Investigating the Role of User Engagement in Digital Reading Environments

Heather L. O'Brien
School of Library, Archival and Information Studies
University of British Columbia
Vancouver, Canada
1-604-822-6365
h.obrien@ubc.ca

Luanne Freund
School of Library, Archival and Information Studies
University of British Columbia
Vancouver, Canada
1- 604-822-0825
Luanne.Freund@ubc.ca

Richard Kopak
School of Library, Archival and Information Studies
University of British Columbia
Vancouver, Canada
1- 604-822-2898
r.kopak@ubc.ca

ABSTRACT
User engagement is recognized as an important component of the user experience, but relatively little is known about the effect of engagement on the learning outcomes of such interactions. This experimental user study examines the relationship between user engagement (UE) and comprehension in varied academic reading environments. Forty-one university students interacted with one of two sets of texts presented in 4 conditions in the context of preparing for a class assignment. Employing the User Engagement Scale (UES), we found evidence of a relationship between students' comprehension of the texts and their degree of engagement with them. However, this association was confined to one of the UES subscales and was not consistent across levels of engagement. An examination of additional variables found little evidence that system and content characteristics influenced engagement; however, we noted that all students' reported increased knowledge, but topical interest for non-engaged students declined. Results contribute to existing literature by adding further evidence that the relationship between engagement and comprehension is complex and mediated.

Keywords
User engagement; comprehension; digital reading environments

1. INTRODUCTION
In digital information environments, user engagement (UE), "a user's response to an interaction that gains, maintains and encourages attention" [18], is viewed as a positive and necessary component of human information interaction. It is a widely held belief that UE mediates outcomes such as knowledge acquisition, yet there is conflicting evidence to support this claim. While we do know that UE is pivotal in fostering long-term relationships between users and systems [43], most work has focused on users' interaction preferences. For example, [19] noted that users preferred more "engaging" educational multimedia systems, but that modality preferences were mediated by task type (searching versus browsing). Other researchers have focused on the role of UE in the adoption of learning technologies [17, 44] and the effect of interactive computer-mediated environments on learning [36, 10]. Yet drawing any larger conclusions about the relationship between UE and learning is inhibited by a lack of consistency in how researchers define and measure both UE and learning.

CHIIR'16, March 13–17, 2016, Carrboro, North Carolina, USA

ACM 978-1-4503-3751-9/16/03…$15.00.
DOI: http://dx.doi.org/10.1145/2854946.2854973

Research on human information interaction and retrieval has tended to focus on user-system interaction outcomes such as the retrieval of relevant documents or user satisfaction, rather than comprehension or learning per se. While some studies measure self-reported topical knowledge and search expertise to document cognitive changes during the search process [20], more robust methods are needed to ascertain the extent to which users develop cognitively and affectively through their interactions with systems. Comprehension, the process of extracting meaning from information [30], is a critical indicator of success for information systems, yet largely under-researched in interactive information retrieval (IIR).

In this research, we focused specifically on comprehension as an outcome of interacting with digital texts. Drawing upon prior research on the design of digital reading environments [36, 23, 26], we created a reading interface with different modes of interaction (note-taking and annotation) and text representation. We investigated whether UE affected comprehension outcomes and whether features of these reading environments influenced UE. Our initial conjecture was that more engaged readers would score higher on comprehension tests and that more interactive and visually rich reading environments would be more engaging. While there was a relationship between UE and comprehension, we found that it was more nuanced than we predicted, leading us to explore the role of user, system and content characteristics to interpret the findings. In this paper, we review relevant prior research and outline our methodology, and then describe our findings in detail. We conclude with a discussion of the findings and illuminate future directions for examining the relationship between UE and comprehension in human information interaction research and applications.

2.PRIOR RESEARCH

2.1. User Engagement (UE)
This research builds upon the work of [33] that defined user engagement as a multidimensional construct comprising the interaction between cognitive (e.g., attention), affective (e.g., emotion, interest), and behavioral (e.g., propensity to re-engage with a technology) characteristics of users, and system features (e.g., usability). UE is characterized by a number of interrelated factors: attention, motivation, control, hedonic and utilitarian

needs, perceived time, attitudes, interest, sensory and aesthetic appeal, interactivity, novelty, challenge and feedback [19, 33]. Researchers are developing and applying various self-report, analytic and physiological methods to capture UE [24]. The User Engagement Scale [35] has been used extensively to investigate UE in various computer-mediated environments, including search, educational technologies, haptic interfaces, and social and communication media, and has found to be reliable and valid [31].

Within computer-mediated environments UE can be distinguished from *student engagement*, which looks at the extent to which a student is involved and invested in his or her own learning [13]. Research on student engagement has shown that more engaged students are more likely to succeed academically, and less likely to drop out of school [12]. In this study, we acknowledge that student engagement serves as the broader context in which UE with digital learning applications takes place. Whereas student engagement measures socio-economic, classroom and student variables that contribute to academic success, UE is more concerned with system (e.g., usability) and user (e.g., motivation) attributes as they relate to interactions with digital media. UE is equated with "stickiness," that is, continued engagement or re-engagement with systems [33, 43]. However, continued engagement with digital technologies cannot be assumed to foster outcomes such as comprehension and learning: the UE "equation" is not well-defined: unlike student engagement, we do not have a good understanding of what predicts UE or its outcomes, and how this varies across different types of systems.

Yet, the general consensus is that UE fosters better outcomes: more products sold, increased loyalty to a search engine, more invested e-learners, etc. Furthermore, interactive system features (e.g., links), embedded multimedia (e.g., video), and social annotation (e.g., "likes" or comments embedded within news websites) are believed to foster UE [3]. However, there is a dearth of literature that explicitly examines the relationships between UE, system features and user outcomes. Does greater engagement with digital systems make us better learners, more productive employees, or more efficient information seekers? In the current study we focus specifically on the relationship between UE and comprehension in a digital reading environment and the role UE plays in facilitating the understanding of texts.

