A Decision Tree Approach to Predict the Online Information Search Performance of 8-13 Year Old Students in French-Speaking Belgium

Audrey Kumps
Educational Engineering and Digital Education Department
Assistant
University of Mons, Belgium
audrey.kumps@umons.ac.be
https://orcid.org/0000-0001-6481-0318

Gaetan TEMPERMAN
Educational Engineering and Digital Education Department
Professor
University of Mons, Belgium
https://orcid.org/0000-0002-0200-350X

Bruno DE LIEVRE
Educational Engineering and Digital Education Department
Department Director
University of Mons, Belgium
https://orcid.org/0000-0001-8843-1582

ABSTRACT
In French-speaking Belgium, one of the new tasks of the teacher will be to develop the informational skills of learners. Therefore, the main purpose of this research is to identify the variables that affect the performance of online information retrieval so that teachers can plan appropriate teaching activities. The statistical Classification and Regression Tree (CART) technique was used to characterize the activity of 260 students aged 8 to 13 according to their results in three search tasks. The results show that certain navigational strategies, as well as the different individual characteristics of the participants in our study, can be considered as key factors that modulate performance. The different findings from our analyses allow us to put forward concrete avenues for the mastery of this skill in an educational context.
Keywords: decision tree ; Improving classroom teaching; Information literacy; quantitative analysis ; predictive analysis.

INTRODUCTION
Based on the observation that the digital transition is profoundly modifying the role of the school, a new reference system of digital competences is being defined for pupils aged 3 to 15 in French-speaking Belgium. The plan is for new teaching missions regarding the use of digital technology. One of these new missions will be to develop learners' information skills, their skills in searching for, evaluating and using online information. But do teachers know the specific strategies of their students in order to adapt their pedagogical interventions?

While the comparison between experts and novices in information retrieval is often analyzed, this study aims to bring out the key factors that impact students' success in online searches from a predictive analysis. From this modeling, it is then easier to adapt pedagogical scenarios to guide students in the development of these information skills through a better understanding of the learning process (Sanzana et al., 2015).

In this context, our research focuses on students' individual characteristics and cognitive and metacognitive strategies in three online search tasks. Specifically, we investigate the relationships between comprehension skills, browsing behaviours, Internet use patterns, age, gender, and solving performance on three information search problems for a sample of 260 students aged 8 to 13. Based on all of these variables, we provide evidence to answer the following research question: Are paper-based comprehension skills, students' Internet use patterns, gender and age or actions taken online predictors of online information search performance of students aged 8-13? The multivariate analysis technique CART (Classification and Regression Tree) allows us to analyse the associations between the different variables considered.

In order to better understand the performance (answers to the search problems posed) of learners in this domain, a theoretical review of the literature is carried out in the next section. This review tends to show that the capacity of young people seems limited and that the variables investigated can influence online search performance. We, therefore, take this finding as a premise.

LITERATURE REVIEW

The information presented above conveys our thoughts on understanding studies of students' performance in online information retrieval and the different variables that can influence it.

Students' Performance in Searching for Online Information

An increasing number of students, regardless of social level, are using the Internet to search for information for academic or extracurricular purposes (Wartella et al., 2015). While its use seems straightforward, it brings new difficulties (Mullis et al., 2017) that can disorient and frustrate students. This is especially the case for this "Google generation", born after the 1990s, for whom it is assumed that they are comfortable with this tool (Boubée & Tricot, 2010). The hypertextual structure tends to increase the cognitive demands of decision making, making it difficult to understand the information (DeStefano & LeFevre, 2007). Indeed, contrary to traditional media, the Internet, with its own navigation properties, is dynamic and heterogeneous (lack of structure, different presentation formats, variable quality of available information). Moreover, the reader has to deal with an ever-increasing number of available information sources. This overload can lead to the user getting lost and requiring a significant amount of time to identify the content corresponding to their information needs.

Despite the massive use of information on the Web by primary and secondary school learners (Smahel et al., 2020), it is generally observed that their informational skills are limited (Dumouchel, 2016) and that their search strategies are not very effective (Macedo-Rouet et al., 2019). They are not skilled in defining the search problem (clarifying task needs, activating prior knowledge, etc.), and have difficulties in specifying the terms of the search, in evaluating the results and documents, and in regulating their informational process. Indeed, an Internet user must both retain what they have seen and anticipate what they will see next. This sometimes leads some students to becoming cognitively disoriented, causing them to go in circles, skim pages and thus lose the initial purpose of their research (Djouani et al., 2015). They sometimes tend to navigate, almost randomly, by focusing on the first links presented, producing a search which lacks fluidity and has numerous backtracks without successfully processing the available information (Bilal & Kirby 2002).

In other words, searching for information online represents a significant cognitive cost (Djouani et al., 2015) given that the Internet user must manage both the characteristics of the environment and their search process. It can cause feelings of disorientation and cognitive overload (Macedo-Rouet & Rouet, 2004) when faced with the flow of information.

According to Lazonder and Rouet (2008), three sets of variables can influence an individual's behavior during an information seeking task: contextual variables, resource variables and individual variables. Our study focuses on these three variables as well as on the process that students implement when they perform an online search.

From a performance explanation perspective, our theoretical research focuses on the impact of paper reading skills, Internet use habits, student sex and age, and the navigation strategies used in online searches.

Role of the Determinants of Effective Online Navigation

Reading Skills

Comprehension models established for so-called "traditional" reading do not necessarily account for the wide range of contexts that readers may encounter in a digital text environment (McNamara & Magliano, 2009). Online reading comprehension is therefore not isomorphic to offline reading comprehension (Leu et al., 2015). Indeed, surfing web pages mixing short articles, video, audio, and animations of all kinds is not similar to careful and deep reading, which one practices on a printed medium (Baccino, 2011). It requires different skills than those used offline, for example the use of search engines, the mastery of hypertext links or the ability to evaluate the reliability of information. Consequently, structural and functional knowledge (Rouet, 2012) additional to paper reading is required to perceive the overall coherence of digital documents.

Moreover, the construction of a mental model from a continuous or discontinuous text requires a different cognitive load. The learner reading a continuous text can remember the content better and is able to draw more conclusions than with a discontinuous text (Schnotz, 1988). In a non-linear presentation, all the information must be mentally restructured in order to build overall coherence. The interconnectedness, interactivity, and flexibility of Internet documents requires a more complex comprehension process which requires additional cognitive effort (Brand-Gruwel et al., 2009).

For online reading to be successful, however, students must search for information that meets their needs, process that information, and connect it to make sense of it (Boucheix et al., 2016). Therefore, it is important to implement "traditional" reading comprehension strategies such as: defining the purpose of reading, checking for difficult words, asking questions about the content, inferring, identifying ideas, summarizing, using non-textual information to construct meaning, scanning the text, and processing the information read in order to build knowledge and grasp meaning (Brand-Gruwel et al., 2009). These skills facilitate the selection of relevant search results based on content rather than superficial keywords (Rouet et al., 2011) and help to avoid getting lost in hyperspace (Salmerón et al.,18). In addition, the ability to understand simple texts (decode and infer) facilitates the localization of relevant information in digital texts, both with and without hyperlinks (Coiro & Dobler, 2007).

Different research shows that the level of paper reading skills is positively correlated with digital reading skills. Thus, students with good single-text reading skills are not only better able to navigate using a consistency lens but are also less distracted by misleading cues (Naumann & Salmerón, 2016). Online readers must therefore juggle different reading strategies: selective reading/full reading; skimming/scanning (Schillings & André, 2020) to use elaborate strategies. The online reader thus relies on metatextual cues to meet their information needs.

