Learning with Trees:
A Non-Linear E-Textbook Format for Deep Learning

Eric Spero, Milica Stojmenović, Ali Arya, and Robert Biddle

Carleton University, Ottawa, Canada,
{eric.spero, milica.stojmenovic, ali.arya, robert.biddle}@carleton.ca,

Abstract. A "deep" approach to education requires considering the non-linear connections between concepts, which is difficult to do with the standard linear textbook format. Guided by the cognitive science literature, we designed a format for a new, non-linear e-textbook format, and implemented a high fidelity prototype. We tested this prototype with end-users, measuring its pedagogical efficacy, usability, and overall likability, in comparison with a linear control. We found no significant differences in learning outcomes between the two conditions, but a significantly greater number of participants preferred the non-linear interface. We suspect that many potential advantages of the non-linear format were negated by our short study. Future work should study the effects of the non-linear interface over a longer period of use.

Keywords: human-computer interaction, educational technology, cognitive science

1 Introduction

Educators are calling for a "deeper" coverage of learning material—one that focuses more on the interactions between concepts, within and across domains [20]. This is a proposed departure from the present dominant "broad" approach, which prefers to instead focus on covering a large number of concepts in relative isolation from each other.

Textbooks are important tools in formal education, and making significant changes to textbooks is a prerequisite for making changes to education as a whole [27]. Textbooks today are designed to support the broad approach [2, 22, 30, 33]: there is a need for deep textbooks.

The relationships between concepts have a non-linear structure: they are hierarchical [25, 36, 39, 1], associative [10, 17, 31], and multidimensional [40, 15]. Communicating non-linear relationships is possible in linear texts, but we think it comes with a cost. For example, hierarchy can be signalled with language and with headers, but we think that organizing concepts into a visual hierarchy would make the relationships more salient to the learner. We think that a non-linear textbook format would better support the deep approach to education.

In this paper, we describe the design and study of a novel e-textbook format which is guided by the cognitive science literature on knowledge representation and organization. Our non-linear format deviates from the standard linear format in two key ways.

First, we separate content into two types: core and peripheral, as seen in Fig. 1. Core content corresponds to domain concepts, and peripheral content grounds these domain concepts in the real world. Second, content is arranged non-linearly. Core elements are organized in a hierarchy, and each core element is flanked to the left, right, and bottom by up to three types of peripheral content. Our format more explicitly communicates the relations between concepts, reducing extraneous cognitive load [34], which frees up mental resources for learning.

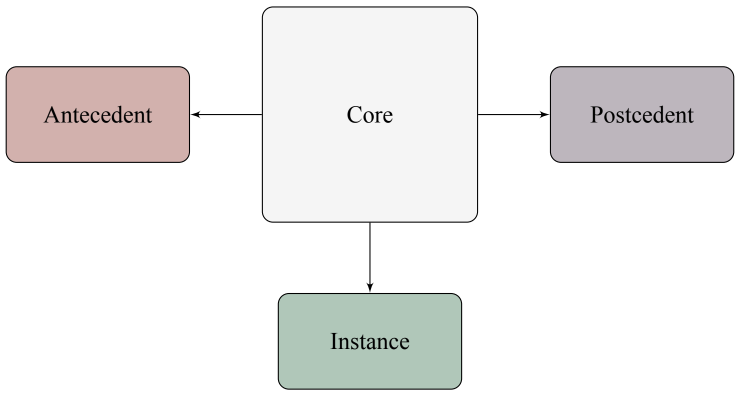

Fig. 1. Our version of a page: a core concept surrounded by grounding concepts

Our research questions were:

1. Can we create a non-linear e-textbook that accords with the cognitive science literature?
2. Does this non-linear textbook better support learning?
3. Do users prefer the proposed non-linear format?

To address the first question, we first conducted a review of the cognitive science literature. Highlights of this literature review are presented in Section 2. This section also reviews the relevant literature on e-textbooks. Guided by this literature review, we then designed and implemented a prototype, which is described in Section 3. After designing the prototype, we came across a study which suggested that our e-textbook interface may present difficulties for certain kinds of learners. In Section 4 we describe this study, and our attempts to examine the challenge it presents. To address the second and third research questions, and to examine the just-mentioned challenge, we designed a study to test the effectiveness of our non-linear interface compared to a linear control. We describe the methodology of the study in Section 5, our findings in Section 6, and discuss their implications in Section 7. Finally, we present concluding remarks in Section 8.

2 Background

2.1 Project 2061 and the Next Generation Science Standards

Project 2061 is a research and development initiative aimed at improving literacy in science, mathematics, and technology through educational reform. The project was created in response to middling scores [11] in science, math, and technology literacy among students in the United States [8]. Project 2061 believe that the science curricula focus too much on breadth, and call for a "radical" reduction in the total number of concepts students are asked to learn [22]. They believe that real science literacy requires making connections between science, mathematics, and technology, with the arts, humanities, and the vocational subjects.

Project 2061 regard the design of new and better instructional material such as textbooks as an instrumental component of educational reform needed to improve STEM literacy in the Unites States. Professional development (i.e. training instructors) is much more expensive in comparison [27].

Partially informed by the work of Project 2061, The Next Generation Science Standards (NGSS) [20] challenge the designers of curriculum material to present ideas in a coherent manner. This means that a) ideas should build upon each other over lessons and units, and b) that students understand how the new content they are learning relates to prerequisite ideas, or other ideas already present in long-term memory.

2.2 E-Textbooks

Although textbook material presented in an electronic format is widely available through retailers like Apple, Amazon, and Google, e-textbooks have not enjoyed the same success as e-books [9]. The typical e-textbook in use today is structurally very similar to print texts, with some additional features added such as the ability to search the text, and support for multimedia content. Studies comparing these kinds of electronic texts and standard print texts typically show no differences in learning outcomes, but a preference for using standard print texts [29, 35, 24]. Four out of five students prefer print texts to digitized texts despite print being more expensive and less portable, because they are more familiar, and they better afford highlighting, dog-earing, and annotation [24]. In a study by Daniel and Woody[9], students using e-textbooks spent significantly more time reading than those using print texts. In general, it seems that the drawbacks of the typical e-textbook compared to print texts do not outweigh the benefits. Authors like Toukonen [35] have argued the current incarnation of e-textbooks do not properly leverage the advantages of digital media technology such as dynamicity and non-linearity.

