Information retrieval from hypermedia, with a navigable overview, at the end of primary school

Sandrine Moschetti- Rome[a], Franck Amadieu[a] and Cécile van de Leemput[b]

[a] *Laboratory CLLE (CNRS, UT2J), University of Toulouse, 5 Allée Antonio Machado, 31058 Toulouse, France*
[b] *Faculty of Psychological Sciences and Education, Department of Work and Consumer Psychology (PsyTC), Solbosch Campus - CP 122, Avenue F.D. Roosevelt, 50 1050 Brussels, Belgium*

Abstract

At the end of primary school, pupils begin to learn how to manage information web search. This activity is complex for young learners. Thus, they carry out their first learning and use in a "protected hypertext environment". Despite this material adaptation, cognitive processes remain complex and most of pupils need navigation aid. This study aimed to examine how a navigable overview of the structure of the hypermedia could support navigation and promote better performance in answering questions from non-linear hypermedia, considering children' personal differences. We hypothesized that both information retrieval skills in a non-linear printed document and motivational aspects (self- efficacy about these skills) would predict navigational behaviour and performance. Moreover, we expected that a navigable overview would help when personal resources are low. A sample of pupils schooled in priority education was examined. Our hypotheses were partially confirmed. Pedagogical applications were discussed.

Keywords [1]

Information retrieval, hypermedia, skills and self-efficacy, navigable overview, children

1. Introduction

At the end of primary school (4th and 5th grades), in Science, History and Geography, children are made aware of web research guided by specific questions[2]. They work in a "protected" digital learning environment [1]: different physical structures of hypermedia, in which content and navigation functionalities are limited by the designer. There are two types of navigation functionalities: keyword querying and/or the selection of relevant links sometimes embedded in a navigation aid like menu or a map [2], [3]. Most of studies showed that a navigational map supports a better comprehension: for a review, see Amadieu and Salmeron [4]. In addition, children learn to manage four different tasks: 1. retrieve and extract factual information; 2. make simple inference; 3. interpret and integrate information with their own knowledge (locate and relate two or three information situated in different nodes); 4. review and evaluate content. In the present study, we focused on non-linear hypermedia where children navigated exclusively by embedded links to perform the first and third tasks. In this context, we aimed to better understand the effects of a navigational overview of the hypermedia structure and personal factors on navigation and

Proceedings of the CIKM 2020 Workshops, October 19-20, Galway, Ireland
EMAIL: sandrine.rome1@univ-tlse2.fr (A.1);
franck.amadieu@univ-tlse2.fr (A.2);
cecile.van.de.leemput@ulb.ac.be (A.3)
ORCID: https://orcid.org/0000-0002-8144-8450 (A. 1);
https://orcid.org/0000-0003-3390-7146 (A.2);
https://orcid.org/0000-0002-6369-2804 (A. 3)

[2] French curriculum, cycle 3, page 6

performance. We synthesized below the state of recent scientific knowledge on (1) cognitive demand for the two tasks studied, in terms of navigation and treatment of information nodes, and on (2) the main personal predictive variables searching information with navigational overview for students at the end of primary school

1.1. Cognitive demand for information retrieval from hypermedia

Compared to a printed linear informative document, a non-linear hypermedia allows a more flexible reading path. However, the hypermedia structure makes information retrieval task more complex, and navigation behaviour are closely related to hypertext reading success [5], [6]. Information searchers have to access relevant information while they don't know the extent of the hypermedia and only one page is visible at a time. An effective strategy consists in discovering the entire document before reading the questions [7]. After, they analyze question and make inferences with their prior knowledge and reading material (*To answer this question, what do I already know and what information could I find in the documents, where and how to access it?*). To locate relevant nodes, they make inferences from hyperlinks about the content of the corresponding nodes. When hypermedia suffer from/has a lack of cohesion, they use their knowledge about the subject matter to fill in the "blanks" and build a suitable text representation. Then, without forgetting the goal of their research, they "flies over" or ignore irrelevant parts and directly access relevant information to answer questions [7]–[9]. Complexity is increased when the task requires locating and integrating information from different nodes. "Very good level pupils" have motivational, attentional, cognitive and metacognitive resources to match up to these cognitive demands. However, for most of them, at the end of primary school, explicit learnings of information retrieval strategies are in the process of being acquired [10]–[12]. Thus, navigation may be "hazardous" and poor performance are usual [10], [13], [14]. It seems that an overview may support information retrieval tasks in a hypermedia for young pupils. But as we detailed above, the effects of this navigation aid according to personal characteristics need more investigations.

