The skills of argument*. New York, NY, US: Cambridge University Press.

Lassonde et al., this volume). To appear in McCrudden, Magliano & Schraw's Edited volume on Relevance-based reading.

Maki, R. H. (1998b). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice (pp. 117–144).* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McCrudden, M.T., Magliano, J., & Schraw, G. (2010). Exploring how relevance instructions affect personal reading intentions, reading goals and text processing: A mixed methods study. *Contemporary Educational Psychology*, *in press.*

McCrudden, M.T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review*, 19(2), 113-139.

McGinley, W. (1992). The role of reading and writing while composing from sources. *Reading Research Quarterly*, *27*, 227-248.

Moore, P. (1995). Information problem-solving: A wider view of library skills. *Contemporary Educational Psychology*, *20*, 1-31.

Mosenthal, P. (1996). Understanding the strategies of document literacy and their conditions of use. *Journal of Educational Psychology*, *88*(2), 314-332.

Mosenthal, P.B., & Kirsch, I.S. (1991). Toward an explanatory model of document literacy.

National Assessment of Educational Progress (1996). *NAEP 1994 U.S. History Report Card: Findings from the National Assessment of Educational Progress*. Princeton, NJ: Educational Testing Service.

National Assessment of Educational Progress. (1998). *NAEP 1998 Writing Report Card: Findings from the National Assessment of Educational Progress*. Princeton, NJ: Educational Testing Service.

Newell, G.E. & Winograd, P. (1989). The effects of writing on learning from expository text. *Written Communication*, *6*, 196-217.

Ozuru, Y., Best, R., Bell, C, Witherspoon, A., & McNamara, D. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction*, 25(4), 399-438.

Perfetti, C.A., Rouet, J.-F., & Britt, M.A. (1999). Towards a theory of documents representation. in H. van Oostendorp & S.R. Goldman (Eds.) *The Construction of Mental Representations During Reading (pp. 99-122).* Mahwah, NJ: Lawrence Erlbaum Associates.

Puustinen, M., & Rouet, J.-F. (2009). Learning with new technologies: Help seeking and information searching revisited. *Computers and Education*, *53*, 1014-1019.

Raphael, T.E. (1984). Teaching learners about sources of information for answering comprehension questions. *Journal of Reading*, January, 303-311.

Rouet, J.-F. (2006). *The skills of document use: from text comprehension to Web-based learning*. Mahwah, NJ: Erlbaum.

Rouet, J.-F., & Coutelet, B. (2008). The acquisition of information search skills by 9 to 13 year-old students. *Applied Cognitive Psychology*, *22*, 389-406.

Rouet, J.-F., & Eme, E. (2002). The role of metatextual knowledge in text comprehension: Issues in development and individual differences. In P. Chambres, M. Izaute and P.J. Marescaux (Eds.) *Metacognition: Process, function and use (pp. 121-134)*. Amsterdam: Kluwer.

Rouet, J.-F., Favart, M., Britt, M.A. & Perfetti, C.A. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and Instruction*, *15*(1), 85-106.

Rouet, J.-F., & Le Bigot, L. (2007). Effects of academic training on metatextual knowledge and hypertext navigation. *Metacognition and Learning*, *2*, 157-268.

Rouet, J.F., & Vidal-Abarca, E. (2002). "Mining for meaning": A cognitive examination of inserted questions in learning from scientific text. In J. Otero, J.A. Leon, & A.C. Graesser (Eds.) *The Psychology of Science Text Comprehension (pp. 417-436).* Mahwah, NJ: Lawrence Erlbaum Associates.

Rouet, J.-F., Vidal-Abarca, E., Bert-Erboul, A. & Millogo, V. (2001). Effects of information search tasks on the comprehension of instructional text. *Discourse Processes*, *31*(2), 163-186.

Saracevic, T. 2007. Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology* 58, no. 13: 1915-1933.

Schraw, G., Wade, S. E., & Kardash, C. A. (1993). Interactive effects of text-based and task-based importance on learning from text. Journal of Educational Psychology, 85, 652–661.

Seigneuric, A., Ehrlich, M.F., Oakhill, J.V., Yuill, N.M. (2000). Working memory resources and children's Reading comprehension. *Reading and Writing*, *13*, 81–103.

Snow, C. & the RAND reading study group (2002). *Reading for understanding. Toward a R&D program for reading comprehension.* Santa Monica, CA : RAND

Stadtler, M. & Bromme, R. (2007). Dealing with multiple documents on the WWW: The role of metacognition in the formation of documents models. *International Journal of Computer Supported Collaborative Learning, 2*, 191-210

Symons, S. & Pressley, M. (1993). Prior knowledge affects text search success and extraction of information. *Reading Research Quarterly*, *28*(3), 250-261.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*, 66–73.