2.2. User Engagement and Digital Learning Environments

Studies of digital learning environments are numerous and span multiple disciplines. Here we focus on a few examples that have explored UE within learning systems. These studies examined the presentation of content, compare different kinds of digital media, or contrast computer and non-computer mediated media. Webster and Ho [44] for instance, sought to determine characteristics of presentation software that were valued by students (challenge, variety, control and feedback), and whether these characteristics could be manipulated in the design of a computer-mediated lecture. The authors captured students' design preferences for presentation software over the duration of two iterative studies, but did not look specifically at the relationship between UE and learning outcomes, e.g., students' comprehension or retention of content. More recently, Denny [7] found that badge-based achievement systems lead to increased participation in an online course management system, as measured by the number of days students interacted with the system and the number of questions posed and answered by students. Denny concluded that badges increased student motivation, but made little difference in student learning, based on the number of questions formulated by students. Results were inconclusive regarding the relationship between learning, UE and interactions with the course management system.

Other work has focused on comparing "human" versus computer-mediated instruction. Hu and Hui [17], for example, compared face-to-face versus computer-mediated (i.e., 20-minute video) instruction on students' ability to perform tasks using Adobe Photoshop. They found that face-to-face instruction resulted in more learning activities than video instruction. Yet, regardless of the medium of instruction, learning engagement predicted learning effectiveness and learning satisfaction. In other words, engaged students were more effective learners, independent of the medium of interaction. Sung, et al. [40] explored the depth of students' interactions with museum exhibits, comparing use of an interactive guidebook with a print worksheet. Results indicated longer and more positive interactions with museum exhibits for those in the interactive guidebook condition; however, there were no differences in learning outcomes as measured by concept maps created by students. Taking another approach, Jacques, et al. [19] drew upon the findings of four studies of multimedia learning software and concluded that users' perceptions of media depended on its type (e.g., videos, animations, photographs), presentation (e.g. aesthetic qualities such as font, color, sound quality) and their perceived control when using the system. However, perceptions were mediated by the user's interaction task and level of interest.

In summary, studies have manipulated how material is presented and showed that students preferred specific modes of interaction or design features, but have not found significant differences in terms of the quality of the output generated by students (concept maps [40], or discussion questions constructed [7]). In light of such findings, it is clear we need not only to understand the effects of specific technologies, but also how individual (e.g., motivation) and contextual (e.g., content, task) factors interact with system characteristics to produce experiential outcomes.

2.3. Digital Reading Environments

The shift from reading in print to reading in digital formats has received considerable attention in recent decades. Early work demonstrated that reading from screens was slower, less accurate, and resulted in performance and comprehension deficits in comparison to print reading [9]. More recent studies have characterized digital reading as non-linear, selective and antithetical to deep and sustained attention [6], running counter to the goals of "active reading", as framed by Adler [1]. Support for active and focused reading has been identified as an important challenge in designing digital reading environments [36]. Hence, many digital reading environments attempt to enable active reading by implementing various modes of interaction, e.g., browsing, scanning, skimming and keyword extraction [6, 25) through affordances such as note-taking, linking and annotating. Interfaces are designed with attention to features such as page layout, legibility, typology [26, 36] and structural cues (e.g. hierarchies, link definitions, etc.) [29]. The main thrust of these efforts is to help readers transition from reading to writing [29] and, more importantly, to engage in active reading that facilitates comprehension. Much of this work is based on an implicit assumption that heightened user engagement in a reading task will foster greater learning outcomes.

Comprehension is the process of extracting meaning from information [30] and is an essential component of learning from texts. Of the many models of text comprehension utilized in this context, Kintsch's Construction-Integration (C-I) Model is widely referenced and well-suited for investigating comprehension in information interaction research [4; 21]. In this model, comprehension results from a multi-staged process where propositions are first formed into a plausible interpretation of linguistic input [construction of a textbase], and then integrated with existing knowledge into a situation model. The text-base relies on micro-level (e.g., decoding, remembering facts) and macro-level (e.g., more global processing of the gist or topic) inferences within the text. Reading consists of increasingly sophisticated parsing of text-based propositions to form a mental representation of the content. This, coupled with prior knowledge and experience, evokes imagery, emotions, and intentions that create the "situation model" and guide current and future text interactions [22].

The C-I Model has been previously applied in IIR research. Cole and Mandelblatt [4] devised an information retrieval (IR) model based on Shannon and Weaver's Communication System Model with three sub-systems: perception, comprehension, and application, and mapped these sub-systems to Kintsch's layers of mental representation (e.g., propositions, micro-level representation, macro-level representation, situation model). At the perceptual level, the information searcher must perceive the signal or series of propositions from the IR system, input the information, and encode it in such a way that it can be processed by the comprehension sub-system. The comprehension system receives this microstructure output and transforms it into macrostructure output and situation model. Finally, the situation model is applied when it reaches the application sub-system. At this stage, the searcher is able to use his/her altered cognition to communicate with the IR system, for example, by refining query terms to achieve more relevant results. The C-I Model was used to guide the design of an information system for undergraduate students in the preliminary stages of their research. It was believed that helping students to articulate their information need and represent it to the IR system would aid in promoting more successful searching behaviors.

While [4] saw the potential of Kintsch's work in IR system development, we envisioned it could assist with the evaluation of IIR experiences, both in terms of its ability to help us construct an instrument to measure comprehension and to understand the relationship between learning outcomes and readers' perceptions of enjoyment and engagement.