1.2. Influences of individual differences and overview, with young students

Prior knowledge of the subject matter has a principal effect on performance. But with young learners, the level of this resource is often very low. Also, reading linear printed document skills are powerful predictors of navigation and performance, and browsing behaviour mediates the relationship between these skills and performance [9], [15], [16]. However, over the past two decades, linear informational materials are uncommon in children's literature. From the beginning of elementary school, students learn to read printed non- linear document with help of meta-textual clues. Although these skills are not yet fully mastered at this age, we hypothesized that they could support a task of information retrieval from hypermedia. Indeed, the structural characteristics of a printed non-linear document have similarities with a hypermedia [17]. Otherwise, studies conducted with high school students have shown that motivational aspects are predictive of web search achievement [18]. For complex academic tasks, motivational factors (self- efficacy about reading) have also been identified as predictors of reading task success [19]–[21]. However, to our knowledge, the predictive power of this personal characteristic has not been tested with children, in the specific context of information retrieval from hypermedia. Moreover, recent studies on children suggest that an overview could help to answer questions from text. Indeed, Salmerón and Garcia [15] found that, under certain conditions, an overview had a main effect on performance in information retrieval, for answering inferential questions. However, they didn't found any interaction in the relationship between skills and performances. Overview effect would be limited when learners' reading comprehension skills were too low: in this case, children couldn't take full advantage of this navigation aid [22]. In contrast, Fesel and colleagues [23] found that overview might support children with low prior knowledge. Furthermore, the interactions between personal variables and

navigation on the one hand and performances on the other hand are still unclear.

1.3. Our study: aims, framework and hypotheses

The present study examined both the predictive power effects of new personal variables (i.e. information retrieval skills in a printed non-linear document and self-efficacy about these skills) and navigable overview on navigation and performance. In addition, we examined the moderator effect of overview in relationships between personal variables and navigation behaviour and performance:

H1: Main effect of the overview. Overall, an overview displays the structure of the hypermedia and allows direct access to relevant pages (avoids the use of backward buttons and intermediate pages opening) and helps to relate information situated on different nodes. So, we expected that, compared to the opposite navigation condition, overview should support more accurate navigation (H1a) and performance should be better (H1b).

H2: Effect of personal characteristics. A printed non-linear document shares common structural characteristics with a hypermedia. Thus, information retrieval skills in a printed non-linear document (here, a double-page) should have a positive influence on navigation (H2a) and performance (H2b). Moreover, we expected to find similar results with previous studies regarding the relationship between motivational aspects and navigation on the one hand and performance on the other hand. In other words, self-efficacy about information retrieval should support navigation (H2c) and performance (H2d).

H3: Interaction effect. An overview could support students with low level skills. Furthermore, considering motivation is related to instructional condition, we hypothesized that overview could also help students with low personal resources. Thus, overview would condition the effect of skills and self- efficacy about information retrieval, on navigation and performance. In other words, without overview, information retrieval skills and self- efficacy about these skills would have a greater effect on navigation (H3a) and performance (H3b).

2. Method

2.1. Design and participants

Before the study started, we informed the headmasters of the experimental schools, the parents and the children. We also obtained the authorizations required by French law. One hundred and nineteen learners from two priority education schools participated at the experience. (M_{age} = 10,06 years, SD = 0,63 ; 61 females et 39 males). Before any data collection, they were divided, into two groups: the test group (with overview) and the control group (without overview). These learners rarely used tablets in class and only to do exercises, take photography and produce simple documents. In addition, they regularly learned how information retrieval works from linear and non-linear printed documents, but never from hypermedia or from web. In addition, students' prior knowledge was very low. Nineteen participants were excluded from the data analysis due to incomplete data (these participants did not attend all sessions in the study).

2.2. Material

2.2.1. Device and hypermedia

Two versions of a hypermedia document have been designed with the software Sparkle 2.6.2 Visual Web Design [24]: with and without overview.

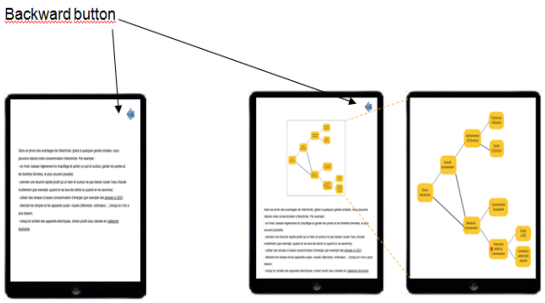

Figure 1: Example of node of the experimental hypermedia. On the left, without overview. At the center, the same node with overview. And, on the right, the overview opened.