Computer Skills

Computer skills are also a relevant factor when searching the web. These involve, among other things, keyboarding skills and navigation skills (Kafai & Bates, 1997). Knowledge of how a search engine works, or the structure of a website, is also relevant to explain difficulties (such as disorientation) with information-seeking activities (Brand-Gruwel et al., 2009).

Individual differences in computer literacy partly explain success on digital reading tasks, even after controlling for the effect of other variables, such as reading skills or navigation efficiency (Hahnel et al., 2016).

In addition, Internet use habits also play a role. As a result, using the Internet to conduct literature searches exerts a positive effect on the effectiveness of an online search. In contrast, using the Internet every day has no impact (ePirls, 2016). Therefore, it is not the amount of time spent online that matters, but how one uses the means of accessing information (Schillings & André, 2020).

Browsing Behaviors

Effective navigation can be defined as the ability to stay within a sequence of relevant pages based on the reader's purpose (Sullivan, 2015). In addition, more linear sequencing, and thus less retrograde (backtracking) behavior, produces more systematic navigation behavior with fewer orientation problems and higher comprehension scores (Salmerón et al., 2018).

In contrast, learners who spend a lot of time interacting with features of the hypermedia environment, or whose navigation path reveals no logical order, demonstrate poorer performance (Lawless et al., 2002). Consequently, one of the essential components for an effective information search on the Internet is the localization of relevant information. Concretely, this activity translates into the selection, in a reflective manner, of the links proposed by a search engine, of the various topics presented by the table of contents of a website, or of the multiple hypertexts leading from one website to another (Leu et al., 2015). Several strategies are considered to be effective in optimally performing this information location activity: generating coherent keywords (Bilal & Kirby, 2002); inferential thinking about hyperlinks (Coiro & Dobler, 2007); rapid identification of relevant information on a web page (Rouet, 2012). Thus, navigation behaviors, such as focusing on relevant pages and choosing a coherent and linear navigation path, seem to be effective strategies.

Sex

Girls and boys seem to act differently when faced with documents available on the web. Girls tend to be more cautious about the resources they consult (Taylor & Dalal, 2016). They also seem to perform better than boys, as shown by the first international assessment of online reading for 10-year-olds: ePIRLS (2016). As a result, the technological advantages that boys often enjoy do not seem to apply to reading nor to learning with information found online (Mullis et al., 2017).

Age

Vanderschanchtz et al. (2014) argue that a child's online problem-solving skills develop over the course of a lifetime. Indeed, a gap of a few years can make a huge difference in cognitive (intellectual) abilities (Nesset & Large, 2009). These authors suggest that children 12 -13 years of age and older can be effective problem solvers, and therefore younger children need extra help. In addition, the gradual shift from childhood to adulthood is accompanied by changes in how the epistemic value of sources is assessed. Older children use a greater diversity of criteria and pay more attention to questions of credibility and authority as a result of their social and cognitive development and advancement in schooling. Nevertheless, when students' information skills practices are analyzed, few differences between primary and secondary students are identified (Beheshti, 2012).

In view of these different empirical data, we can hypothesize that the selected variables (individual characteristics, paper reading performance, information search performance, number of strategies used during the search, the time given to each stage of the search, reported practices on Internet use habits) impact the effectiveness of online searches and thus can be predictive factors of performance. The purpose of our study is to examine the interactions between these different variables in explaining the degree of effectiveness in online information searches. We define search efficiency by the accuracy and completeness of the answers given to the proposed tasks.

METHODOLOGY

In the following text, we describe the methodology that was used to test our initial hypothesis and thus to bring out the explanatory variables of the performance of 8- to 13-year-old students during three online information search tasks. We specify the choice of statistical analysis, the sample considered, the tasks proposed, and the variables observed.

Statistical Analysis

The multivariate analysis technique CART (Classification and Regression Tree) allowed us to analyze the associations between the different variables considered. This statistical method builds predictors with a regression tree. It is based on a binary division of groups successively on the basis of a statistical criterion. Starting with the total sample, each variable is examined to determine how well it divides students into two groups. Each node is then divided using the same procedure. As the process progresses, students are divided into smaller and smaller nodes. The similarity of the outcome measured within each node increases, and at the same time, the difference in measurement between nodes also increases. The nodes that cannot be divided are the terminal nodes and these represent groups of students with common characteristics. As a result, each of these groups can fully describe the characteristics of the individuals and the CART procedure defines an estimated mean of the dependent variable for them (Schiattino & Silva, 2008).

In order to interpret the results obtained through the CART analysis technique, our groups (student profiles), according to each of the tasks proposed, are presented in the analysis of our results. The emergence of the predictor variables according to the three tasks performed is presented in our discussion.

Participants

Our sample consisted of 260 native French-speaking students in French-speaking Belgium. Ninety-six of them (43 girls) were in primary 3 or 4; 88 students (42 girls) in primary 5 or 6; 76 (30 girls) in secondary 1 or 2. Our participants were non-repeaters, without any particular disorder, from schools located in French-speaking Belgium. None of the teachers reported organizing specific online information-seeking activities with their students in their classrooms.

Instrumentalization and Data Collection

Students participated individually in two sessions of about 30 min. The data collection was conducted in two phases: (1) a paper reading comprehension test and a questionnaire about their Internet use habits, and (2) the completion of three online information search tasks.

Paper Reading Comprehension Test

This test assessed narrative comprehension to cover reading processes, such as inference. It consisted of a text and six questions: three questions assessing literal comprehension (score out of 4 for each question) and three questions assessing inferential comprehension (score out of 3 for each question). This protocol was applied the day before the assessment of their information seeking skills and lasted 30 minutes. The same story was proposed to our three age groups (1: 8-9 years old; 2: 10-11 years old; 3: 12-13 years old), but the text was longer for each progressive age group.
In order to verify that our test was a tool capable of consistently measuring the abilities of the subjects in our sample, we paid attention to its internal consistency. Analysis of the reliabilitý of the test revealed that Cronbach's alpha was equivalent to 0.705. Hogan (2007) recommends that this should be between 0.7 and 0.9 to judge it as satisfactory. We can therefore consider this comprehension measurement instrument as reliable, i.e., it has the capacity to measure, in a coherent way, the comprehension skills of the subjects in our sample.

Self-reported Practices on Internet Usage Habits

The purpose of this data collection was to obtain students' reported practices regarding Internet use at home, as well as in school settings. It consisted of six items, such as "I use the Internet to watch videos", and the students were asked to rate themselves on a 10-point Likert scale ranging from 0 (meaning "Never") to 10 (for "Every day"). Students completed this questionnaire immediately after completing the "paper" reading comprehension test.

Online Information Search Tasks

These activities were carried out on a tablet application that we created, "Schoolgle", with the same characteristics as the Google search engine. The recording of each student's tablet screen, as well as an observation grid, allowed us to highlight the frequencies of appearance of the strategies used, as well as the time allocated to the different stages of the search.