More recently, in response to the issues with this first type of e-textbook, efforts have been made to create e-books that leverage the interactive nature of computers. Interactive texts (i-texts) contain less text, and add animations of key concepts, more questions, and interactive tools [13]. A new system proposed by Miller and Ranum [19] incorporates video, code editing, execution, and visualization inside the textbook. A commercial line of i-texts called zyBooks [13] have recently gained popularity [41]. zyBooks present text interleaved with interactive examples of the learning material. The approach taken by i-texts is complementary to our approach: our non-linear text could one day accommodate—and may well benefit from—interactive material.

DeStefano and LeFevre [12] review studies on hypertext reading, and discuss the advantages and disadvantages relative to linear texts. Hypertext is more demanding of working memory resources than linear text because it requires that readers make a decision before selecting a link. The researchers do not specify if this additional load is extraneous (i.e. unnecessary), or if it aided learning. They note that in many cases where hypertexts appear to impair learning, it is because these hypertexts exceed the working memory capacity of the reader. Many hypertexts used visual overviews of the content to reduce the cognitive load burden placed on readers, but these visual overviews were helpful only when their structure matched the inherent structure of the domain. Structuring texts hierarchically seemed to aid learning, whereas other structures (e.g. allowing readers to navigate a semantic network of concepts) seemed to impair learning.

2.3 Review of Cognitive Science Literature on Concepts

E-textbooks should make use of the advantages of dynamicity and non-linearity. To do this in a way that does not overwhelm readers' working memory capacity, e-textbook designers should strive for harmony between the text's format and the mind. We conducted a review of the cognitive science literature on concepts, and this subsection summarizes some key findings.

Knowledge is Categorical. Concepts are a unit of knowledge about the world. Through concepts we divide the complex, continuous world into simple, finite categories, thereby lightening the load of perception [26, 5]. Concepts form the basis for thought and communication [16]. A fundamental property of knowledge is its categorical nature [4]. A conceptual system is not a collection of holistic images like a camera. Rather, it is a collection of category knowledge, where each represented category corresponds to a component of experience, not to an entire holistic experience.

Knowledge is Hierarchical. There is evidence that concepts are subdivided hierarchically in two ways: taxonomically, and partonomically [36]. A taxonomy organizes things by kind: a McIntosh is a kind of apple, and an apple is a kind of fruit, but the reverse is not true. A partonomy organizes things by part: a piston is a part of an engine, and an engine is a part of a car, but the reverse is not true. This is true of abstract concepts as well, such as governments: a democracy is a kind of government, and a government consists of the legislative, judicial, and executive branches. Events have a hierarchical structure as well [39, 1].

Meaning and Use. Philosophical pragmatists hold that, phenomenologically speaking, the meaning of objects and categories is subjective and situation-dependent [6]. Objects get their meaning from their relation to a goal. For example, a chair is canonically viewed as a tool used for seating, but in other situations it can have a different meaning: it can be a tool used for standing-on, or an obstacle that impedes motion. This idea underlies a number of important intellectual works of the last century, including Gibson's Affordance Theory [14]. Gibson argues that the act of perceiving leads one toward a course of action: when we perceive a door handle, we do not perceive the object in-itself, but rather the ways in which we can interact with it.

Grounding Abstract Concepts. Traditionally, there is thought to be a sharp distinction between concrete concepts and abstract concepts. This view is supported by so-called concreteness effects, which are well-established cognitive processing advantages for concrete concepts over abstract concepts: concrete concepts are linked to stronger memories, and they are accessed and comprehended more quickly than abstract concepts [5]. However, research by Schwanenflugel, Shoben, and colleagues (e.g. [28]) showed that this processing advantage for concrete concepts disappears when people are provided with an instantiating situation for abstract concepts. Providing concrete examples of abstract concepts has also shown to be an effective teaching tool [3, 23].

Cognitive Load Theory. Cognitive load theory describes the role that working memory limitations play in the learning process [34]: learning is an information processing activity, humans are limited information processors, and so it is important to manage the information workload placed on students.

Some workload is necessary and desirable, and some is not. What makes workload desirable is whether or not it contributes to the acquisition of new concepts and skills. *Intrinsic* load is the workload inherent to the material being learned; it is the minimal workload associated with learning a given piece of information. The driving force behind this type of workload is element interactivity, which is the total number of elements that must be considered at one time in order to understand some piece of information. Learning a concept that is high in element interactivity requires that a number of elements are held in working memory simultaneously. *Extraneous* load is workload that is unnecessary and therefore detrimental to learning. An example of this type of workload is requiring that students search for some piece of information when it could just be provided for them. The crucial management of extraneous workload becomes especially important when element interactivity is high, because the learning task will test the capacity of working memory. Good instructional texts should keep extraneous cognitive load to a minimum.

The deep approach to education is high in element interactivity, so minimizing extraneous workload is even more important than in the broad approach.

3 Design and Implementation

Our primary design goal was to harmonize the structure of our text with the conceptual structure of its content. We think this will aid the learning process by reducing extraneous workload in two ways.

The first is by *offloading mental work onto the environment*. Concepts are complex—they are multidimensional, hierarchical, and categorical—and the interrelations between domain concepts and other concepts are brought to the forefront in the deep approach.

If the information regarding how concepts relate to each other can be represented in the visual environment, it will free up cognitive resources which can be spent on some other aspect of the learning task at hand, thereby enhancing learning.

The second is by *minimizing thought about the text's structure that is not also thought about its content*. To navigate instructional materials, students must consider its structure. Wherever the structure of the instructional material matches the structure of the content, any effort exerted thinking about the structure of the instructional material is also effort spent learning the content. Conversely, wherever there is disharmony between the two structures, the student will be exerting effort that does not help them learn the content. For this reason, we think that the designers of instructional materials should strive for structural harmony between texts and content.

3.1 Key Features

This goal of harmonizing the structure of the text and its content is realized in our design in two major ways: we (i) divide content into two major categories: *core* and *peripheral*; and (ii) allow for the non-linear navigation of content.