The two versions included texts (nearly 600 words), graphical elements and data tables

distributed on 10 nodes. Hyperlinked words and backward buttons were embedded in each node. The versions differed in the presence or not of an overview. Overview was implemented at the top of each node. It showed exactly the hypermedia structure. Like the city map, a red point indicated where the reader was in his navigation path (Fig.1). To navigate, in both versions of the hypermedia, participants could click on hyperlinks in the text and backward buttons. In addition, the test- group could click on the Overview to display it full screen and make it interactive. The hypermedia were read on iPAD tablets (9.7" screen size) belonging to schools.

2.2.2. The task

The task consisted of retrieving relevant information from hypermedia in order to answer three questions: one explicit question (i.e. factual; the answer required finding one piece of information located in one node) and two implicit questions (i.e. inferential; the answer required finding and relating two or three pieces of information located in different nodes). All questions were given at the beginning of the session. Students wrote their answers with paper and pencil.

2.3. Measures

2.3.1. Individual differences

Prior knowledge. Prior knowledge about production and consumption of electricity was assessed by 9 open ended questions during a collective session. For each question, the scoring consisted of 0 point for a wrong answer or no answer, 0.5 for a partial correct answer, and 1 point for a correct answer.

***Information retrieval in non-linear printed document skills*.** We used a test adapted from PIRL assessment tool [25]. The document included texts, pictures, graphics and data tables. In accordance with our problematic, we focused on retrieving and extracting factual information and, making simple inference. Students answered four explicit or implicit questions. Each item was scored as following: 0 for no answer or wrong answer; 0.5 for a partial answer; 1 for a correct answer. A mean score was computed from the four items.

Self-Efficacy about information retrieval in non-linear printed document. Data about self-efficacy was collected from self-ratings based on Bandura's Children's Self-Efficacy Scale [26]. They indicated the strength of their efficacy beliefs on a 100-point scale, ranging in 10-unit intervals from 0 ("I feel I can't do it at all") to complete assurance, 100 ("I feel I can do it at all"). Questions were related to reading and information retrieval from expository printed non- linear document: for example, "*To which extent do you feel able to read an expositive document that includes texts, pictures, graphics and data tables?"* (translated from French). Chronbach's alpha = 0.78. A mean score of self-efficacy was computed from the items ratings.

2.3.2. Dependent variables

Performance at the task. The scale was: 0 point for wrong answer or no answer; 0.5 point for a partial correct answer, and 1 point for a full correct answer. A mean score of performance was computed from the three items.

N***avigation accuracy.*** The quality of navigation was assessed from accuracy ratio [27]: the total number of target nodes of the predefined path visited / the total number of open nodes (excluding clicks in order to open the overview page, for the test group). We computed the ratio for the global path. It was equal to 1 if all relevant nodes were visited and equal to 0 if none of them were visited. A mean score of navigation accuracy was computed from the three accuracy ratios.

2.4. Procedure

The study was included in a Science and Technology learning sequence about production and consumption of electricity which lasted nine sessions (2 sessions per week, on average). The experimenter conducted all sessions. Prior knowledge, skills and self-efficacy were measured in the first session. The 2nd and 4th sessions consisted in a familiarisation with hypermedia and experimental conditions (hypermedia about production of electricity). The 6th session was the experimental session (hypermedia about consumption of electricity). All questions were given on questions- answers sheet (paper and

pencil format) before starting the information retrieval. Participants wrote their answers respectively under each question. They could process questions in any order and they could come and go between the hypermedia and the questions as much as they needed. They could ask help at the experimenter for reading questions and writing answers. Finally, there was no time constraint. During 5th, 7th, 8th and 9th sessions, over the experimentation, the experimenter proceeded to structure new knowledge and finally assessed the level of acquisition about this knowledge.

3. Results

First, we checked with *t*-tests that the groups did not differ on the cognitive factors assessed. As expected, there were no differences between groups in their scores for prior knowledge, $t(85) = -.88$; $p = .38$, skills, $t(85) = 0.75$, $p = .45$; and self-efficacy, $t(85) = 1.68$, $p = .09$.