These different steps were based on the competency-based model "Information Problem Solving while using Internet [IPS-I]" (Brand-Gruwel et al., 2009). This model defines five components: (1) The definition of problem solving begins with the recognition of an information need. It includes the tasks of reading the problem, formulating questions, activating prior knowledge in memory, clarifying requirements, and determining what information is needed. (2) Information retrieval where the user must select a search engine, a search strategy, specify the terms of the query and evaluate the results. (3) Information scanning where users scan the information, evaluate it and retain the relevant information. (4) Information processing where information is read in detail, evaluated and retained to develop content. (5) Information organization for which the users formulate the problem, structure the relevant information and describe the product, realize it and develop content. As online reading is seen as iterative and metacognitive, the IPS-I model also foresees four regulating activities during the research process: orientation, monitoring, steering and evaluation. More precisely, regulation intervenes to manage the task achievements according to the material and temporal conditions; manage and adjust one's information search and analysis behaviors; evaluate the credibility, recency, relevance of the retrieved information and evaluate the product resulting from one's search after having processed the information. These different strategies were translated into observable behaviors and integrated into our observation grid in order to highlight the frequency and time of occurrence

The online information retrieval tasks included three questions related to three different themes; namely "Georges Rémi", "Nosebleeds" and "Dahu". After having produced a query in order to answer a given question, the students were asked to select one or several references among those predefined by the experimenter and to process the information. The entire search procedure was recorded via the tablet. There was no set time given to complete the activity. A score from 0 to 6 was given for each theme. One point was given for each expected answer. Therefore, zero corresponds to no expected answer or no answer. A score of 1 for a single response corresponds to the expected answer, and so on.

According to Dumouchel (2016), there are different levels of difficulty in information retrieval tasks. Therefore, our three activities were of progressive complexity (Table 1). Task 1 "Georges Rémi" was considered to be of low complexity, since it did not require information transformation. The students had to find the requested information directly via the consulted link. Task 2 "Nosebleeds" was considered to be of "moderate complexity" since it required a summary of the information. Not everything that is read is important to meet the information need. Nevertheless, the comparison of several sources enriches the answer. The last task proposed (Task 3: "Dahu") was the most complex task. It required that the information identified be evaluated, transformed and confronted in order to correctly answer the question asked.

Table 1 : The three tasks according to their level of complexity

TASK	LEVEL OF COMPLEXITY
Who is Georges Rémi? (Pseudonym, Profession, Date of birth, Known works, Nationality)	Low complexity Locate and collect information
Your school has decided to set up a brand-new project "Learning to give first aid". Your class chooses to work on nosebleeds and to write a factsheet on the correct actions to take. How should you intervene?	Moderate complexity Locate and collect information Articulate and summarize content
This morning, we heard about a strange animal: the Dahu. Having never had the chance to see one, we decide to go and find it. It is possible to see them in the French mountains. Draw a cross on the map below to show the different places where we must go to see this animal.	High complexity Locate and collect information Articulate and summarize the content Confront, evaluate, interpret ideas.

Variables Considered and Research Question

As previously stated, to address our goal of predicting the performance of 8- to 13-year-old students on three online information search challenges, we considered several types of data. Thus, nominal data (individual characteristic: sex), metric data (paper reading performance, information search performance, number of strategies used during the search, time given to each step of the search, reported practices on Internet use habits) and ordinal data (individual characteristic: age group) were considered and combined through CART analysis. Table 2 shows all of the independent and dependent variables taken into account to predict the scores on the three tasks performed.

Table 2 : Dependent and independent variables

INDEPENDENT VARIABLES	DEPENDENT VARIABLES
Personal characteristics Sex: Female (1) - Male (2) Age: Group 1 (8-9 years old) - Group 2 (10-11 years old) - Group 3 (11-12 years old) Reading test Explicit test score: score from 0 to 4 Implicit test score: score from 0 to 3 Declared practices Use of the Internet to watch videos: Score from 0 to 10 Use of the Internet to search for information: Score from 0 to 10 Use of the Internet to write messages/emails: Score from 0 to 10 Use of the internet to play games: Score from 0 to 10 Use of the Internet to do homework: Score from 0 to 10 Use of the Internet to search for a definition: Score from 0 to 10 Online behavior (time in seconds) Time to define the problem Time to search for information Time given to scanning the information Time given to processing the information Time for organizing information Online behavior (frequency of occurrence) Number of times the task is read Number of keywords defined Number of phrases defined as search query Number of questions defined as search query Number of spelling mistakes made in queries Number of links chosen based on heuristic indices Number of Wikipedia links chosen Number of relevant links chosen Number of irrelevant links chosen Number of advertising links chosen Number of forum links chosen Number of "image" or "video" links consulted Number of backtracking performed when searching for information Number of page starts scanned Number of pages fully scanned Number of backtracking performed when scanning information Number of pages processed Number of returns to a previously viewed link	Research task score Task 1: score from 0 to 6 6 answers to be given - one point per correct answer Task 2: score from 0 to 6 6 answers to give - one point for each correct answer Task 3: score from 0 to 6 4 possible answers - 0, 2, 4 or 6 points awarded

Number of backspaces performed while processing information
Number of answers to the problem with word by word copying of the information
Number of answers to the problem with reformulation of the information
Number of backtracking operations performed when organizing the information

Based on the set of variables presented above, we will provide elements of an answer to the following research question: are paper-based reading comprehension skills, students' internet usage habits, sex and age or actions taken online predictors of the online information retrieval performance of students aged 8 to 13?

Protection of personal data
To ensure ethical research practice that protects the safety, privacy and confidentiality of our participants, we have anonymised the names of participants in each activity by replacing them with a number.

ANALYSIS OF RESULTS

Online Information Search Performance
The average score achieved by the 260 learners on the first task, which required reading information with no need to transform it to answer the question posed (low difficulty), was 3.84 points out of 6 (64% average) with a standard deviation of 2.018 points. Twenty-five percent of the students (65 students) did not pass the task (i.e., they did not get 3/6) and 23.07% (60 students) achieved the maximum score. It was also noted that 31 students (11.92% of the sample) achieved a score of zero.

When the information required confronting and summarizing to answer Task 2, the average dropped to 1.58 out of 6 (26.3% average) with a standard deviation of 1.72 points. One hundred and eighty-two students (70% of the sample) did not achieve the average (3 points out of 6), including 112 who had a zero score. Only six students (2.31%) managed to get 6/6.

Finally, for the last task where the information found must be evaluated, transformed and confronted in order to correctly answer the question, our sample obtained the lowest average: 1.44 (24% average) with a standard deviation of 1.37. For this challenge, 93.46% (243 students) did not pass the activity, and only 3.46% (nine students) achieved a perfect score. Ninety-eight students (10.27%) scored zero out of six.

Explanatory Factors
Relationships between the Dependent and Independent Variables
To meet our objective, we first calculated the strength of the relationships between our dependent and independent variables using Spearman's correlation coefficient.

Table 3 allows us to highlight several interesting elements. First, only the three variables "age", "reading test scores" (explicit, implicit) and "intensity of Internet use for homework" were positively correlated with the scores on the three proposed tasks (highlighted in grey in Table 3). Some variables, on the other hand, were not correlated at all with performance on any of the tasks. This was the case for the variables: "intensity of internet use for information search", "number of sentences defined as search query", "number of spelling mistakes made in queries", "number of links chosen based on heuristic cues", "number of Wikipedia links", "number of advertising links chosen", "number of backtracking performed when organizing information" and "number of backtracking to a previously viewed link".