Core and Peripheral Content. We identify two broad categories of content in instructional materials. The first type of content pertains to domain concepts. Examples of domain concepts from psychology include: operant conditioning, behaviorism, and working memory. We call these concepts *core* concepts because we believe they are the focus of an educational unit, and that learning them is the primary goal. Since we believe that domain concepts have a important status in the classroom, we believe that core content should likewise have an important status in instructional materials.

The second category of content is content that provides grounding context for the core content, thereby enriching its understanding. These are ideas that relate the abstract domain concept to the real world. Examples of these kinds of content include instances of the concept, the history of the concept, and ways in which the concept can be usefully applied. We identify three categories of enriching content: *antecedents*, *postcedents*, and *instances*.

Antecedents (Inputs). While each student encounters a domain concept at a particular instant in time, a concept is something that stretches across time, and its past can tell us something about its present. Antecedents are the category of concepts that come prior to (chronologically or causally) the core concept. Examples of antecedent content in scientific disciplines include descriptions of key research, and paradigmatic assumptions.

Postcedents (Outputs). The second category of enriching content involves the ways in which the concept can be usefully applied in the real world. Since applying the concept comes after mastering it, from the reader's point of view this category of concepts mirror antecedents; we therefore call these postcedents. If a concept's use is an important element of its meaning (if not meaning itself), this category of concepts will be of great interest to learners. And if the primary goal of education is to foster the development of skills that are socially useful—which we hold—then this category of concepts is essential.

Instances (Real-World Examples). The concepts students are asked to learn are all abstractions of instances. There is evidence that abstract concepts are easier to understand when they are linked to a grounding situation. An example of an instance of the abstract concept truth is a legal verdict. The three types of enriching content provide three different types of grounding for each abstract core concept. We think that linking abstract ideas with ideas relating this knowledge to the real world will help readers develop better situation models than they might without this information.

We consider the proposition that these three types of enriching concepts are important for learning to be uncontroversial as they already feature prominently in textbooks. For example, in psychology textbooks, a domain concept is often accompanied by real-world scenarios instantiating the concept, key studies providing empirical substantiation for the concept are often summarized, and information regarding how the concept can be applied in one's life. Our proposed design gives these enriching concepts an elevated status in the text's structure. For each core concept, students will be able to quickly see a number of different ways in which it relates back to the material world.

Non-linear Organization/Navigation. According to our literature review, concepts have a non-linear structure. That is, they are hierarchical and multidimensional. However, learning materials are typically presented linearly. In linear texts, hierarchy is signalled (a) linguistically through sentences such as "x is a kind of y", and (b) typographically through headers (e.g. chapter, section, subsection). The dimensions of a concept, when they are provided, are presented in-line with the core domain content.

We think a linear ordering of material that is inherently non-linear is problematic for a number of reasons. First, strict linearity forces a single ordering of learning material. This is a problem because it seems to us that there is no single optimal way to arrange learning material: what will be optimal for one student will be sub-optimal for others, and what will be optimal for one student at one time under one set of circumstances may for that same student be sub-optimal at some other time under other circumstances. A non-linear organization of content will allow students the flexibility of choosing the order that suits them at the time of reading. A linear arrangement of non-linear material also introduces a number of sources of extraneous workload [34]. This problem is especially severe in the deep approach to learning. We discuss this in more detail in Section 4.1.

We think that, rather than present concepts in a linear order and signal their underlying non-linear features with language and typography, it is better to simply present the material non-linearly.

We think we avoid the pitfalls of non-linear texts covered in the review by DeStefano and LeFevre [12] for three reasons. 1. We structure the material according to the domain hierarchy; 2. we take care to limit navigational freedom so that working memory capacity is not exceeded; and 3. we provide a hierarchical visual overview of the content.

3.2 The Model

We propose the following model for an e-textbook interface, incorporating the key design concepts just discussed.

There are many different ways of positioning core and peripheral elements with respect to each other. We elect the arrangement depicted in Fig. 1. Core concepts are given a central position reflective of their preeminent status in education, and they are flanked by enriching concepts to the right, bottom, and left. We place antecedents to the left, and postcedents to the right of the core because these three elements comprise a timeline of past, present, and future, and in the English-speaking world we think of time as flowing from left to right [37]. We place instances beneath the core because an instance is a taxonomical concept, and taxonomies are intuitively vertical [25], with instances being the lowest element. For each domain concept, students will quickly and easily be able to see up to three kinds of grounding information.

Core-peripheral clusters (the objects depicted in Fig. 1) are organized into a hierarchy. Students navigate through the text by moving up and down this hierarchy. A visual representation of the hierarchy—that is, a tree graph—should be provided to give students an overview of the concepts they will be asked to learn, and to allow quick and easy navigation to various locations in the tree.

3.3 Early Prototypes

The sketch in Fig. 2 shows our earliest attempt at an interface that is consistent with the model just described. At this stage, we intended on providing both a partonomy, which would be descended by clicking buttons in the middle of the sketch, and taxonomy, which would be descended by clicking nodes on the extreme bottom of the page. Making a sharp distinction between partonomy and taxonomy made the interface crowded and confusing, while not offering a clear educational advantage, so this distinction was dropped in later iterations.

A high fidelity prototype featuring a simpler design is shown in Fig. 3. We conducted a cognitive walkthrough [32] with two human-computer interaction experts, who made a number of recommendations, the most important of which was to include a visual overview of the text structure to prevent overwhelming the working memory resources of readers. The experts also recommended the inclusion of text search, and slide-in/out animations for showing peripheral content, to reinforce the impression that this content is positioned either to the left, right, or beneath the core content.

3.4 High Fidelity Functional Prototype

We implemented the functional prototype of the interface, shown in Figs. 4, 6, and 5 using JavaScript, HTML, and CSS. Text corresponding to the currently selected domain concept is shown in the middle of the screen. Clicking on buttons on the periphery triggers an animation where a panel “slides-in” from the side of the window the button is on. The panel features either peripheral content (left, right, bottom), or the “treemap” (top).