Utilizing the C-I model of comprehension in this study, we hypothesized that one way that UE may contribute to comprehension and learning is through its relationship with the concept of flow, a state characterized by deep absorption in an activity [5]. UE shares qualities with flow, including intrinsic motivation, focused attention and control [19]. Kintsch [22] articulated the relationship between learning and flow saying, "flow is not the goal of instruction, learning is, which is hard work, but the flow may provide motivation to engage in the hard work of deliberate practice" (p. 230). Thus UE, like flow, may be a mediating variable in the learning equation, inspiring learners to invest in the "hard work" of learning.

2.4. Current Study

UE is a desirable goal for interactive systems and is believed to foster positive user outcomes. The reviewed literature suggests that people interact more in more engaging computer-mediated environments. But such systems do not necessarily affect greater gains in learning [7, 40]. Student motivation [17], interest, and the nature of the learning task [19] may intervene between UE and learning outcomes.

One of the dilemmas in documenting learning outcomes is that studies are focused on different technology-mediated interactions (presentation software, course management software), settings and time frames. The latter is particularly crucial if we consider that learning is a process. Another issue is the variety of measures employed in attempting to measure learning. Some researchers have focused on perceptions of UE and learning, rather than on more objective indicators. Comprehension, which is of interest to us, is typically assessed through fact-based questions in multiple-choice or true and false formats that ask readers to recall information or make inferences [27] or to create written text summaries [39]. A major challenge with comprehension measures is that the variable content of the learning environment makes it impossible to develop standardized metrics.

We adopted Kintsch's C-I Model to guide us in the development of suitable measures of micro (local) and macro (global) level comprehension. We created a digital reading environment with four reading conditions that varied in interactivity (note-taking, linking) and text presentation (plain versus in-context cues). Based on the inconsistencies in the literature regarding the relationship between UE and learning outcomes, we sought to address the following questions: Do more engaged readers achieve better comprehension outcomes? What role do user, system and content characteristics play in facilitating UE?

3. METHOD

This study was conducted with 41 student participants. We designed a simulated task scenario based on a typical academic task of preparing for class by completing assigned readings. Readings were focused on a general interest "Technology and Society" theme. Students interacted with three digital readings of mixed genres (journal article, website, and popular press article) from one of two article sets: Digital Activism (7874 words) and Human-Robot Interaction (8325 words). Groups of students were assigned to interact with one article set, and with one of four versions of the experimental reading interface (see Table 1). The reading interface manipulated the level of interactivity (presence or absence of interactive features) and presentation style ("plain text" or "in-context"). The interactive conditions included a small number of hyperlinks that connected relevant ideas within the reading set, and had the Diigo Toolbar installed so that participants could annotate and highlight text; the non- interactive interface had neither of these features. The "in-context" condition presented the texts in their original format together with images and any original formatting features, whereas the "plain text" interface displayed articles in a single font and standard format with no images. Articles were presented as individual Web pages in a browser with a vertical scrollbar available for within-page navigation. A menu page contained the task description and hyperlinks to each article in the set and was used to navigate between readings.

3.1. Participants

The students recruited for this study (21 F; 17 M; 4 not specified) were 19-24 (n=21) and 25-29 (n=14) years old. There were undergraduate (n=22), masters (n=10) and doctoral students (n=5) who were enrolled in a range of degree programs; four did not specify student status. The majority reported spending more than 30 minutes per day reading for academic (n=39) or personal (n=30) purposes. Participants used digital media about half the time for academic reading (M= 49%; range: 10-95%).

Table 1. Summary of experimental conditions

Interface Condition	Human Robot Interaction	Digital Activism
Plain text with interactive features	Group 1 (*n*=5)	Group 2 (*n*=4)
Plain text without interactive features	Group 3 (*n*=5)	Group 4 (*n*=6)
In-Context with interactive features	Group 5 (*n*=5)	Group 6 (*n*=6)
In-Context without interactive features	Group 7 (*n*=5)	Group 8 (*n*=5)

3.2.Instruments

We used a variety of performance, self-report, and comprehension measures. In this paper, we focus specifically on the self-report measures for UE, topical knowledge, interest, and motivation, and comprehension. Students were asked to rate their knowledge and interest (4-items) on a 7-point Likert scale before and after the reading task. These were intended to gauge whether students perceived changes in their topical knowledge or interest as a result of interacting with the texts.

The User Engagement Scale (UES), consisting of 31-items rated on a 7-point Likert scale, was used to measure UE [35]. The UES comprises six sub-scales: perceived usability, aesthetic appeal, felt involvement, novelty, focused attention and endurability. It has been used to measure UE in computer-mediated domains such as web search, e-commerce, social networking systems and online news, and has demonstrated good reliability and validity; however, the use of factor analysis in some studies has pointed to a four-factors, rather than a six-factor UES [34, 46]; we elaborate on the implications of this below in the "data analysis and preparation" section.

The Adult Reading Motivation Scale (ARMS) (21 items rated on a 5-point Likert scale) was used to investigate attitudes toward reading [38]. It includes dimensions of reading for recognition, reading to do well in other areas of life, reading for efficiency, and reading as part of one's self identity. Motivation has been shown to precipitate UE [19, 33, 44], and is an affective component of learning experiences. The Nelson-Denny Standardized Reading Test was used to assess students' general reading level [2], and allowed us to collect baseline reading ability data of participants.