Then, some online search behaviors also showed significant correlations to the score obtained on one of the three challenges. Thus, for the low complexity task, 15 variables, such as "number of relevant links opened", "time spent organizing the answer to the question asked", etc., were significantly correlated to the score obtained. We also noticed that, for six variables, including "number of irrelevant links opened", "time given to the search for information", and "number of keywords defined", the correlation coefficient was negative, showing us a simultaneous variation in the opposite direction between the performance score and the variable considered.

For Task 2 (moderate complexity), 20 variables were significantly correlated to the score, including "sex", "number of links (relevant and forums) opened", "number of backtracking performed when scanning and processing information", and "number of pages read in depth". Only the variables "sex" and "use of the Internet to watch videos" showed a negative relationship

Finally, for the last challenge that required verification, confrontation and reformulation of information, 13 variables (including "number of readings of the problem posed", "number of pages fully scanned" and "number of pages fully scanned") were seen as significantly correlated with performance. For these variables, the correlation coefficients were positive except for "number of selected forum links".

Table 3 :The strength of the relationships between the dependent and independent variables (Spearman correlation)

INDEPENDENT VARIABLES	DEPENDENT VARIABLES		
	Score Task 1	*Score Task 2*	*Score Task 3*
PERSONAL CHARACTERISTICS			
Sex	0.063	-0.136	0.014
Age	0.409***	0.337***	0.290***
TEST READING			
Explicit test score	0.360***	0.278***	0.228***
Implicit test score	0.340***	0.214***	0.291***
REPORTED PRACTICES			
Use of internet to watch videos	-0.139*	0.142*	0.024
Use of internet to search for information	0.049	0.060	0.054
Use of the internet to write messages/emails	0.129*	0.088	-0.011
Use of internet to games	-0.157*	-0.096	-0.023
Use of the internet to do schoolwork	0.265***	0.212***	0.190**
Use of the internet to search for a definition	0.124*	0.138*	0.070
ONLINE BEHAVIOR (actions)			
Number of times the task was read	-0.035	0.118	0.126*
Number of defined keywords	-0.137*	0.091	0.261***
Number of sentences defined as search query	-0.092	0.073	0.031
Number of questions defined as search query	-0.092	0.130*	-0.103
Number of spelling mistakes made in the queries	-0.057	0.045	0.030
Number of links chosen according to heuristic indices	-0.003	0.060	0.044
Number of selected Wikipedia links	0.063	0.098	0.092
Number of relevant links selected	0.234***	0.431***	0.097
Number of irrelevant links selected	-0.201**	-0.036	-0.00
Number of selected advertising links	-0.099	-0.076	-0.057
Number of selected forum links	-0.012	0.249***	-0.132*
Number of "image" or "video" links consulted	-0.032	0.096	0.261**
Number of backtracking performed during the search for information	-0.103	0.069	0.130*
Number of scanned page starts	-0.068	0.142*	0.063
Number of fully scanned pages	-0.027	0.184**	0.212***
Number of backtracking performed when scanning the information	-0.066	0.202***	0.133*
Number of pages processed	0.136*	0.309***	0.056
Number of returns to a previously viewed link	-0.001	0.057	0.055
Number of backtracking performed during the processing of the information	0.089	0.244***	0.051
Number of answers to the problem with word-to-word copying of the information	-0.034	0.189***	0.094
Number of answers to the problem with reformulation of the information	-0.034	0.195**	0.056

Number of backtracking performed when organizing information	0.081	-0.053	-0.095
ONLINE BEHAVIOR (time)			
Time given to the definition of the problem	-0.239***	0.048	0.046
Time given to search for information	-0.348***	0.062	0.070
Time to scan information	-0.010	0.153*	0.178**
Time given to information processing	0.087	0.426***	0.174
Time given to the organization of information	0.194**	0.307***	0.201**

*. p ≤ 0,05.
**. p ≤ 0,01.
***. p ≤ 0,001.

Selection of Variables
After highlighting the variables that were significantly correlated with the challenge score considered, we identified the correlations that can be qualified with an intensity said to be at least average, i.e. greater than 0.2 (Cohen, 1988), in order to discard the almost null relationships. Then, for each data set, we removed the variables correlated with each other, keeping the one where the Spearman's Rho was of the highest intensity. Indeed, the CART analysis procedure being a regression model, it obeys the same conditions of application. As a result, only variables that were not correlated with each other were used to construct the regression trees.

In order to explain our approach to variable selection, we can illustrate it with the help of the example developed below (Table 4): the collection of data on performance on the paper reading comprehension test. This includes the score of the comprehension of explicit and implicit information for the three requested tasks.

A) The six correlation coefficients between the score on a task and the score on the explicit and implicit test were significant and greater than 0.2 (columns 3, 4 and 5 of Table 4). All six were therefore likely to be introduced into the regression model. B) As the explicit and implicit test scores were derived from the same measurement, we checked that they were not correlated. Applying a Spearman correlation showed us that these two variables could be seen as correlated (rs = 0.242: p < .001: column 1) and therefore could not both be entered into the CART procedure. We therefore had to select one of the two per task. C) For Task 1, we kept the explicit test score with respect to the intensity of Spearman's Rho coefficient with the dependent variable (explicit rs= 0.360 vs implicit rs=0.340). In the same way, for Task 2, it was also the score of the explicit test that was retained (rs explicit = 0.278 vs rs implicit = 0.214). On the other hand, for Task 3, it was the implicit test score that was used for the CART procedure (rs explicit = 0.228 vs rs implicit = 0.291).

Table 4 : Variable selection process for the "paper reading comprehension test" data collection

	Paper Reading Comprehension Test			
	Implicit test score	Score Task 1	Score Task 2	Score Task 3
Explicit test score	0.242***	*0.360****	*0.278****	0.228***
Implicit test score		0.340***	0.214***	*0.291****

NOTE. The items in italics were chosen to be introduced into the CART given the higher Spearman's Rho intensity
***. p ≤ 0,001.

Following this selection approach for each data collection, five variables were selected to predict performance on Task 1: "age", "score on the explicit reading test", "reported practices: use of the Internet to do schoolwork", "number of relevant links chosen" and "time spent searching for information".

For Task 2, six variables were selected as possible predictors of the score obtained: "age", "score in the explicit reading test", "declared practices: use of the Internet to do schoolwork", "number of relevant links chosen", "number of backtracking operations carried out when scanning the information" and "time spent processing the information".

Finally, for the last task, six variables were taken into account: "age", "implicit reading test score", "number of keywords defined", "number of image or video media consulted", "number of pages fully scanned" and "time given to the organization of information".

CART Procedure: Regression Tree Analysis

From the selected variables, a CART procedure was applied giving a unique regression tree for each task performed. The reading of the tree obtained for Task 1 is fully explained below. For the trees obtained for the two other information retrieval tasks performed by our sample, the variables that allow our subjects, and the student profile that obtains the best performance, to be classified are specified.