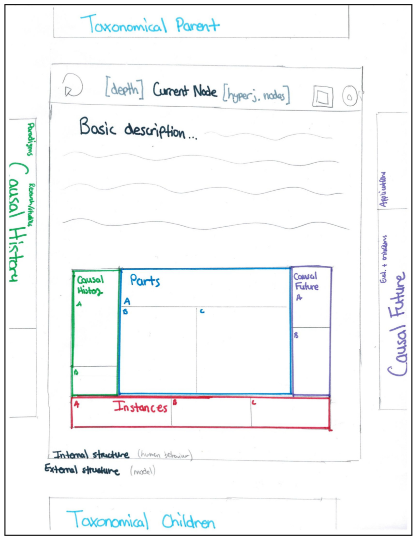

Fig. 2. Paper prototype: early design iteration

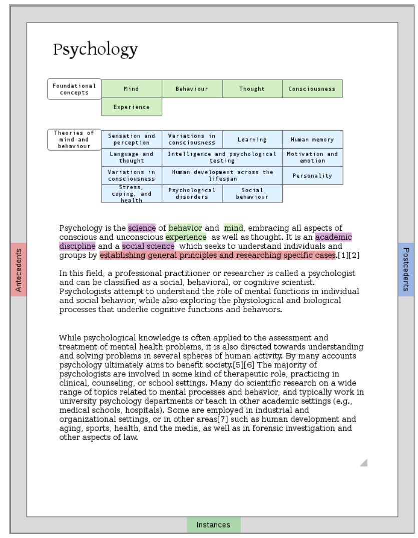

Fig. 3. High fidelity prototype: a refinement of the sketch in Fig. 2

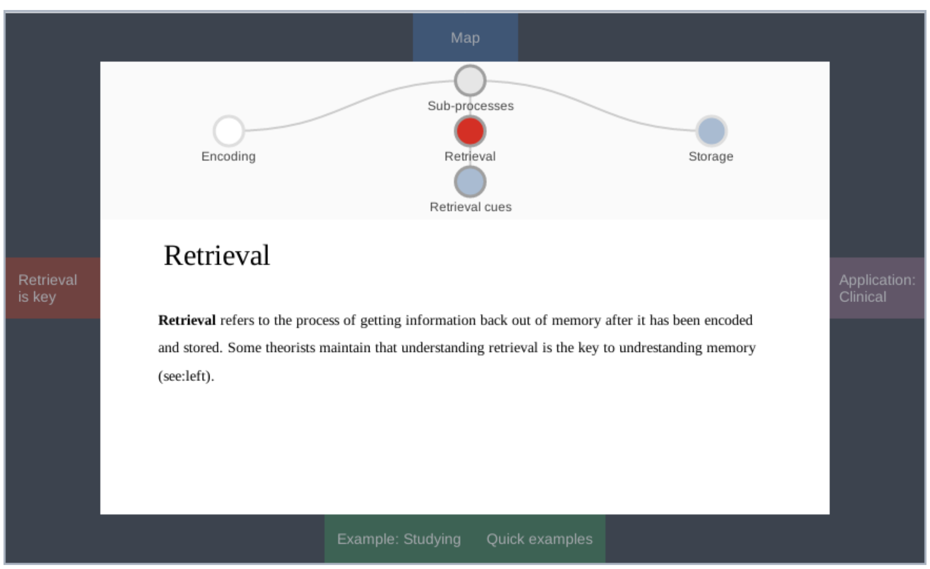

Fig. 4. Our functional, high fidelity, prototype for a non-linear e-textbook.

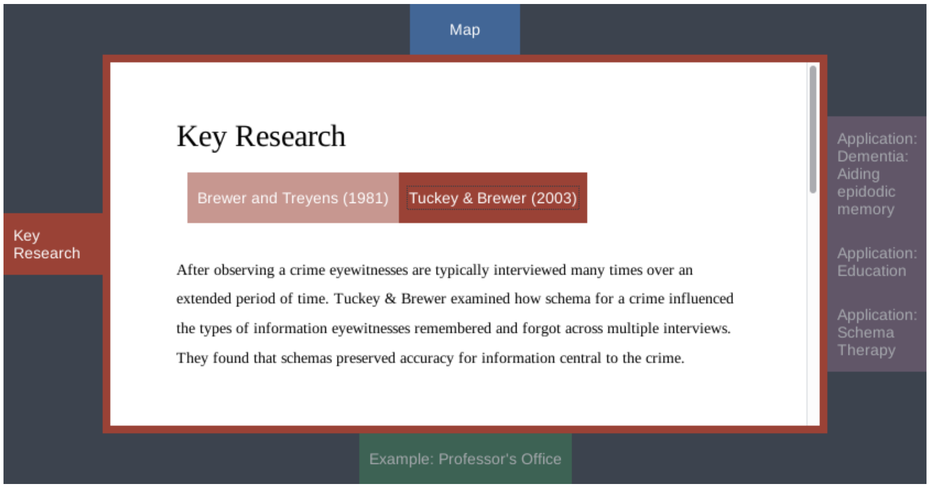

Fig. 5. Viewing peripheral content: the result of clicking the "Key Research" button