We developed and refined a reading comprehension test based on Kintsch's C-I model [21]. Comprehension questions for each article set were designed to assess microstructural and macrostructural text comprehension. For all of the comprehension questions there were "correct" and "incorrect" responses. *Microstructural items* tested recall (eight factual true or false items) and understanding of the concepts presented in the texts through the Sentence Verification Technique (SVT) [37]. The SVT for each set of articles consisted of twelve sentences (four per article), two of which accurately represented the semantics of the text in the form of one exact phrase and one paraphrase, and two of which falsely represented the content of the article by changing the meaning or inserting something incongruous. Participants were asked to indicate whether or not each sentence accurately represented the content. *Macrostructural items* asked students to connect ideas presented in the three texts and consisted of six summary statements (SS) that varied in centrality and importance [21, 42]. Students were asked to select the three statements that best depicted the central themes in the article set.

3.3. Procedure

Each two-hour session took place in a seminar room. Morae software recorded on-screen activity, including URLs visited (task menu and articles), time spent on each page, scrolling, typing, and mouseclicks. Laptops were pre-set to display the texts in one of the four conditions per session. All other materials (instructions, questionnaires) were provided in print format. A researcher was present to greet students and facilitate the session. Following completion of the informed consent procedure, students (in groups of 4-6) completed a demographic and reading habits questionnaire, and the ARMS. Following this, those in the interactive conditions participated in a tutorial of the Diigo Toolbar. Next, participants were asked about their pre-task knowledge and interest in the topic of the assigned article set. They then moved on to the reading task, beginning from an online menu page that contained the task scenario and links to the three articles. They were asked to browse the readings within a 30-minunte time frame. Immediately after the reading task, students completed a post-task questionnaire consisting of items related to their knowledge of and interest in the topic *after* interacting with the readings, the UES, and the comprehension test related to each reading set. Lastly, students completed the Nelson-Denny Standardized Reading Test. At the conclusion of the study, participants were thanked and paid an honorarium.

4. DATA ANALYSIS AND PREPARATION

4.1. User Engagement Scale (UES)

Responses to the UES were first examined to assess data quality and consistency of the sub-scales. Missing values ranged from 0 to 2 per item. The exception to this was one aesthetic appeal item ("I liked the graphics and images in this set of readings") that had nine missing items. The lower response rate was likely due to the fact that readings did not contain graphics and images in some conditions; this item was removed from further analysis.

A standard approach when using an experiential scale such as the UES would be factorization to test its robustness before proceeding to data analysis. Factorization was not appropriate here given the small sample size. Therefore, we based our groupings of items on prior work [34, 46] that recommend the use of four-factors, rather than six, as most appropriate for the UES. Specifically, three of the six sub-scales (perceived usability, aesthetic appeal, and focused attention) have consistently resulted in distinct factors across various administrations of the UES, but the endurability, felt involvement, and novelty sub-scales have tended to combine to form one factor, which we refer to here as "Interest" (INT). Thus, we grouped the 31 items as follows: 1) endurability + felt involvement + novelty (INT) (2) perceived usability (USAB), 3) aesthetic appeal (AESTH), and 4) focused attention (ATTEN), and checked the reliability of these groupings.

AESTH and ATTEN had excellent Cronbach's alpha values [7]. The value of USAB was not optimal (0.66). Through an examination of the inter-item correlations for this sub-scale, two items were removed ("I could not do some of the things I needed to do during this reading experience" and "This experience was mentally taxing"); internal consistency was subsequently improved (0.704). Although INT had a good Cronbach's alpha value (0.9), the Endurability item, "This reading experience did not work out the way I had planned," was poorly correlated with other items in the grouping, and was therefore removed. This increased the alpha value to 0.914, which indicated some redundancy. One further item was removed ("This experience was worthwhile") after examining the inter-item correlations to reduce the amount of overlap for this factor. Thus 26 items were used to create four groupings. The descriptive and correlation statistics for the sub-scales are summarized in Tables 2 and 3, respectively, for the four UES groupings.

Table 2: Descriptive statistics of UES item groupings

UES	*M* (*SD*)	# items	α	Skewness (*SE*)	Kurtosis (*SE*)
INT	4.37 (0.96)	9	0.898	0.075 (0.338)	-0.218 (0.75)
USAB	5.25 (0.81)	6	0.704	-0.06 (0.374)	-0.254 (0.733)
AESTH	4.01 (1.24)	4	0.84	0.147 (0.378)	-0.146 (0.741)
ATTEN	2.51 (0.73)	7	0.894	-0.204 (0.378)	-0.593 (0.741)

Table 3: Correlations for UES groupings (Sig. at p<0.001)**

UES	INT	USAB	AESTH
INT	1		
USAB	0.512**	1	
AESTH	0575**	0.348**	1
ATTEN	0.74**	0.206	0.291

Score distributions were then examined to determine appropriate approaches to data analysis. Kurtosis statistics showed a fairly flat distribution for all of item groupings, but INT and AESTH were positively skewed, whereas USAB and ATTEN were negatively skewed. The mean USAB score was high, whereas ATTEN was quite low (based on the seven-point Likert scale range). This analysis confirmed it was best to examine sub-sets of the UES as opposed to calculating a composite score for engagement based on all 26 items, and, given these findings and the small sample size, that non-parametric statistics were most appropriate.

We examined the validity of the UES by looking at the association between the UES item groupings and a single post-task self-report item: "What was your level of engagement with the content of the articles that you read for the experimental task?" Spearman's rank correlation coefficient showed positive associations between responses to the level of engagement question and the UES item groupings: INT (r_s=0.662, p=0.00), ATTEN (r_s=0.5, p=0.001), and USAB (r_s=0.48, p=0.002); AESTH was marginally significant (r_s=0.309, p=0.056). These correlations validated the use of these groupings in subsequent analyses.