Task 1: Low Complexity - Locate and Retrieve Information

"Who is Georges Rémi? (Pseudonym, Profession, Date of birth, Known works, Nationality)"

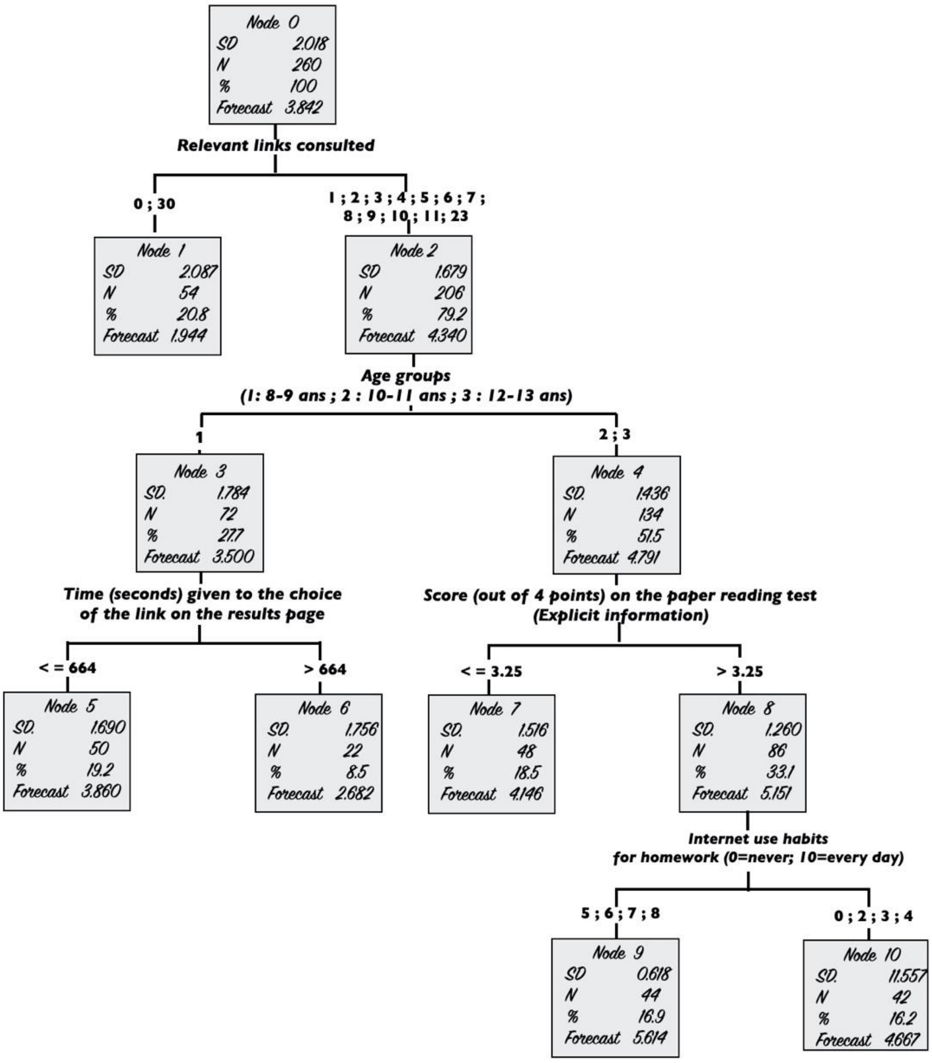

Figure 1: Decision tree obtained with the CART procedure for Task 1

For Task 1, the root node groups the 260 students who obtained an average of 3.84 points out of 6. This was divided by the variable "number of relevant links accessed". The node on the left "0 or 30 relevant links selected" has 54 students with a predicted score of 1.944. This node is a terminal node called Group 1. The right node "1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or 23 links chosen" has 206 students with an average score of 4.340. This first division shows us that not viewing any relevant links, or viewing 30 relevant links, is less effective in terms of performance than choosing 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ,11 or 23 links.

This right node was subdivided by the variable "age". Node 3 includes 72 students of "age 1 (8-9 years)" with an average of 3.5, and Node 4 contains 134 students of "ages 2 and 3 (10-13 years)" with an average of 4.791. Thus, for students who selected 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ,11 or 23 relevant links, age can also predict the score on information retrieval Task 1. The 8-9-year-old students performed worse than 10-13-year-old learners.

Node 3 gives rise to two terminal nodes: Groups 2 and 3. This Node 3 was split by the variable "time spent on information retrieval". Thus, the 50 students who took 664 seconds or less to enter their query and choose a link to consult are in Group 2 (with an average prediction of 3.86) and the 22 others who did so in more than 664 seconds form Group 3 (2.682 average). The performance for these groups increased when the time allowed was less than or equal to 664 seconds (just over 11 minutes).

Node 4, which includes the 134 students aged 10-13, is split in two by the "explicit reading test score" variable. Forty-eight of these students scored 3.25 or less out of a possible 4.5 on the explicit reading test. These students are in a terminal node with an average of 4.146 and form Group 4. The remaining 86 students (Node 8), who scored more than 3.25 on the explicit reading test, have a mean of 5.151.

Node 8 is the origin of the last two terminal nodes. It is split in two by the variable "reported intensity of Internet use for homework". On one side are those who assigned a value of 5, 6, 7 or 8 to this variable (Node 9) and on the other are those who assigned 0, 2, 3 or 4 (Node 10). The students in Node 9 make up Group 5 with 44 subjects and an average of 5.614. There are 42 students in Node 10 and they have an average of 4.667.

With the terminal nodes, we can identify six groups/profiles of students:
Group 1: students with the lowest average score of 1.944 out of 6 on the information retrieval test. This group is characterized by students who consulted 0 or 30 relevant links during their search. It represents 20.8% of the sample.

Group 2: Students who scored 3.86 out of 6. This group is defined by students who consulted 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ,11 or 23 relevant links, were 8-9 years old, and needed 664 seconds or less to enter their query into the search engine and to choose the link(s) to consult. This group represents 19.2% of the sample.

Group 3: students who scored 2.682 out of 6. This group of students has the same characteristics as Group 2, but these students spent more than 664 seconds searching for information. This group represents 8.5% of the sample and is the smallest of the six profiles identified.

Group 4: Students scoring 4.146 out of 6. This group is determined by students who consulted 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ,11 or 23 relevant links, were aged 10 to 13 and scored 3.25 or less out of 4.5 on the explicit reading test. This group represents 18.5% of the sample.

Group 5: Students scoring 5.614 out of 6. This represents the profile with the highest performance on the information retrieval test. Students in this group consulted 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or 23 relevant links, were between 10 and 13 years old and scored more than 3.25 out of 4.5 on the explicit reading test. In addition, they reported using the Internet for homework moderately or frequently. This group represents 16.9% of the sample.

Group 6: Students who scored 4.667 out of 6. This group of students has the same characteristics as Group 5. However, this group reported spending little or no time on the Internet to complete their assignments. This group represents 16.2% of the sample.

Task 2: Moderate complexity - Locate and retrieve information - Articulate and summarize content
Your school has decided to set up a brand-new "Learning to give first aid" project. Your class chooses to work on nosebleeds and to write a factsheet on the correct actions to take. How should you intervene?