Thiede, K.W., Griffin, T. D.,Wiley, J., & Redford, J. S. (2009). Metacognitive monitoring during and after reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser, (Eds.), *Handbook of metacognition in education (pp. 85–106).* New York: Routledge.

Vidal-Abarca: To appear in McCrudden, Magliano & Schraw's Edited volume on Relevance-based reading.

Voss, J.-F. & Wiley, J. (1997). Developing understanding while writing essays in history. *International Journal of Educational Research*, *27*, 255-265.

de Vries, E. & de Jong, T. (1999) The design and evaluation of hypertext structures for supporting design problem solving. *Instructional Science*, 3-4, 285-302.

Wagner, R.K. & Sternberg, R.J. (1987). Executive control in reading comprehension. In B.K. Britton & S.M. Glynn (Eds.) *Executive Control Processes in Reading (pp. 1-22)*. Hillsdale, NJ.: Lawrence Erlbaum Associates.

Wiley, J., Goldman, S., Graesser, A., Sanchez. C., Ash, I. & Hemmerich, J. (2009) Source evaluation, comprehension, and learning in internet science inquiry tasks. *American Educational Research Journal,* 46, 1060-1106.

Wiley, J. & Voss, J.F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding not just memory for text. *Journal of Educational Psychology*, *91*, 301-311.

Wilson, D. & Sperber, D. (2002). Relevance theory. *University College London Working Papers in Linguistics, 14*, 249-287.

Retrieved June 3, 2010 from http://www.phon.ucl.ac.uk/home/deirdre/papers.html

Wineburg S.S. (1991). Historical problem solving: a study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, *83*, 73-87.

Wineburg, S.S. (1994). The cognitive representation of historical texts. in G. Leinhardt, I. Beck & C. Stainton (Eds.) *Teaching and Learning in History (85-135)*. Hillsdale, NJ: Erlbaum.

Winne, P.H. & Hadwin, A.F. (1998). Studying as self-regulated learning. In D. Hacker, J. Dunlowsky & A. Graesser (Eds.) *Metacognition in educational theory and practice (pp. 277-304)*. Mahwah, NJ: Erlbaum.

Wolfe, C.R., Britt, M.A., & Butler, J.A. (2009). *Argumentation Schema and the Myside Bias in Written Argumentation*. *Written Communication*, 26, 183-209.

Wolfe, C.R. & Britt, M.A. (2008). The locus of the Myside Bias in written argumentation. *Thinking and Reasoning*, *14*(1), 1–27.

Table 1. A simple scenario of multiple document use.

Lisa, a 9th grader, is working on a science project on global warming. The science teacher has asked Lisa and her teammates "to find, review and discuss different perspectives on the issue of whether human activity contributes to the global increase in temperatures on Earth, and write a short essay addressing this issue". Based on a search engine query, Lisa locates three web pages that seem of interest:

(D1) A web site on "human activities and climate change" designed by a group of 10th grade students at a public high school as part of their coursework. The students concluded that climate change is largely due to natural circumstances.

(D2) A passage in the online version of a news magazine commenting on the release of a new study by a climate research institute. The passage cites the study as providing new evidence of a causal link between human activity and global warming.

(D3) A scholarly article reporting on the new study of the production of greenhouse gases as part of human activities released by the climate research institute. The study concludes that human production of greenhouse gases is most likely a significant factor of global warming.

For the next hour or so, Lisa engages in a variety of reading acts that will lead her to write a one-page draft essay addressing the teacher's assignment. She will then share her draft with her teammates for discussion.

Table 2. Three types of external resources involved in using documents

Resource	Definition	Examples
Task specifications	Verbal statements conveying instructions, questions, directions or assignments.	"review and discuss documents on the issue of human activity and climate change"
	Non-verbal constraints such as time available, cost and potential limitations in accessing the information.	Looking up documents for 10 minutes vs. a whole afternoon. Language in which the text is written.
	Type and level of performance expected, potential consequences of the activity.	Study for own interest or in order to prepare for an exam.
Information resources	Document search, categorization and access devices	List returned by a search engine based on the request "human activity and climate change"
	Documents available, including source and content information	A scholarly article on the production of greenhouse gazes as part of human activities released by the climate research institute
	Other, non-documentary resources	Lisa's teacher and teammates
Reader-generated resources	Verbal and graphical information produced by the reader upon or after reading.	Notes, highlights, clips, summaries, diagrams, and the actual written response to the task requirements, if any (e.g., answer, essay, solution)

Table 3. Permanent and transitory cognitive resources involved in using documents.