4.2. Comprehension assessments

Table 4 shows descriptive statistics for the comprehension measures. Responses to the comprehension questions were scored by comparing them with the correct responses to the true or false (T/F), sentence verification technique (SVT) and summary statements (SS) questions. We looked at the number of errors made by students for each of the T/F, SVT and SS assessments. Scores were calculated by subtracting the number of errors from the maximum number of correct answers. Since there were different maximum scores for each test based on the number of questions, we attempted to give each test equal weight in computing the microstructural and total comprehension scores. For *microstructural comprehension*, T/F and SVT each comprise half of the score, and T/F, SVT and SS were equal to a one third of the *total comprehension score*. These latter measures are expressed as percentages.

Table 4. Descriptive statistics for total comprehension, true/false, SVT and summary statements

Comprehension	*M*(*SD*)	Median	Range	Max Score
T/F	6.04 (1.53)	6	3-8	8
SVT	7.92 (1.63)	8	4-11	12
SS	5.09 (1.13)	6	2-6	6
T/F + SVT	70.83 % (14.39)	72.91%	41.67-95.83%	100%
T/F + SVT + SS Score	73.35 % (12.59)	76.92%	46.15-96.15%	100%

The UES results were converted into low, medium and high engagement groups using the percentiles of each variable based on the median. This approach was chosen given the non-normal distribution of the item groupings, as previously described. As shown in Table 2, the mean for ATTEN was low (2.51) compared to USAB (5.01); AESTH and INT closer to the middle of the seven-point Likert scale range.

5. RESULTS

5.1.User Engagement and Comprehension

A summary of the means and standard deviations for the overall comprehension score for the three levels of engagement on each of the factors is presented in Table 5. The means vary by group, but a strong general trend is not discernable. Kruskal Wallis Tests show a significant difference in the Total Comprehension score for the INT factor, but not for ATTEN, USAB, and AESTH. A more fine-grained analysis of comprehension outcomes across levels of INT (Table 6) shows that scores on the True/False questions varied significantly (P=0.052), which affected the overall Microstructural comprehension (T/F + SVT) and Total Comprehension scores (Table 5). The SVT and SS scores did not vary significantly. The means and standard deviations indicate a non-linear relationship in which the T/F scores were higher for the High and Low INT groups than in the Medium group. A similar pattern repeats across all comprehension measures, with the highest mean scores achieved by those with the lowest level of INT.

Table 5. Descriptive statistics for total comprehension for three levels of user engagement

	INT M (SD)	AESTH M (SD)	ATTEN M (SD)	USAB M (SD)
Low	76.92 (12.3)	70.9 (14.09)	73.84 (10.69)	71.94 (13.44)
Medium	62.01 (12.4)	78.84 (10.13)	78.32 (11.7)	71.67 (12.55)
High	72.43 (9.25)	69.93 (10.71)	70.19 (11.72)	75.96 (12.06)

Table 6. Descriptive statistics (M, SD) and Kruskal Wallis Test (df=2) for low, medium and high INT for comprehension measures

INT	T/F	SVT	SS	T/F + SVT	Total Comprehension
Low	6.25 (1.43)	8.37 (1.45)	5.37 (1.2)	73.95 (13.6)	76.92 (12.40)
Medium	4.62 (1.59)	6.75 (1.83)	4.75 (1.03)	57.03 (14.81)	62.01 (12.4)
High	6.25 (1.28)	7.66 (1.49)	4.91 (1.16)	71 (10.63))	72.43 (9.96)
Kruskal Wallis	$\mathbf{x^2{=}5.91}$**, p=0.052**	$x^2{=}4.68$, p=0.09	$x^2{=}2.73$, p=0.25	$\mathbf{x^2{=}6.57}$**, p=0.037**	$\mathbf{x^2{=}6.61}$ **p=0.037**

In response to our first research question, these results provide some evidence that engagement and comprehension are related, in that participants with either a high or low level of Interest (INT), measured in terms of their felt involvement, and the endurability and novelty of the reading experience, achieved better outcomes on the comprehension test than those with a medium level of interest. Specifically, they had better outcomes on the True and False questions, an indication of surface level comprehension and recall. Other dimensions of engagement, specifically the aesthetic appeal and perceived usability of the reading environment, and the extent to which the readers focused their attention on the task, were not associated with comprehension outcomes.

5.2. User, System and Content Characteristics

This section presents the results of analyses to determine whether characteristics of participants (demographic variables, reading ability, and general reading motivation), the experimental system (interactivity and text presentation style), or content (reading set, pre- and post-task knowledge and interest) offer insights into the findings with respect to engagement and comprehension. We test for relationships between these variables and engagement, focusing on the INT dimension of engagement, which displayed an association with comprehension.

5.2.1.User Characteristics

We were interested in whether student status or age affected self-reported engagement. Older or more advanced students may have greater familiarity with the task of preparing for class, the scenario featured in our experiment; one might assume they have developed efficient strategies for navigating texts and could therefore become more engaged. Six participants (two in the low and one in the medium groups) did not provide their age, gender or student status. Chi-square tests showed no significant differences in INT for student status (e.g., undergraduate, Masters, etc.), x^2=5.62(6), p=0.443; gender, x^2=3.62(2), p=0.163; or age group, x^2=5.08(4), p=0.278. However, it is worth noting that 4 of the 5 doctoral students who participated in the study were in the High INT group.