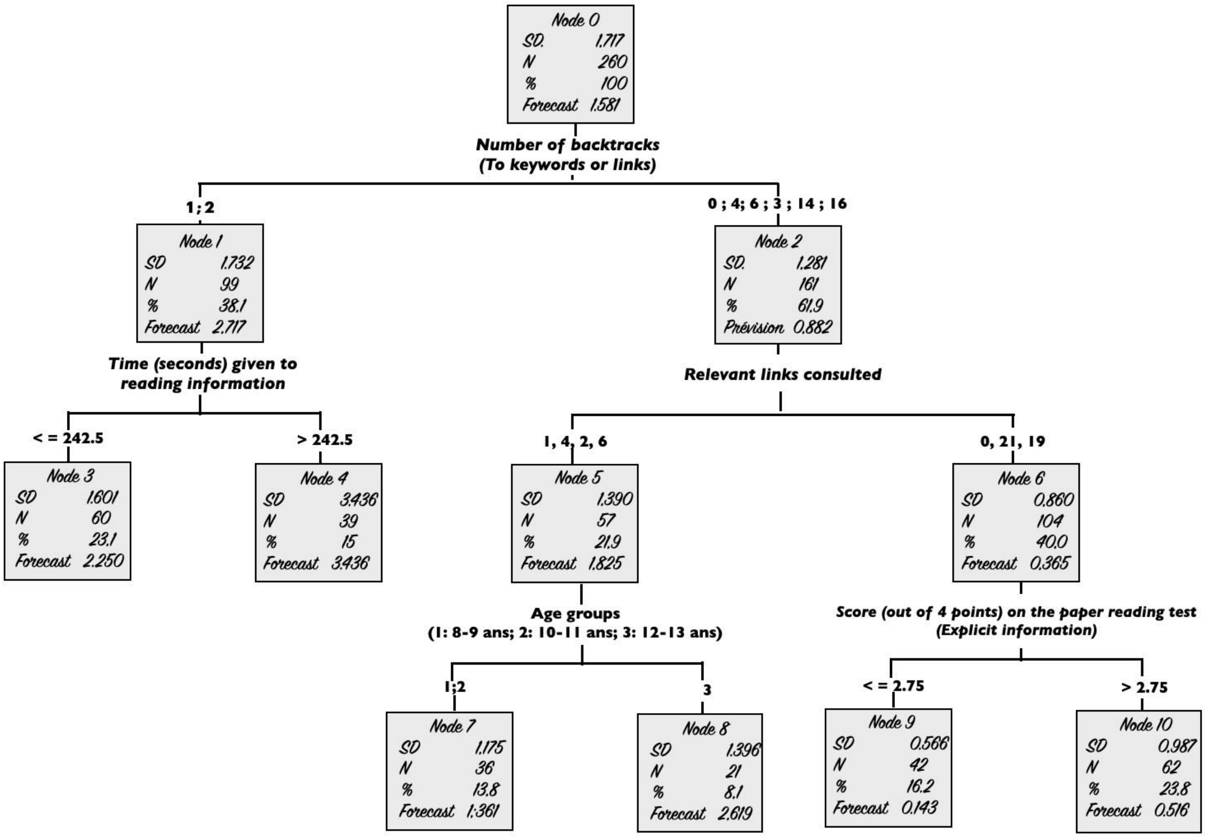

Figure 2: Decision tree obtained with the CART procedure for Task 2

Following the same reading of the tree as for Task 1, we can see that the first variable that divides our sample in two is the number of times the student returned to the keywords and links after scanning the information. As a result, students who did not go back at all, or who went back 3, 4 , 6, 14, or 16 times performed worse than those who went back only 1 or 2 times.

The time spent processing the information and the number of relevant links consulted distinguishes the students in a second step. For learners who had 1 or 2 returns to keywords and links after scanning the information, allowing more than 242.5 seconds (about 4 minutes) for in-depth processing of the information found, had a better performance than those who allowed less time. For students who did not return to the keywords or links, or who returned 3, 4, 6, 14 or 16 times, the number of links chosen influenced their score. Thus, those who visit 0, 19 or 21 links will have a lower score than those who visit 1, 2, 4 or 6 links.

The very last separation is made with the variables "age" and "explicit reading test score". The "age" variable had an impact on performance for the group of students who either did not go back after scanning the information or went back 3, 4, 6, 14, or 16 times and consulted 1, 2, 4, or 6 relevant links in their search. Students aged 12-13 performed better than those aged 8-11. The variable "score on the explicit reading test" reflects the score of students who either did not go back after scanning the information or who went back 3, 4, 6, 14 or 16 times; who consulted no relevant links or, on the contrary, consulted a lot of links (19 or 21 links). Learners who scored above 2.45 out of 4 on the explicit information comprehension test performed best.

This CART tree for performance on the second task produces six terminal nodes. These nodes allow us to define six profiles/groups of students to predict resolution of an online information retrieval task.

Only one of these groups scored above average on Task 2. This group includes students who scored 3.436 out of 6 on this online search task. It is characterized by students who went back 1 or 2 times after scanning the selected information and who spent more than 242.5 seconds reading the information. This group represents 15% of the sample.

Task 3: High complexity - Locate and retrieve information - Articulate and summarize content - Confront, evaluate, interpret ideas.

This morning, we heard about a strange animal: the Dahu. Having never had the chance to see one, we decided to go and find it. It is possible to see them it in the French mountains. Draw a cross on the map below to show the different places we must go to see this animal.

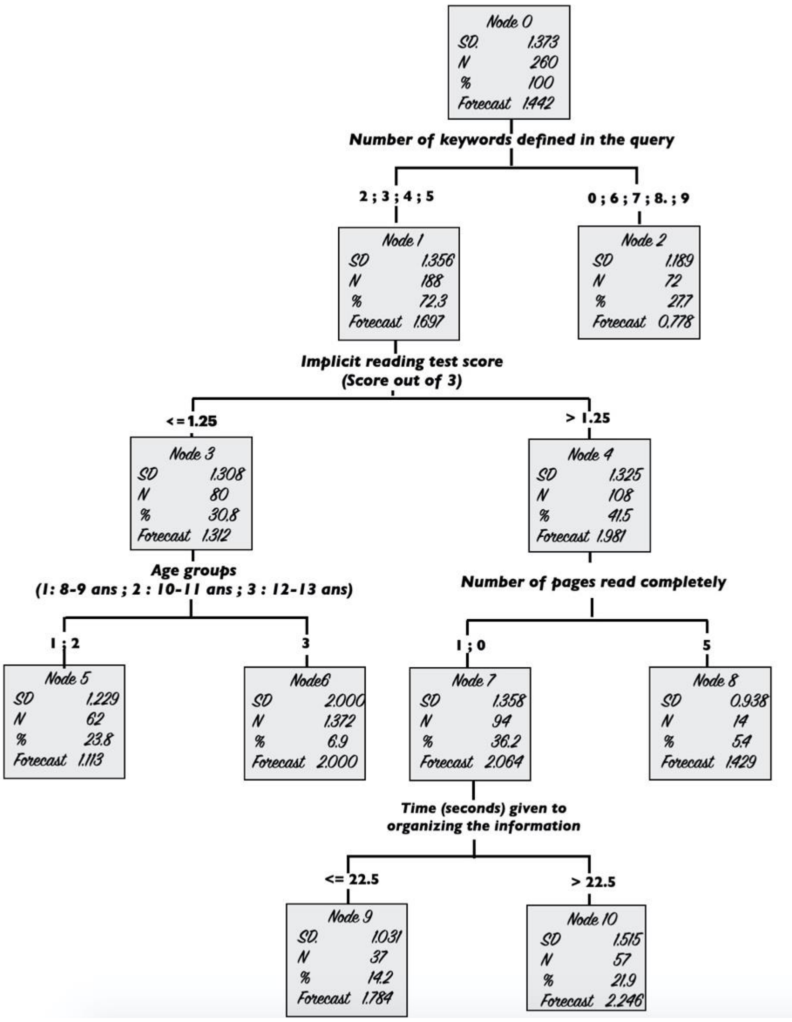

Figure 3: Decision tree obtained with the CART procedure for Task 3

As for Tasks 1 and 2, the CART tree shows us the variables that subdivide our sample and make groups/profiles of students emerge. For Task 3, there are five variables that subdivide our sample and thus predict performance: "number of keywords defined in the query", "implicit reading test score", "age", "number of pages read completely", "time given to the information processing stage".