In order to test the effects of guidance condition (i.e. with vs. without overview) and personal differences (skills; self-efficacy), we performed multiple linear regression analyses with interaction terms, on each dependent variable. Skills and self-efficacy were entered as z-standardized variables. The guidance condition was entered as a contrast-coded dummy variable. In our model, we entered prior knowledge as a control variable but we expected a non-significant effect because level was very low in both groups, M = .14, SD = .17 and for test-group, M = .17, SD = .20.

3.1. Navigation accuracy

A quick descriptive analysis of the navigation historic showed that all participants of the test-group had systematically use overview for their navigation. The model for the navigation accuracy scores appeared significant, $R_{corr}^2 = .21$, $F(5,80) = 5.58$, $p < .001$. As expected, overview supported better navigation accuracy $\beta = .93$, $t(85) = 4.78$, $p < .001$. But in contrast with our hypotheses, neither skills influenced navigation $\beta = .25$, $t(85) = 1.93$, $p = .056$, nor self-efficacy, $\beta = .27$, $t(85) = 1.93$, $p = .057$. And, no interaction was observed: neither between overview and skills, $\beta = -.11$, $t(85) = -.54$, $p = .59$ nor between overview and self- efficacy, $\beta = -.26$, $t(85) = -1.34$, $p = .18$.

3.2. Performance

The model for the performance scores revealed a significant model, $Rcoor^2 = .31$, $F(6,79) = 7.44$, $p < .00$. In accordance with our hypotheses, overview supported performances, $\beta = .91$, $t(85) = 4.96$ $p < .001$. Performance were also improved by skills, $\beta = .36$, $t = 2.85$, $p = .006$ and by self-efficacy, $\beta = .38$, $t(85) = 2.84$, $p = .006$. Moreover, prior knowledge had a significant effect but at the limit of the conventional threshold, $\beta = .19$, $t(85) = 2.05$, $p = .044$. About interactions, our hypotheses were partially confirmed. There was no interaction between skills and guidance, $\beta = -.32$, $t(85) = -1.76$, $p = .083$. But the interaction between self-efficacy and guidance was significant, $\beta = -.40$, $t(85) = -2.21$, $p = .003$. Navigable overview moderated self- efficacy effect. Overview sustained participants with lower level of self-efficacy (Fig.2).

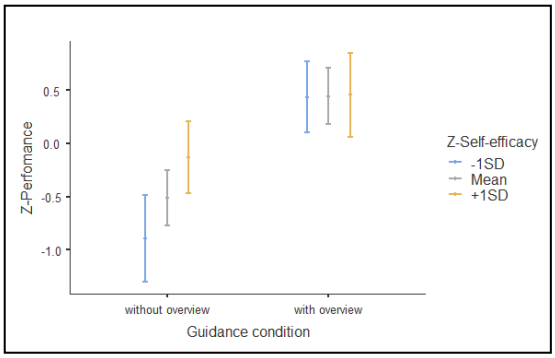

Figure 2: Predicted value of performance; effects of self-efficacy according to guidance condition

4. Discussion

The study investigated the effects of a navigable overview and individual characteristics (skills about information retrieval from printed non-linear document; self- efficacy about these skills) on navigation and performance: main effects and interactions. Our assumptions were partially confirmed but tended to support or complement previous studies.

4.1. Predictive variables of information retrieval with young learners

Predictive variables of navigation: our study confirmed the main effect of an overview on the quality of navigation [4], [22]. In addition, while the test-group had the choice to navigate via the links embedded in the node or via the overview, all participants systematically used overview. Thus, the acceptability of this navigation aid would be very strong with children. It would help to make lexical or semantic correspondences between the question keywords and relevant nodes. It allowed direct access to the target pages (avoiding intermediate pages). Over this design factor, we tested new personal variables: skills about information retrieval from a printed non-linear document and self- efficacy about these skills. We found no main effect related to personal variables. Maybe participants would not have transferred their cognitive and motivational resources on navigation strategies in hypermedia; they would not have identified the implicit structural similarities between a printed non-linear document and a hypermedia.

Predictive variables of performance**:** our study confirmed that cognitive resources have a major effect on performance. Like linear reading skills [9], [15], [16], non-linear reading skills would help to process the selected pages and link them. Our study also highlighted the importance of taking into account motivational aspects when the task is complex and, above all, the influence of instructional choices on motivation [21]. Participants who felt they had low self-efficacy would have perceived that overview could give them autonomy.