We also examined the relationship between INT scores and reading ability and motivation. We might hypothesize that people with lower reading ability or lower motivation would be less engaged in the reading task scenario due to the cognitive challenge of engaging with the texts or disinterest in reading. We saw no difference between INT ratings and students' Nelson Denny reading grade, x^2=1.985(2), p=0.371, or reading strength, x^2=0.695(2), p=0.707. Despite the fact that this was a sample of university students, there was a range of reading abilities amongst participants. The mean Nelson Denny reading grade was 14.24 (range: 4.09-18.8). Thus, the sample did not consist solely of "strong readers" according to the Nelson Denny Test, yet this variability in reading grade and strength did not affect perceived engagement. Schutte and Malouff saw motivation as a "central feature" of reading engagement that impacts reading competency and both the function and social components of reading [38, p. 470]. We expected a positive association between participants' reading motivations, as measured by the Adult Reading Motivation Scale and their engagement in the texts, but this was not supported, x^2=3.53(2), p=0.171.

5.2.2. System Characteristics

Chi-square tests also showed no significant differences according to whether students were assigned to an interactive or non-interactive condition, x^2=3.96(2), p=0.138, nor a plain text or in-context condition, x^2=0.76(2), p=0.682. However, the frequency distributions across the four system environments are suggestive of some trends (Table 7). Almost half of those in the Low INT group were in the simplest reading environment and the majority of those in the High INT group were in one of the interactive conditions, in which they were provided with tools for highlighting and annotating text while reading.

5.2.3.Content Characteristics

Aspects of content, such as the topic, format or writing style, may influence UE [32]. In this study there were no statistically significant differences in INT between the two sets of articles, x^2=4.77(2), p=0.092. However, more participants were in the high INT group for the human-robot interaction article set, which had four people in the medium, and nine people in the high group. The digital activism set of readings had twelve people in the low, four in the medium, and five in the high INT groups. This suggests that the content of the readings had some impact on participants' engagement.

When comparing participants' self-reported interest in the content (post-task) with their level of engagement based on the UES, Kruskal Wallis tests show significant differences between the low, medium and high INT groups, x^2=9.45(2), p=0.009 (Table 8). This result validates our interpretation of the UES INT factor as representing the user interest component of engagement. Further, patterns across the three groups in pre and post task reports of interest in and knowledge of the content offer some insight into this finding (Tables 8 and 9). Mean pre-task levels of interest and knowledge in the topic of the articles are lowest for the Medium

INT and highest for the High INT groups, but the level of interest of both groups increased after reading. In contrast, the mean level of interest among those in the Low INT group, which started out quite high, decreased after reading.

Table 7: Frequency of system conditions by levels of INT

INT	*n*	Non-Inter-active, Plain	Inter-active, Plain	Non-Inter-active, In-Context	Inter-active, In-Context
Low	16	7	1	3	5
Medium	8	1	2	4	1
High	14	3	5	1	5

Table 8: Pre- and post-task interest for low, medium and high INT groups

INT	*n*	Pre-task interest *M*(*SD*)	*Post-task Interest* *M*(*SD*)	*Change in Interest*
Low	16	4.69 (1.44)	3.44 (1.36)	-1.25 (1.52)
Medium	8	4.25 (1.5)	4.88 (0.99)	0.62 (1.4)
High	12	4.83 (1.68)	5.58 (0.99)	0.75 (2.09)

The difference between students pre- and post-task self-reported knowledge was not statistically significant: x^2=5.008(2), p=0.082, and all groups, on average, reported gaining knowledge over the course of the study (Table 9). This gain was less pronounced for the low INT group, which had lost interest over the course of the readings, as compared to the medium and high groups, which both became more interested as they read (Table 9). It is worth noting that the Medium INT group, which achieved the lowest outcomes on the comprehension tests, reported a lower level of topical knowledge prior to reading and the highest knowledge gains of all the groups. Thus, while they may have scored lowest on the comprehension measures, they evaluated their own learning as having increased from when they began the experiment.

It is recognized that prior knowledge is key to formulating queries and devising tactics in information search [41]. Extrapolating to the digital reading environment, students' lack of topical knowledge may have resulted in less understanding of how to interact effectively and efficiently with these particular sets of readings. The effort they expended to make sense of the texts in the limited time frame of the study may have affected their INT ratings. In this case, learning gains were made, but they may have experienced more cognitive load than either the low or high INT groups, who began with higher self-reported topical knowledge, and evaluated their experience as less engaging. Their comprehension score may have been equally affected by their lack of prior subject knowledge. If digital reading strategies are comparable to information seeking more generally, where there is movement from the general to the specific [41], students may have spent their time orienting themselves to the topic rather than focusing on the detailed information of the texts which they were required to understand to complete the comprehension tests.

Table 9: Pre- and post-task knowledge for low, medium and high INT groups

INT	*n*	Pre-task knowledge *M*(*SD*)	*Post-task Knowledge* *M*(*SD*)	Change in knowledge
Low	16	2.44 (1.2)	3.56 (1.2)	1.12 (1.14)
Medium	8	2.00 (0.92)	4.25 (1.48)	2.25 (1.28)
High	12	2.83 (2.16)	4.75 (1.54)	1.91 (1.67)

In response to our second research question, we found limited evidence of an influence of user, system or content characteristics on user engagement. Statistical testing on this small sample showed no significant results for most of the variables tested; however, the descriptive data suggest some possible trends that merit further investigation: doctoral student participants were primarily in the High INT group, as were the majority of those in the Interactive System condition, and more of the readers of one article set (human-robot interaction) than the other. Different patterns with respect to levels of topical knowledge and interest across the three INT groups and over time point to a possible explanation for the non-linear relationship between INT and comprehension observed in this study.

6. DISCUSSION

We examined the relationship between user engagement, as measured by the User Engagement Scale, and comprehension, as measured through true/false and sentence verification items that addressed micro-level (local) comprehension, and sentence summaries intended to capture macro-level (global) understandings of texts. We reiterate here that our sample was relatively small and our findings should be treated as exploratory. Rather than identifying a strong effect of engagement on comprehension, our results echo those of previous research in the educational domain, summarized in section 2, which found partial or inconclusive associations between engagement and learning. Nevertheless, these results contribute to our understanding of human information interaction as a complex phenomenon, and offer some insights for future research.