The tree shows us that the first division of the sample is done at the level of the variable "number of keywords defined in the query". Thus, students who defined 2, 3, 4 or 5 keywords achieved a better score than those who did not, or than those who defined 6, 7, 8 or 9 keywords. For those who defined fewer keywords, the CART procedure divides them with respect to the implicit comprehension test score. Learners with a score higher than 1.25 out of 3 performed better than those with a score lower than that.

The variable "age" discriminates students who define few keywords and who score less than or equal to 1.25 out of 3 on the implicit comprehension test. As a result, subjects who were 12-13 years old finished Task 3 with a better performance than those who were 8-11 years old.

The variable "number of pages read entirely" differentiates students who defined few keywords but who scored more than 1.25 out of 3 on the implicit comprehension test. The tree shows us that those who read one or no web pages completely have a better predicted score than those who read 5.

The last distinction is revealed in the variable "time given to the information organization stage". It shows the score of students who defined 2, 3, 4 or 5 keywords, had a score higher than 1.25 out of 3 on the implicit comprehension test and who read one or no web pages entirely. Thus, giving more than 22.5 seconds to the organization of information predicts a higher score than those who give a shorter time or equal to 22.5 seconds. Thus, this profile obtains the best score prediction (2.246 out of 6) for our entire sample. It should be noted that for this task, none of the student profiles shown in this tree achieves sufficient score prediction.

DISCUSSION

In order for teachers to implement pedagogical practices that are considered effective, we chose to highlight the interaction between variables identified as predictive of students' performance in searching for information on the Internet through the CART decision tree procedure. To do so, we collected data at the level of individual participant characteristics and at the level of the browsing process recorded via a tablet tool for 260 students aged 8-13 performing three online information search tasks. These three tasks were designed with regard to the findings of Dumouchel (2016) and Fournier (2007) who demonstrate that there are different levels of difficulty in this type of task.

As a result, our three activities were characterized by progressive complexity and different comprehension processes. The first was a task considered of low complexity, as it did not require any transformation of the information found to answer the proposed question (finding and retrieving information). The second task was of "moderate complexity" because it required summarizing and comparing information (finding and collecting information, comparing it with other sources and summarizing it). Finally, the last task was considered to be of high complexity because it required that the information found be evaluated, transformed and confronted in order to correctly answer the question asked (retrieve, take information, interpret it, integrate it and examine and evaluate the content).

Our analysis of the results highlights some interesting findings. First of all, in terms of student performance, we can confirm that using the Internet to search for information did not seem easy for students aged 8 to 13 (Dumouchel, 2016; Boudée & Tricot, 2010; Macedo-Rouet et al., 2019). We can also provide a clarification to this finding. While about 23% of our subjects scored less than 50% when they had to retrieve information (Task 1), when it came to summarizing this information and confronting it with other sources (Task 2), 70% failed. And, when it came to evaluating the source of the information and compiling it with other sources (Task 3), 94.5% of the learners did not pass the task. Thus, not surprisingly, it was the task requiring the least complex processes that was the most successful. We can also conclude that the sum of the processes to be put in place in order to respond to the proposed task seemed to be correlated to the difficulties encountered by the students.

Then, when predictive variables could be selected, we can notice that, for our sample, the variables "age", "paper reading skills", "Internet use habits", "browsing behaviors" and "time given to the search steps" are the common explanatory factors for the three proposed tasks. This is consistent with the results of empirical research in the field. However, contrary to the findings of Taylor and Dalal (2014), the international ePIRLS assessment (2016), and Mullis et al. (2017), the variable "sex" did not impact performance. In our experiment, girls did not seem to stand out compared to boys in the different online information-seeking challenges we proposed, unlike in the previously cited research.

Moreover, our analyses allow us to refine the impact of these variables considered as explanatory according to the nature of the tasks presented.

First, only the variable "age" is seen as an explanatory factor for the three tasks requested. This finding can be justified by the evolving nature of a skill (Tardif, 2006). Online problem-solving skills are therefore no exception to the rule, and a difference of a few years can translate into a difference in performance. Our results show that, for an online search task that requires retrieving information, being older than 9 years old has a positive influence on search results. On the other hand, when the information must be summarized, linked to other sources (Task 2), as well as evaluated and confronted (Task 3), it is the students over 11 years old who perform better.

Reading comprehension performance in "paper" reading is also a predictor of success in online information retrieval. Thus, a student with good comprehension skills for explicit "paper" information will have an easier time retrieving isolated information, as well as during tasks where they are asked to summarize and articulate this information with other documents. On the other hand, in order to be effective when it is necessary, in addition, to interpret, integrate, examine and evaluate the content, it is the students who are competent in implicit comprehension who perform best. Our results thus add to the list of studies that show that paper reading skills improve digital reading (Sung et al., 2015).

The variable "reported practices: using the Internet to do schoolwork" categorized students for the first two tasks (low and moderate complexity level). Therefore, the fact that a student is used to using the Internet to complete school activities at home positively impacts their performance.

Time is also a factor in our modeling. We notice that the ability to quickly choose the query to enter into the search engine, as well as the link to consult, has a positive impact on the results for the low complexity task. For the information search Task 2, which requires comparing and summarizing content, it is the time spent on in-depth processing of the information that has an influence. If the reader spends more than 4 minutes on this task, they are more likely to obtain a better result. The time spent organizing the response to the information need also predicts performance. Thus, taking time to organize one's ideas and construct one's answer are important abilities in solving a high complexity task.

Our three CART trees also highlight four navigational strategies as predictive factors. First, the "number of relevant links chosen" differentiated students for the first two tasks (low and moderate complexity level). The ability to identify and consult relevant links during a search therefore positively influences the results. It should be noted that consulting too many relevant links has a negative impact on the task score. We can hypothesize a cognitive overload to explain this finding. Secondly, the "number of backtracking performed while scanning the information" was identified in Tree 2. Thus, being able to realize when scanning the information consulted that it is not relevant to our information need and modifying our choice of link or query appears to be a profitable metacognitive strategy. The variables "number of keywords defined" and "number of pages fully read" also explain the quality of the performance of the complex task: interpreting, integrating, examining and evaluating the information. Therefore, choosing keywords well, and therefore defining few (less than 6), influences the success of students aged 8 to 13. On the other hand, it is not necessary to read all the pages consulted on the web in order to perform well. In fact, the best scores were attributed to students who processed the entirety of only one page or no pages during their search.

CONCLUSIONS

From a methodological point of view, we can highlight that decision tree modeling is a relevant tool for identifying the key variables in a process that explains the quality of task performance. Moreover, taking into account the interaction of these variables can allow the design of contextualized interventions according to the student's profile. Thus, while it can be used, CART prediction seems to be an effective analysis strategy for the field of education.

Thus, our different findings allow us to make several recommendations to both teachers and educational policy makers responsible for the new reference materials for teaching online information search skills.

It is crucial to make the link between paper-based and digital comprehension strategies visible. Providing specific activities for skills teaching in the classroom avoids a patchwork of (home) learning. In addition, teachers need to plan and break down the teaching of online research, taking into account that it is acquired gradually. They must also focus on learning about the structure of different web pages in order to perceive the reading of them as well as the organization of the information (and avoid reading the whole thing slowly), and the organization and presentation of the information found. This is important since these variables are seen as predictors of the performance of students' searches.