Resource	Definition	Examples
Permanent resources (*)		
General world knowledge	The reader's prior knowledge of the topic at hand and the contents of the text	Lisa's knowledge of climate, climate change, and human activities that may affect climate
Prior task experience	The reader's experience and proficiency in the activities required to fulfill the task	Lisa's reading and comprehension ability; the number and diversity of her document-based learning experiences
Working memory skills	The capacity and effectiveness of short- and long-term memory mechanisms	Lisa's verbal span; her planning and executive control abilities
Transitory resources		
Task model	Reader's understanding of the task requirements	Lisa's understanding of the assignment; her awareness of time constraints and information resources available; her goals and action plan
Document model	Readers' understanding of the documents' sources and contents, and of how the documents relate to each other	Lisa's comprehension of the magazine article; her awareness of the respective authority of a high school student vs. a research institute on the topic; her understanding of where the documents stand on the issue.

(*) Only the most relevant cognitive resources are represented in Table 3, as a complete account would exceed the scope of the present chapter.

NB. Table 4. Dimensions, components and processes of relevance assessment in document-based activities

Dimensions	Components	Processes
Contribution to reader goals	Topicality	Lexical match
		Semantic overlap
	Reliability	Source assessment
		Corroboration
	Usefulness	Goal structure and task product monitoring
Cost of accessing and processing	Access	Number of steps, usability of tools required for access
	Processing	Readability (including prior knowledge requirements)
		Length

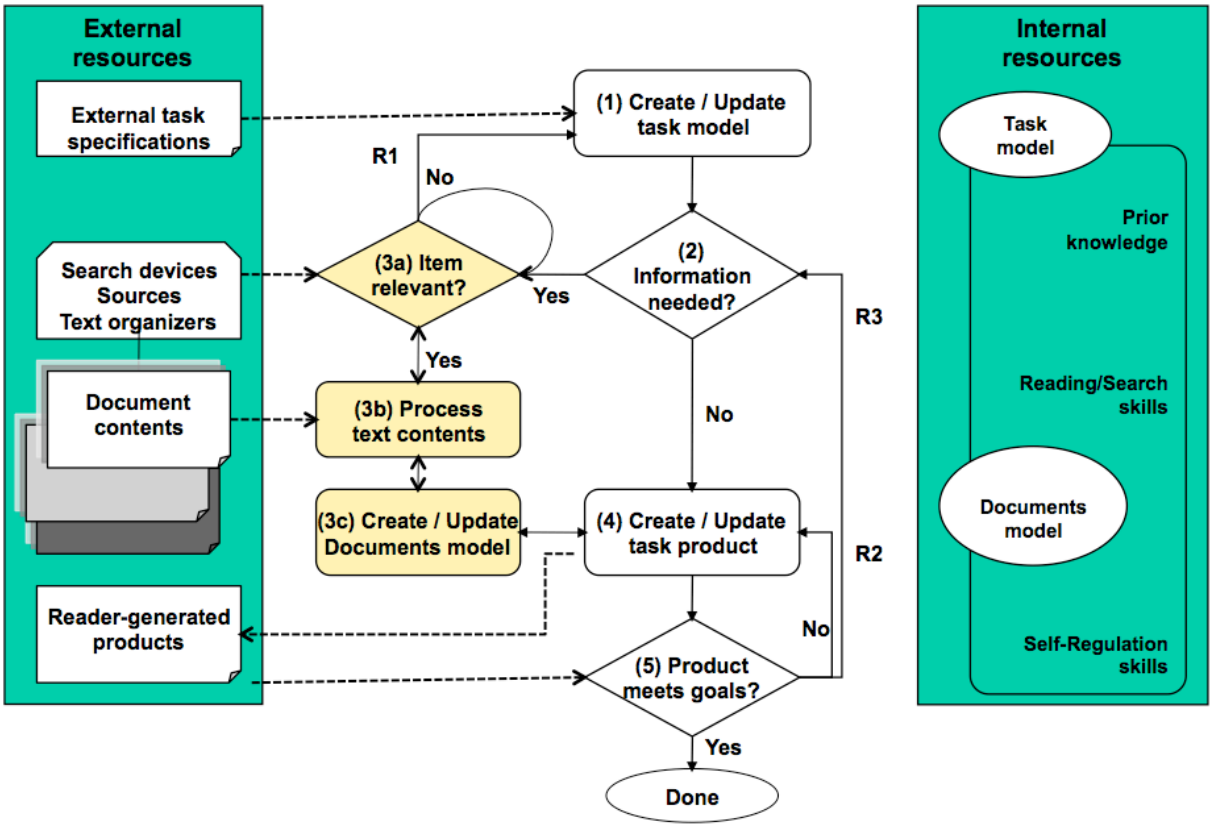

Figure 1. The MD-TRACE model of multiple document use (adapted from Rouet, 2006).