We observed a positive association between one of the four UES subscales – INT – and comprehension at the micro-level and overall. Given that this was a reading task conducted in a relatively simple and familiar web environment, it makes sense that the INT factor, representing the novelty, felt involvement and endurability of the reading experience, would be associated with comprehension, while the more system-oriented components of engagement, AESTH and USAB, were not. If we had compared highly complex systems with more pronounced design features, it is possible that these factors would have had an effect. The lack of an association between comprehension and focused attention (ATTEN) is more surprising. We had anticipated that a higher degree of perceived focus during the study would contribute to comprehension, in keeping with the notion that text comprehension is associated with effort on the part of the reader [20]. It may be that, given the relatively low scores on this factor (M=2.91, SD=0.73), there was not enough variation to see a difference between groups.

The only individual test score associated with INT was the True/False questions, which measure surface learning [11]. From this, we can infer that readers who were highly engaged in the textual content were more likely to remember facts gleaned from the texts; however, this association only held true for the Medium and High INT groups. Surprisingly, the Low INT group performed as well as the High INT group on the True/False questions, indicating that the relationship between engagement and comprehension may be non-linear or mediated by other variables.

To better understand the relationship between UES ratings and comprehension, we looked at user characteristics (demographics, reading ability and reading motivation), the experimental condition to which people were assigned and content characteristics (article sets, user interest and prior knowledge). We hypothesized that people who were farther along in their academic careers, stronger readers, or those who rated reading as a more favorable (enjoyable) activity would be more engaged [38]. Further, we expected that the experimental condition to which people were assigned would affect UE based on how the different reading environments shaped the user experience. Surprisingly, clear relationships between these variables and engagement did not emerge, although trends in the data suggest that level of education, the provision of interactive reading tools, and the nature of the content may be associated with engagement. Thus designing reading and learning environments requires consideration of many variables beyond the system's affordances [17], and experience may be shaped by factors beyond the developer's control.

Data on participants' knowledge of and interest in the content of these texts provided the most insight into the observed association between engagement and comprehension. While all participants reported similar levels of knowledge and interest before starting the reading task, the Low INT group lost interest during the task and reported learning less, while the Medium and High groups increased in both interest and knowledge. Two important questions arise from this: what caused the decline in interest for the Low INT group, and how did they manage in spite of this, to achieve comprehension outcomes as good as or better than the other groups?

One explanation for this is that students in the Low INT group may have been working hard to learn [22] and had good outcomes, but a less positive experience. The effect of that experience may have caused them to under-estimate how much they actually learned. This conjecture is supported by the observation that many of the participants in the Low INT group were in the simplest, and possibly least engaging, reading environment (plain text/non-interactive), and from Whitman's [45] findings of students' experiences with either a baseline or interactive tutorial: learning gains were achieved for students using both systems, but the interactive tutorial resulted in faster completion without sacrificing their ability to succeed or their enjoyment. Results of a separate analysis of the current study's dataset showed that readers in the plain text/non-interactive system condition had the strongest comprehension outcomes overall, but also tended to spend more time reading [14]. Future work should include assessment of cognitive effort to measure the effects of challenge on UE and comprehension in digital reading environments.

Another explanation is that students in the low INT group, who had quite a high level of prior knowledge, lost interest because the texts failed to add anything new to their existing knowledge of the topic, resulting in boredom, disengagement and a negative self-perception of their learning. Despite this, they completed the assigned task and their knowledge of the topic enabled them to do well on the comprehension tests. This corresponds to findings in the online news domain showing that novel content was imperative for initiating and sustaining engagement for some news readers [32]. Finally, it is possible that there were different types of motivations inherent within these groups. For example, the low INT group may have been motivated by the extrinsic reward of the honorarium at the end of the experiment, whereas the high INT group may have been intrinsically motivated or interested in the subject matter of the texts. The concentration of doctoral students, predominantly highly self-motivated learners, in the High INT condition suggests this may be one factor at play.

If we focus on evidence of learning, rather than comprehension, the Medium and High INT groups reported higher gains in knowledge than the Low INT group, even though the Medium group was least successful on the comprehension test. This supports a relationship between UE and learning, at least in terms of self-perception: those who rated their novelty, felt involvement and endurability higher claimed to have learned more. This also points to the limitations of comprehension as a static outcome measure and the need to employ more individualized process measures to assess learning in the context of human information interaction. For example, recent work [16] used students' facial expressions and postures to disambiguate engagement and frustration with intelligent tutoring systems over seven sessions, defining specific expressions, postures and event sequences, such as interactions with human and non-human tutors, that characterized the nature of students' experiences; these were corroborated with self-reports. This research shows the potential for designing more responsive systems, but also for triangulating process and outcome, and subjective and objective data.

Clearly, prior knowledge plays a key role in comprehension generally, and, although we did not observe differences at the macro-level of comprehension through the Sentence Summary questions, we expect that it is particularly important for these (the?) more integrative components of comprehension [22]. However, due to sample size, it was not possible to perform regression analysis using prior knowledge as a co-variate. Another limitation of our experiment is that we looked at comprehension over a two-hour period, rather than an actual academic class where scaffolding content and collecting data at multiple points in time might provide more accurate measures of comprehension at the macro-level. Future work should address these limitations.