REFERENCES

Baccino, T. (2011). Lire sur internet, est-ce toujours lire ? », *Bulletin des Bibliiothèsques de France*, 5, 63-66

Beheshti, J. (2012). Teens, virtual environments, and information literacy. *Bulletin of the American Society for Information Science and Technology*, *38*(3), 54-57.

Bilal D., & Kirby J. (2002). Differences and similarities in information seeking: Children and adults as Web users. *Information Processing & Management*, 38(5), 649–670.

Brand-Gruwel S., Wopereis I., & Walraven A. (2009). A descriptive model of information problem solving while using internet, *Computers & Education*, 53, 1207-1217.

Boucheix, J.M., Amadieu, F., & Tricot, A. (2016). Les technologies numériques au service des apprentissages, In *Cerveau et apprentissage, Dossier "Les sciences cognitives et l'éducation"*, Fondation la main à la pâte

Boubée, N., & Tricot, A, (2010) *Qu'est-ce que rechercher de l'information ?* Villeurbanne : Presses de l'Enssib

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Coiro, J., & Dobler, E. (2007). Exploring the online reading compre- hension strategies used by sixth-grade skilled readers to search for and locate information on the internet. *Reading Research Quarterly*, *42*(2), 214–257. https://doi.org/10.1598/RRQ.42.2.2

DeStefano, D., & LeFevre, J.-A. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior, 23*(3), 1616–1641. https://doi.org/10.1016/j.chb.2005.08.012

Djouani, M., Caro, S., Boucheix, J.-M. (2015). Un dispositif de prévisualisation qui améliore la navigation : comparaison entre une tablette tactile et une souris 3 D. *RIHM Revue des interactions humaines médiatisées*, 16 (1).

Dumouchel, G. (2016). *Les compétences informationnelles des futurs enseignants du Québec sur le Web* (Thèse de doctorat, Université de Montréal). Thèses.fr.

Fournier, H. (2007). *Stratégies de recherche et de traitement de l'information dans des environnements informatiques et sentiment d'efficacité personnelle des futurs enseignants à l'égard de ces stratégies* (Thèse de doctorat non publiée). Université du Québec, Montréal.

Hahnel, C., Goldhammer, F., Naumann, J., & Kröhne, U. (2016). Effects of linear reading, basic computer skills, evaluating online information, and navigation on reading digital text. *Computers in Human Behavior*, 55, 486–500. https://doi.org/10.1016/j.chb.2015.09.042

Hogan, T. P. (2007). *Psychological Testing: a practical introduction* (2e éd.). New-Jersey : Wiley.

Kafai Y., Bates M. (1997). Internet Web-searching instruction in the elementary classroom: Building a foundation for information literacy. *School Library Media Quarterly*, 25(2), 103–111.

Lazonder, A. W. & Rouet, J. F. (2008). Information problem solving instruction: Some cognitive and metacognitive issues. *Computers in Human Behavior*, 24, 753–765.

Lawless, K. A., Mills, R., & Brown, S. W. (2002). Children's hypermedia navi-gational strategies. *Journal of Research on Computing in Education*, 34(3), 274-284

Leu D. J., Forzani E., Rhoads C., Maykel C., Kennedy C., & Timbrell N. (2015). The new literacies of online research and comprehension: Rethinking the reading achievement gap. *Reading Research Quarterly*, *50*(1), 37–59.

Macedo-Rouet M., Potocki A., Scharrer L., Ros C., Stadtler M., & Salmerón, L. (2019). How good is this page? Benefits and limits of prompting on adolescents' evaluation of Web information quality. *Reading Research Quarterly, 54*(3), 299–321.

Macedo-Rouet, M. & Rouet, J. (2004). Vulgarisation scientifique: les revues en ligne. *Hermès, La Revue*, 2(2), 61-68. https://doi.org/10.4267/2042/9464

McNamara, D. S., & Magliano, J. (2009). *Toward a comprehensive model of comprehension.* In B. H. Ross (Ed.), *The psychology of learning and motivation: Vol. 51. The psychology of learning and motivation* (297–384). Elsevier Academic Press.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 international results in reading*. Boston College, TIMSS & PIRLS International Study Center.

Nesset, V. et Large, A. (2009). Elementary school students, information retrieval, and the Web. In M. Pagani (dir.), *Encyclopedia of multimedia technology and networking* (2e éd., p. 469-476). Hershey, PA: Information Science Reference.

Naumann, J., & Salmerón, L. (2016). Does Navigation Always Predict Performance? Effects of Navigation on Digital Reading are Moderated by Comprehension Skills. *The International Review of Research in Open and Distributed Learning*, *17*(1). https://doi.org/10.19173/irrodl.v17i1.2113

Rouet J.-F. (2012), Ce que l'usage d'internet nous apprend sur la lecture et son apprentissage, *Le français aujourd'hui*, 3 (178), 55-64.

Salmerón, L., Strømsø, H.I, Kammerer, Y., Stadtler, M., & van den Broek, P. (2018). Comprehension processes in digital reading. In M. Barzillai, J. Thomson, S. Schroeder, & P. van den Broek (Eds.), *Learning to read in a digital world*, (91-120). Amsterdam: John Benjamins

Sanzana, M.B, Garrido, S.S & Poblete, C.M (2015). Profiles of Chilean students according to academic performance in mathematics: An exploratory study using classification trees and random forests, *Studies in Educational Evaluation* 44, 50–59

Schiattino, I., & Silva, C. (2008). Arboles de Clasificacio ́n y Regresio ́n: Modelos Cart. *Ciencia & Trabajo*, 10, 161–166.

Schnotz, W. (1988). Instructional implications of text processing research. *Eur J Psychol Educ* 3, 111–121.

Schillings, P. & André, M. (2020). Analyse Clinique des obstacles rencontrés par des élèves de 4e et 5e années primaires face à un texte de lecture numérique. [Communication orale]. Didactifen 2020. Liège.

Smahel D., Machackova H., Mascheroni G., Dedkova L., Staksurd E., Olafsson K., & Liningstone S. (2020). *EU Kids Online 2020. New European study on children and the internet in 19 countries.* http://www.lse.ac.uk/media-and-communications/research/research-projects/eu-kids-online/eu-kids-online-2020

Sullivan, D. (2015, 16 janvier*). Google still doing at least 1 trillion searches per year.* http://searchengineland.com/google-1-trillion-searches-per-year-212940

Sung, Y.-T., Wu, M.-D., Chen, C.-K., & Chang, K.-E. (2015). Examining the online reading behavior and performance of fifth-graders: Evidence from eye-movement data. *Frontiers in Psychology*, 6, 1–15

Tardif, J. (2006). *L'évaluation des compétences : documenter le parcours de développement*. Montréal, QC : Chenelière Éducation.
Taylor, A.& Dalal, H. A. (2014). Information literacy standards and the World Wide Web: Results from a student survey on evaluation of Internet information sources. *Information Research, 19*(4).
Vanderschanchtz, N., Hinze, A. & Cunningham, S.J (2014). Sometimes the internet reads the question wrong: Children's search strategies & difficulties. *Proceedings of the American Society for Information Science and Technology, 51(1).* DOI: 10.1002/meet.2014.14505101053
Wartella, E., Rideout, V., Zupancic, H., Beaudoin-Ryan, L., & Lauricella, A. R. (2015). *Teens, health, and technology: A national study*. Evanston, IL: Center on Media and Human Development, School of Communication, Northwestern University.