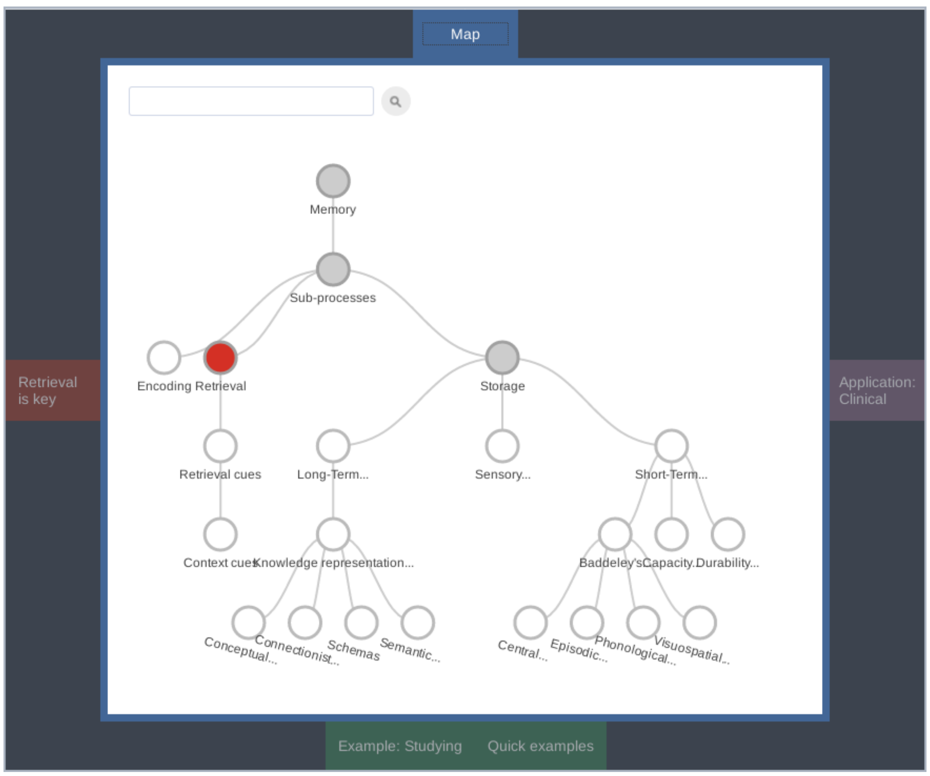

Fig. 6. The "treemap": a navigable, hierarchical table of contents

Fig. 5 shows the result of clicking on a peripheral button. Content pertaining to the peripheral concept, in this case an important prior study, is shown in the middle of this pane. Multiple peripheral concepts can be included on a single pane. When this happens, buttons are placed above the text area. When clicked, the displayed text underneath changes accordingly.

What we call the "treemap" is a visual representation of the the hierarchy linking together core concepts, which we implemented using d3.js [21]. Figure 6 shows the full treemap, which is made visible after clicking the "Map" button at the top of the screen. The currently selected node is shown in red, and visited nodes are greyed-out. Clicking a node allows users to navigate to its corresponding concept. In the top-left of the treemap pane is a text box which allows users to perform a text search of the e-textbook.

A miniaturized version of the treemap is shown at the top of the main page (Fig. 4), showing the parent, sibling, and child nodes of the currently selected node. The mini-treemap shows users their local environment at a glance, and lets them easily move through the tree in single steps.

3.5 Answering Research Question 1

Our first research question was: Can we create a non-linear e-textbook that accords with the cognitive science literature? At this stage, we felt we were able to answer "yes". Our design has a hierarchical and multidimensional structure, and it provides several types of grounding for abstract concepts.

3.6 Linear Interface

To examine the effects of non-linearity on support for learning in our study (described later in Section 5, we created a second interface (Fig. 7 with the same look-and-feel as the non-linear interface, except with linear organization and navigation. The mini-treemap and the treemap were removed. In place of the treemap was a numbered list of concepts which, when clicked, brought the user to that part of the text. We replaced the basic functionality of the mini-treemap (i.e. incremental movements through the text) with forward and backward buttons on the left and right periphery.

4 Challenge to the Design

The model just presented is a text of higher *coherence* than the standard linear text format: The present model more explicitly shows the structural relations between entities than standard linear texts. Since the goal of education is to foster learning, and learning involves making connections between ideas, one would assume that a more coherent text would offer an educational advantage over less coherent texts.

However, research by McNamara et al. [18] suggests that high coherence texts might not always be better. In their study, they found that high coherence texts helped all learners for more superficial forms of learning (essentially text memorization). When it came to deep learning (i.e. developing a "situation model" [40] of the text), low knowledge

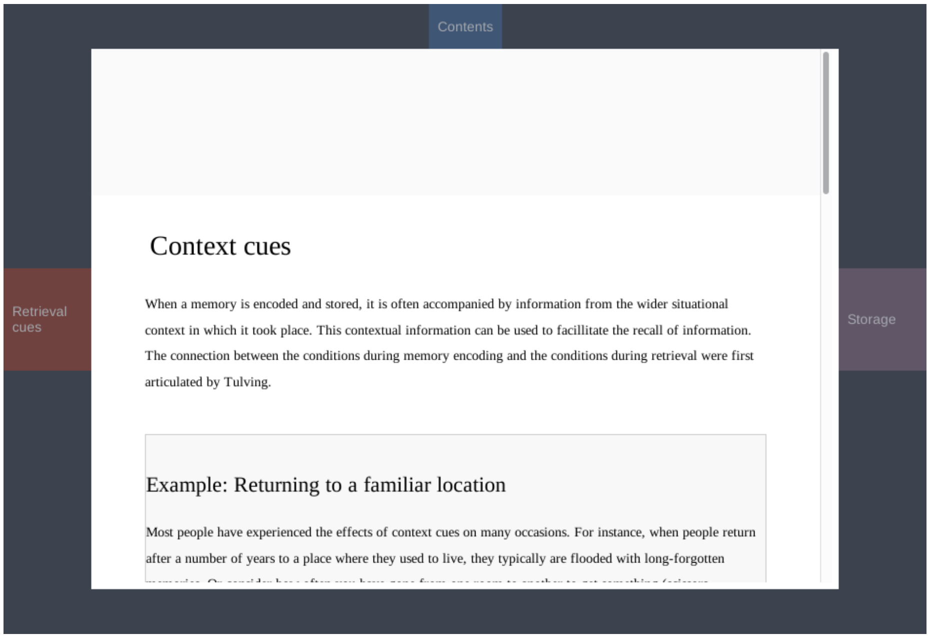

Fig. 7. The linear control interface. The mini-treemap is removed, and peripheral content is placed beneath the core content inside grey boxes.

learners benefited from the high coherence text, but high knowledge learners learned more with a low coherence text.

McNamara et al. argue that high knowledge participants were harmed because they did less active processing when using the high coherence texts. Low knowledge participants, this argument goes, are helped by the high coherence texts because it provides key knowledge that they do not already have. The high knowledge learners learn more from low coherence texts because they are already in possession of the required background knowledge to form a rich situation model, and a low coherence text forces them to pay attention to the relations between the concepts presented in the text.

4.1 Examining the Challenge

Sweller, Van Merrienboer, and Paas [34] offer an alternative explanation for the results shown by McNamara et al. [18]. They say that high knowledge learners demonstrate shallower learning when using high coherence texts because, for them, these texts introduce *extraneous workload*. The information conveying the relations between concepts is redundant for high knowledge learners, and reading this information is therefore a waste of cognitive resources.