4.2. Limits of the study and further works

About navigation accuracy: we could not be sure that a low level reflected a lack of efficiency: it could be explained by a navigation behaviour motivated by the pleasure of handling the material over the task. Indeed, as mentioned earlier about this research, participants have no previous experience with reading hypermedia; it was possible that the novelty effect had negatively influenced navigation accuracy. About dependant variables, we computed the average navigation performance and navigation accuracy. Yet, previous studies [7], [8], [16] suggested that the effect of predictive variables would differ depending on the type of question. Regarding motivational aspects, we measured self-efficacy before participants were informed of the task conditions. However, it was possible that self-efficacy may have evolved after the sessions of familiarization. Thus, for future research, it would be appropriate to collect a measure of self-efficacy after the exploratory phase and after explaining to each group the task and de navigation condition. However, despite these limitations, this study allowed us to identify directions for instructional recommendations. First, teacher should explicitly teach the similarities and the differences between different reading materials. It would help children to transfer and adapt their personal resources acquired in printed non linear document to hypermedia. Until students have acquired sufficient personal resources, designers and teachers should ensure that reading materials are equipped with a navigation aid like an overview. But in order to children benefit fully from overview, they need to be provided with explicit learning about it.

5. Conclusion

The study supported that both personal and design factors predict digital reading with young students. For an information retrieval task, a navigational overview could be an effective aid. Moreover, we highlighted the importance of both cognitive and motivational variables about reading a non-linear document.

6. Acknowledgements

This research was funded by the National Agency for Research (LETACOP project). We thank Pierre-Vincent Paubel, engineer, for his help in collecting and treatment of logs files.

7. References

[1] M. Macedo-Rouet et J.-F. Rouet, « Documents multiples : processus et difficultés de lecture », *Canopé*, n° n° 48, p. p.13, juill. 2011.

[2] N. Gauducheau, F. Cuisinier, et C. Garitte, « Utilisation d'hypermédia et recherche d'information chez les enfants », in *Ergonomie des documents électroniques*, Presses Universitaires de France., Paris: Presses universitaires de France, 2008, p. 103-128.

[3] J.-F. Rouet, C. Ros, A. Goumi, M. Macedo-Rouet, et J. Dinet, « The influence of surface and deep cues on primary and secondary school students' assessment of relevance in Web menus », *Learn. Instr.*, vol. 21, nº 2, p. 205-219, avr. 2011, doi: 10.1016/j.learninstruc.2010.02.007.

[4] F. Amadieu et L. Salmerón, « Concept Maps for Comprehension and Navigation of Hypertexts », in *Digital Knowledge Maps in Education*, D. Ifenthaler et R. Hanewald, Éd. New York, NY: Springer New York, 2014, p. 41-59.

[5] J. Naumann et L. Salmerón, « Does Navigation Always Predict Performance? Effects of Navigation on Digital Reading are Moderated by Comprehension Skills », *Int. Rev. Res. Open Distrib. Learn.*, vol. 17, nº 1, févr. 2016, doi: 10.19173/irrodl.v17i1.2113.

[6] L. Salmerón, H. I. Strømsø, Y. Kammerer, M. Stadtler, et P. van den Broek, « Chapter 4. Comprehension processes in digital reading », in *Studies in Written Language and Literacy*, vol. 17, M. Barzillai, J. Thomson, S. Schroeder, et P. van den Broek, Éd. Amsterdam: John Benjamins Publishing Company, 2018, p. 91-120.

[7] L. Salmerón, J. Naumann, V. García, et I. Fajardo, « Scanning and deep processing of information in hypertext: an eye tracking and cued retrospective think-aloud study: Scanning and deep processing in hypertext », *J. Comput. Assist. Learn.*, vol. 33, nº 3, p. 222-233, juin 2017, doi: 10.1111/jcal.12152.

[8] J.-F. Rouet, « What was I looking for? The influence of task specificity and prior knowledge on students' search strategies in hypertext », *Interact. Comput.*, vol. 15, nº 3, p. 409-428, juin 2003, doi: 10.1016/S0953-5438(02)00064-4.

[9] C. Hahnel, F. Goldhammer, J. Naumann, et U. Kröhne, « Effects of linear reading, basic computer skills, evaluating online information, and navigation on reading digital text », *Comput. Hum. Behav.*, vol. 55, p. 486-500, févr. 2016, doi: 10.1016/j.chb.2015.09.042.

[10] J.-F. Rouet et B. Coutelet, « The acquisition of document search strategies in grade school students », *Appl. Cogn. Psychol.*, vol. 22, nº 3, p. 389-406, avr. 2008, doi: 10.1002/acp.1415.