If this explanation is accurate, we believe our high coherence text can avoid the pitfalls observed in McNamara et al. because we reduce extraneous workload compared to a standard linear text in three ways.

The first is by *reducing opportunities for thinking about the structure of text that is not also thought about the structure of content*. When navigating a textbook, the students must frequently think about the text's structure. For example, when searching for a particular concept, the student thinks about chapters, sections, and pages. Whenever the structure of the text does not match the structure of the concepts it contains, this is extraneous workload—wasted thought. Since the structure of our non-linear format is more harmonious with the structure of the concepts it contains, whenever this happens the student is learning, and the opportunities for this kind of wasted thought are reduced. The second is by *showing structural relations spatially, not linguistically*. Our non-linear format signals hierarchical information through the spatial position of elements, which we think will interfere less with the reading task—a linguistic activity.

Our study, described in the following section, is designed in-part to examine the potential issues raised by McNamera et al. For this preliminary investigation into the effectiveness of our proposed interface, due to constraints of time and space, we decided to examine only the effects of non-linearity, setting aside the effects of dividing content into core and several kinds of periphery.

5 Methods

With this study we aimed to address our second and third research questions, which are: *does this non-linear textbook better support learning?* and *Do users prefer the proposed non-linear format?* As such, we were interested in comparing our non-linear design with a linear control in terms of learning outcomes, both subjective and objective, usability, and overall likability.

The results of a study by McNamara et al. (see: Section 4) raise the possibility that our high-coherence design may result in shallower learning for high knowledge learners compared to the standard linear format. To address this potential problem, we a) sorted participants into high- and low-knowledge groups according to the results of a test of general knowledge of psychology, and b) asked questions designed to target both deep and shallow forms of learning in our objective learning outcome measures.

Including two rounds of pilot testing, we tested our design with over 40 participants. Only the main study (after pilot testing) is described here. This study was approved by the Carleton University Research Ethics Board.

5.1 Participants

Twenty-six participants (13 female, 13 male) volunteered to participate in this study. Twelve had backgrounds in cognitive science, and the other half had an assortment of backgrounds. Participants were alternately assigned to either the non-linear ($n = 13$) or linear ($n = 13$) condition.

5.2 Materials

Participants read content using either the non-linear format, or the linearized control. The e-textbooks were presented using Mozilla Firefox featuring add-ons to eliminate all

GUI elements except for the main display frame. The System Usability Scale (SUS) [7] was used to gauge participants' perceptions regarding the interface's usability.

Participants completed two multiple choice tests designed by us. The first was taken before the reading session and assessed background knowledge in psychology. The other was taken after the reading session, and assessed knowledge of the material they had just read. Both tests consisted of multiple choice questions (16 and 19 items, respectively). We decided against having overlapping questions between the two tests (i.e. no questions that were asked in the background assessment test were asked again in the post-reading test) to avoid priming participants.

5.3 Procedure

Participants completed a 16-item questionnaire assessing background knowledge in psychology. We later used the results of this test to sort participants into high- and low-knowledge groups.

Next, participants were presented with a demo interface, which was the interface they would be using in the main task except featuring placeholder text. Participants were given a brief verbal description of the interface and how to navigate it, while participants practiced using the interface. Once participants were comfortable with the demo interface, they were presented with the 'real' interface featuring content from an introductory psychology textbook [38]. Participants were asked to read all of the content contained in the interface, which took about 20 minutes. We observed the participant's interactions with the interface indirectly through the laptop's main screen.

After the reading session, participants were given a post-test on the material they read, which featured questions targeting deep and shallow learning. Participants then completed the SUS scale.

Following the completion of the post-task questionnaires, we conducted a semi-structured interview with participants about their experience using the interface, their attitudes towards reading and textbooks, and about their approaches to learning.

Finally, we showed participants the interface they *did not* use in the main task. After giving a quick demonstration, we invited participants to try the interface themselves. We then asked which interface they preferred.

6 Results

The purpose of this study was to address our second and third research questions. We address the findings for each below.

6.1 Answering Research Question 2

To answer the second research question, we asked participants to complete a 19-item multiple choice test of the material they read during the reading session. The test featured three kinds of questions: 11 text based questions, four bridging inference questions, and four problem solving questions. Text-based questions assess the reader's representation of the text (shallow learning). The latter two question types assess the reader's representation of the situation described by the text (deep learning). We are mostly interested in the deep learning questions.

A summary of the results of the post-reading scores is shown in Table 1. We divide participants into two groups, low and high knowledge, based on their performance in the pre-test assessing background knowledge in psychology. Participants who scored greater than the median score (*Mdn*=0.5/1) were placed in the high knowledge group ($n = 13$), and participants who scored less than the median score were placed in low knowledge group ($n = 13$). We compare test scores by question type for low knowledge and high knowledge learners, for both interface conditions. For sake of completeness, we include the combined scores of low and high knowledge participants (under "Combined"), and the overall test scores (i.e. regardless of question type) (in the rows labelled "Overall").

Table 1. Proportion of Correct Responses in the Postreading Test for the Two Conditions by Knowledge and Question Type

	Linear	Non-Linear
High knowledge	$n = 7$	$n = 6$
Text based	.68	.62
Bridging inference	.61	.62
Problem solving	.50	.42
Overall	.62	.58
Low knowledge	$n = 6$	$n = 7$
Text based	.56	.56
Bridging inference	.33	.61
Problem solving	.42	.39
Overall	.48	.53
Combined	$n = 13$	$n = 13$
Text based	.62	.59
Bridging inference	.48	.62
Problem solving	.46	.40
Overall	.56	.55

A number of one-tailed independent samples t-tests were conducted to compare the performance of high and low knowledge participants in the two conditions for each type of question (text-based, bridging inference, and problem solving), where "performance" means the proportion[1] of correct scores on the post-reading test. For all combinations of question types and knowledge levels, *there were no significant differences between the linear condition and the non-linear condition.*

Finally, we examined the learning effects for the two conditions, where 'learning effects' means the post-reading test score minus the psychology background knowledge

[1] We treat the scores as continuous data.

test score. Boxplots of the learning effects by condition are shown in Fig. 8. Learning effects were greater in the non-linear condition than in the linear condition. However, the notches indicating the 95% confidence interval around the median overlap, suggesting that these results are not significant.