[11] R. Cerdán, R. Gilabert, et E. Vidal-Abarca, « Selecting information to answer questions: Strategic individual differences when searching texts », *Learn. Individ. Differ.*, vol. 21, nº 2, p. 201-205, 2011, doi: 10.1016/j.lindif.2010.11.007.

[12] M. G. Cataldo et J. Oakhill, « Why Are Poor Comprehenders Inefficient Searchers? An Investigation Into the Effects of Text Representation and Spatial Memory on the Ability to Locate Information in Text », *J. Educ. Psychol.*, vol. 92, nº 4, p. 791-799, 2000.

[13] J.-F. Rouet et A. Potocki, « From reading comprehension to document literacy: learning to search for, evaluate and integrate information across texts / De la lectura a la alfabetización documental: aprender a buscar, evaluar e integrar información de diversos textos », *Infancia Aprendiz.*, vol. 41, nº 3, p. 415-446, juill. 2018, doi: 10.1080/02103702.2018.1480313.

[14] J.-F. Rouet, *The skills of document use: from text comprehension to Web-based learning*. Mahwah, NJ: L. Erlbaum Associates, 2006.

[15] L. Salmerón et V. García, « Reading skills and children's navigation strategies in hypertext », *Comput. Hum. Behav.*, vol. 27, nº 3, p. 1143-1151, mai 2011, doi: 10.1016/j.chb.2010.12.008.

[16] J. Naumann et F. Goldhammer, « Time-on-task effects in digital reading are non-linear and moderated by persons' skills and tasks' demands », *Learn. Individ. Differ.*, vol. 53, p. 1-16, janv. 2017, doi: 10.1016/j.lindif.2016.10.002.

[17] É. Bautier, J. Crinon, C. Delarue-Breton, et B. Marin, « Les textes composite: des exigences de travail peu enseignées? », *Repères*, nº 45, p. 63-79, juin 2012, doi: 10.4000/reperes.136.

[18] J. Naumann, « A model of online reading engagement: Linking engagement, navigation, and performance in digital reading », *Comput. Hum. Behav.*, vol. 53, p. 263-277, déc. 2015, doi: 10.1016/j.chb.2015.06.051.

[19] Y. S. Lee et M. Jonson-Reid, « The Role of Self-Efficacy in Reading Achievement of Young Children in Urban Schools », *Child Adolesc. Soc. Work J.*, vol. 33, nº 1, p. 79-89, févr. 2016, doi: 10.1007/s10560-015-0404-6.

[20] H.-F. Shang, « Online Metacognitive Strategies, Hypermedia Annotations, and Motivation on Hypertext Comprehension », *Educ. Technol. Soc.*, nº 19 (3), p. 321–334, 2015.

[21] J. T. Guthrie, A. Wigfield, et W. You, « Instructional Contexts for Engagement and Achievement in Reading », in *Handbook of Research on Student Engagement*, S. L. Christenson, A. L. Reschly, et C. Wylie, Éd. Boston, MA: Springer US, 2012, p. 601-634.

[22] L. Salmerón et V. García, « Children's Reading of Printed Text and Hypertext with Navigation Overviews: The Role of Comprehension, Sustained Attention, and Visuo-Spatial Abilities », *J. Educ. Comput. Res.*, vol. 47, nº 1, p. 33-50, juill. 2012, doi: 10.2190/EC.47.1.b.

[23] S. Fesel, E. Segers, et L. Verhoeven, « Individual variation in children's reading comprehension across digital text types », *J. Res. Read.*, vol. 41, nº 1, p. 106-121, févr. 2018, doi: 10.1111/1467-9817.12098.

[24] D. Trambusti et D. Wilcox, *Sparkle 2.6.2 Visual Web Design*. 2011.

[25] « PIRLS 2016 : évaluation internationale des élèves de CM1 en compréhension de l'écrit. Évolution des performances sur quinze ans. », p. 4.

[26] A. Bandura, « Guide for constructing self-efficacy scales », in *Self-efficacy beliefs of adolescents*, Greenwich, CT: Information Age Publishing., vol. 5, F. Pajares et T. Urdan, Éd. 2006, p. 307-337.

[27] F. Amadieu, C. Bastien, et T. André, « Les méthodes on-line 1: analyse des parcours », in *Ergonomie des documents électroniques*, A. Chevalier et A. Tricot, Éd. PUF, 2008, p. 251-270.