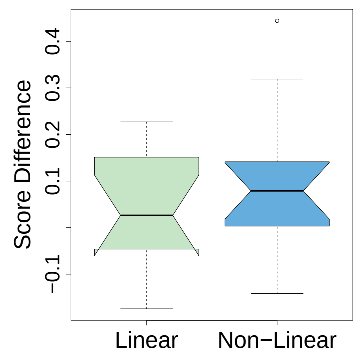

Fig. 8. Learning effects by condition

6.2 Answering Research Question 3

System Usability Scale Scores. To see if there was any difference in perceived usability between the two conditions, we asked all participants to complete the SUS questionnaire, and then compared the results for each condition. The SUS is generated from Likert-scale questions, which means we look for differences between the scores for the two conditions using the non-parametric Wilcoxon rank-sum test. The Wilcoxon rank-sum test indicated that there was no significant difference between the SUS scores for the linear condition (*Mdn*=50.5) and the non-linear condition (*Mdn*=50), U=71.5, p=0.5.

Interface Preference. During debriefing, at the end of the study session, participants were shown and given a description of the interface they did not use, and were asked which they preferred. 18/25 (72%) of participants said that they preferred the non-linear interface to the linear interface. A chi-square test of independence showed that the difference between the total number of participants who preferred the non-linear interface ($n = 7$) was significant, χ^2 (1, N=25)=8, p=.005.

6.3 Post-Experiment Interview: Key Themes

Participants who preferred the non-linear interface tended to appreciate its organization, with several noting that it helped show how the various ideas "fit together". Many participants appreciated that the text was categorized into smaller units than one would typically find in a standard textbook. Many participants noted difficulties with the standard prose format, where content is provided in big, relatively undifferentiated blocks of text.

Some felt that this format was “intimidating”, and others felt going through many pages of these big blocks of text to be “monotonous” and “boring”, and that these texts seem to “go on and on”. A number of participants noted that the ideas presented in standard texts tended to get mixed up in their heads, and that they felt that the larger organizing structure provided by the treatment condition would help prevent this from happening. They found that organizing the content in a hierarchical tree made it appear more manageable. A few of the participants who preferred the treatment interface said that they did not enjoy reading in general, and that this interface made reading less difficult for them.

We asked a number of participants if they preferred text in textbooks to appear in smaller or larger chunks, and all said that they preferred smaller chunks of text. We thought this result was favorable to our non-linear design, which breaks up text into smaller chunks than standard linear texts as a consequence of the division of content into *core* and three kinds of *peripheral*, and keeping this content in separate locations in the UI.

Participants who preferred the linear interface tended to like its straightforwardness, and found the non-linear interface “confusing”. They generally seemed to find the freedom of choice in terms of navigation overwhelming.

7 Discussion

We observed no significant differences between the two conditions in terms of objective or self-reported learning outcomes. Yet, a significantly greater number of participants preferred the non-linear interface to the linear interface.

Contrary to our expectations, we did not find that the non-linear interface offered a learning advantage. In this section, we discuss what we feel are the most likely reasons for this. Overall, we suspect that the reading session was too short and the amount of text participants read was too little for the advantages of the non-linear interface to become apparent.

Not enough time for reflection. The treemap allows readers to see how all of the ideas in a domain “fit together”, which should be good for deep learning. We think that participants were too busy reading to take advantage of this benefit. Each reading session was relatively short: approximately 20 minutes, and the amount of content contained in the interfaces meant that participants had to read at a relatively quick pace. In all except one case they did not have enough time to revisit content, let alone to reflect on how things “fit together”.

The costs of learning a new interface outweighs the benefits. Our non-linear interface is a new way of viewing and interacting with text, whereas the linear text was modeled on pre-existing e-textbooks, which in turn are modeled on physical print media. Learning something new, whether it is text or how to use a new interface, requires the deployment of working memory resources. Our text required more learning, which means there are fewer resources available for learning the content. It is possible that the benefits offered by the non-linear organization of content was overshadowed by the costs associated with learning the new interface.

Affective advantages negated. We think that some features of our non-linear interface would offer an affective advantage. For example, the non-linear interface does not enforce any particular path through the text: students must decide for themselves which path to take. This may give students an increased sense of agency, which could imbue their traversal through the material with more personal meaning. Our interface also breaks up text into smaller pieces than standard linear texts, which seemed to be a relief for a number of our participants who felt overwhelmed by large blocks of text. Learning outcome effects of affective advantages like these would develop over longer periods of time, and in any case would likely be too subtle to be detected by the data gathering methods we employed.

Small sample size. Our main study had 26 participants, which meant there were only six or seven participants in each knowledge level-interface condition group. The study by McNamara et al. [18], which we modeled our study after, had 56 participants—more than twice the amount. Perhaps increasing the total number of participants would have yielded significant results.

8 Conclusion

In this paper, we presented the design, implementation, and subsequent study and analysis of a novel non-linear e-textbook format. Our design was guided by the cognitive science literature on concepts. Out e-textbook format differs from standard textbooks by organizing content non-linearly, and by making a strict distinction between core and peripheral content. We conducted a user study where we compared our non-linear design with the standard format in terms of learning outcomes, usability, and overall likeability. We found no significant differences between the two interfaces for learning outcomes and usability, and found that our interface was better-liked by participants. We feel that many of the advantages of our non-linear format will only become apparent after long periods of use.

We were able to design and implement a functional prototype concording with the cognitive science literature on the structures of concepts. We are encouraged by the fact that participants preferred this novel non-linear design, and that the two interfaces seemed to provide similar support for learning in spite of the fact that the non-linear design was unfamiliar. For many participants, this design seemed to fulfill a need that linear texts were not providing: many students are looking for a non-linear alternative to the standard format. Future work should examine the effects of this non-linear format over longer periods of use.

References

1. Abbott, V., Black, J.B., Smith, E.E.: The representation of scripts in memory. Journal of Memory and Language 24(2), 179–199 (1985)
2. American Association for the Advancement of Science: AAAS's Project 2061 report: Big biology books fail to convey big ideas (Jun 2000), https://www.eurekalert.org/pub_releases/2000-06/AAft-AP2r-2606100.php

3. Atkinson, R.K., Derry, S.J., Renkl, A., Wortham, D.: Learning from examples: Instructional principles from the worked examples research. Review of educational research 70(2), 181–214 (2000)
4. Barsalou, L.W., Simmons, W.K., Barbey, A.K., Wilson, C.D.: Grounding conceptual knowledge in modality-specific systems. Trends in cognitive sciences 7(2), 84–91 (2003)
5. Barsalou, L.W., Wiemer-Hastings, K.: Situating abstract concepts. Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thought pp. 129–163 (2005)
6. Bleazby, J.: Overcoming relativism and absolutism: Dewey's ideals of truth and meaning in philosophy for children. Educational philosophy and theory 43(5), 453–466 (2011)
7. Brooke, J., et al.: SUS - a quick and dirty usability scale. Usability Evaluation in Industry 189(194), 4–7 (1996)
8. Budiansky, S.: The trouble with textbooks. ASEE Prism 10(6), 24 (2001)
9. Daniel, D.B., Woody, W.D.: E-textbooks at what cost? performance and use of electronic v. print texts. Computers & Education 62, 18–23 (2013)
10. Deese, J.: On the structure of associative meaning. Psychological Review 69(3), 161 (1962)
11. Desilver, D.: Pew Research: U.S. students' academic achievement still lags that of their peers in many other countries (2017), http://www.pewresearch.org/fact-tank/2017/02/15/u-s-students-internationally-math-science/
12. DeStefano, D., LeFevre, J.A.: Cognitive load in hypertext reading: A review. Computers in human behavior 23(3), 1616–1641 (2007)
13. Edgcomb, A., Vahid, F., Lysecky, R., Lysecky, S.: Getting students to earnestly do reading, studying, and homework in an introductory programming class. In: Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education. pp. 171–176. ACM (2017)
14. Gibson, J.J.: The ecological approach to visual perception: classic edition. Psychology Press (2014)
15. Johnson-Laird, P.: Mental models. toward a cognitive science of language, inference and language (1983)
16. Jonassen, D.H.: On the role of concepts in learning and instructional design. Educational Technology Research and Development 54(2), 177 (2006)
17. Keil, F.C.: Explanation, association, and the acquisition of word meaning. Lingua 92, 169–196 (1994)
18. McNamara, D.S., Kintsch, E., Songer, N.B., Kintsch, W.: Are good texts always better? interactions of text coherence, background knowledge, and levels of understanding in learning from text. Cognition and Instruction 14(1), 1–43 (1996)
19. Miller, B.N., Ranum, D.L.: Beyond PDF and ePub: toward an interactive textbook. In: Proceedings of the 17th ACM annual conference on Innovation and technology in computer science education. pp. 150–155. ACM (2012)
20. National Research Council, et al.: A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. National Academies Press (2012)
21. Ogievetsky, V., Heer, J., Bostock, M.: D^3 data-driven documents. IEEE Transactions on Visualization & Computer Graphics 17, 2301–2309 (09 2011), doi.ieeecomputersociety.org/10.1109/TVCG.2011.185
22. Project 2061: Benchmarks on-line: About benchmarks (2009), http://www.project2061.org/publications/bsl/online/index.php?intro=true
23. Ranzijn, F.J.: The effect of the superordinate concept and presentation form of examples on concept learning. Computers in human behavior 5(2), 95–105 (1989)
24. Rockinson-Szapkiw, A.J., Courduff, J., Carter, K., Bennett, D.: Electronic versus traditional print textbooks: A comparison study on the influence of university students' learning. Computers & Education 63, 259–266 (2013)

25. Rosch, E.: Principles of categorization. Concepts: Core Readings 189 (1999)
26. Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., Boyes-Braem, P.: Basic objects in natural categories. Cognitive Psychology 8(3), 382–439 (1976)
27. Roseman, J.E., Herrmann-Abell, C.F., Koppal, M.: Designing for the next generation science standards: Educative curriculum materials and measures of teacher knowledge. Journal of Science Teacher Education 28(1), 111–141 (2017)
28. Schwanenflugel, P.J., Shoben, E.J.: Differential context effects in the comprehension of abstract and concrete verbal materials. Journal of Experimental Psychology: Learning, Memory, and Cognition 9(1), 82 (1983)
29. Shepperd, J.A., Grace, J.L., Koch, E.J.: Evaluating the electronic textbook: Is it time to dispense with the paper text? Teaching of Psychology 35(1), 2–5 (2008)
30. Shield, M., Dole, S.: Assessing the potential of mathematics textbooks to promote deep learning. Educational Studies in Mathematics 82(2), 183–199 (2013)
31. Sloman, S.A., Rips, L.J.: Similarity as an explanatory construct. Cognition 65(2-3), 87–101 (1998)
32. Spero, E., Stojmenović, M., Arya, A., Biddle, R.: Creating a non-linear textbook format to facilitate deep learning. In: INTED2019 Proceedings. 13th International Technology, Education and Development Conference, IATED (2019), forthcoming
33. Stern, L., Roseman, J.E.: Can middle-school science textbooks help students learn important ideas? findings from project 2061's curriculum evaluation study: Life science. Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching 41(6), 538–568 (2004)
34. Sweller, J., Van Merrienboer, J.J., Paas, F.G.: Cognitive architecture and instructional design. Educational Psychology Review 10(3), 251–296 (1998)
35. Toukonen, K.: The dynamic electronic textbook: Enhancing the student's learning experience. Ph.D. thesis, Kent State University (2011)
36. Tversky, B.: Parts, partonomies, and taxonomies. Developmental Psychology 25(6), 983 (1989)
37. Tversky, B., Kugelmass, S., Winter, A.: Cross-cultural and developmental trends in graphic productions. Cognitive psychology 23(4), 515–557 (1991)
38. Weiten, W., McCann, D.: Psychology: Themes and Variations, 4th Canadian ed. Nelson Education Limited (2015), https://books.google.ca/books?id=Z6vEjwEACAAJ
39. Zacks, J.M., Tversky, B.: Event structure in perception and conception. Psychological Bulletin 127(1), 3 (2001)
40. Zwaan, R.A., Radvansky, G.A.: Situation models in language comprehension and memory. Psychological Bulletin 123(2), 162 (1998)
41. zyBooks: zyBooks: about us (2018), http://www.zybooks.com/about-us/