Interest in content, and how it affects learning, is also worthy of further exploration. This study pointed to interest, whether measured through self-report data or through the UES INT factor, as a key component of both engagement and learning for human information interaction tasks. Flowerday, et al. [11] distinguished topical (trait) and situational (state) interest; both promote deeper learning and engagement with reading materials. They examined self-reported topical and situational interest in an experiment that manipulated perceived choice in the reading of an expository text. Situational interest, more so than topical interest, increased UE and resulted in more positive affect. In our study, we used the scenario of preparing for class to create situational interest and observed that some aspects of UE were positively associated with pre and post task changes in knowledge and interest. In considering how to stimulate interest, designers of digital reading environments should focus on situational interest, particularly since topical knowledge is largely unpredictable. This may mean going beyond retaining the visual or structural features of genre, as we did in this study, and experimenting with visual salience through the size and placement of text [28]. More guidance on manipulating visual salience and attention, specifically in the area of academic reading environments, could be better revealed through eye tracking; recent research has demonstrated the success of this method for examining other IIR phenomena, including relevance [15].

One important aspect of human information interaction that was not addressed in this study is persistence: the motivation and willingness to continue interacting with the learning environment and to complete the task. Because of the experimental setting, participants were unlikely to stop reading, whether or not they were highly engaged in the task. However, in a real world academic setting, persistence is more likely to be intertwined with engagement, and may then have a more direct effect on comprehension and learning: a student who stops reading, or skips more quickly through a text, is much less likely to learn from it. This points to the need to conduct more naturalistic studies of the relationship between engagement and learning in the context of digital reading, as has been done in the past in information seeking research [for example, 41].

7. CONCLUSION

In this paper, we found that students in the low and high INT groupings had comparable microstructural and total comprehension scores; we did not find a relationship between macro-level (global) comprehension and UE. We explored connections between UE and individual (demographics, reading ability, general reading motivation), and system (interactivity, para-textual cues) characteristics, and the article sets, but no significant differences were detected. This may be due to our sample size and the small number of participants in each condition. However, we did note statistical differences when we examined self-reported INT and pre and post task interest and knowledge. While all groups reported increased knowledge as a result of interacting with the texts, the interest of the low INT group declined. Overall, the directionality of the relationship amongst the variables in not clear: for some participants, interest in the texts and engagement with the task was associated with text comprehension, while for other students there was no association between their perceived experience and comprehension scores.

Results suggest that engagement, arising from a user's interest in content, can facilitate comprehension and learning in academic environments, but that engagement is not an essential component for learning to occur. It is therefore not appropriate to assume that more engaging systems will necessarily produce better learning outcomes: the relationship is more complex than this. As these results are not conclusive, more research is needed to explore other factors, such as motivation, with larger samples and in more naturalistic settings. Future work might focus on design features that increase visual salience as a way to increase situational interest, measuring cognitive load as a means of interpreting users' perceptions of the interaction, and using eye-tracking methods to better appreciate readers' experiences. Replicating this research with a larger sample would allow for the construction and testing of predictive models to better articulate the connection between UE and comprehension in digital reading environments.

In this research we opted to examine students' reading experiences quantitatively, and the group administration of our experiment prevented us from collecting qualitative post-session interviews to learn more about students' perceptions of the tasks, reading environments, and levels of engagement. The combination of qualitative and quantitative methods in future work will facilitate a more holistic understanding of the relationship between engagement, interest, and learning. However, as noted in other research [14; 46], measuring learning in IIR contexts is challenging and has been limited to quizzes and written summaries that test recall and recognition rather than learning. Here we sought to build a measure rooted in the learning sciences that was specific to the texts students were exposed to in the study; the UES and Adult Reading Motivation Scale are established measures against which to examine this novel comprehension measure. As research interests in this area continue to grow, so do the possibilities for validating measurement approaches with physiological and qualitative data.

8. ACKNOWLEDGMENTS

Thank you to our research participants and graduate research assistance Florian Ehrensperger, Elizabeth Shaffer, Kimberly Buschert, Ariel Deardorff and Devin Soper. This research was generously supported by a Hampton Fund Research Grant from the University of British Columbia.

9. REFERENCES

1. Adler, M.J. 1940. *How to Read a Book: The Art of Getting a Liberal Education*. Simon and Schuster, New York, NY.
2. Brown, J.A., Fishco, V.V., and Hanna, G. 1993. *Nelson–Denny Reading Test: Manual for Scoring and Interpretation.* Riverside Pub, Rolling Meadows, IL.
3. Chung, D. S. 2008. Interactive features of online newspapers: Identifying patterns and predicting use of engaged readers. *J J Comput Mediat Commun, 13*, 658–679.
4. Cole, C. and Mandelblatt, B. (2000). Using Kintsch's discourse comprehension theory to model the user's coding of an informative message from an enabling information retrieval system. *JASIS&T, 51*(11), 1033-1046.
5. Csikszentmihalyi, M. 1990. *Flow: The Psychology of Optimal Experience.* Harper and Row, New York, NY.
6. Cull, BW. 2011. Reading revolutions: Online digital text and implications for reading in academe. *First Monday*, *16*(6).
7. Denny, P. 2013. The effect of virtual achievements on student engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 763-772). ACM. DOI= 10.1145/2470654.2470763.
8. DeVellis, R. F. (2003). *Scale Development: Theory and Applications* (2nd ed). Sage, Thousand Oaks, CA.
9. Dillon, A. 1992. Reading from paper versus screens: An empirical review of the literature. *Ergonomics, 35*, 1297-1326.
10. Eysink, T.H.S. and de Jong, T. 2012. Does instructional approach matter? How elaboration plays a crucial role in multimedia learning. *J Learn Sci, 21*(4), 583-625.
11. Flowerday, T., Schraw, G., and Stevens, J. 2004. The role of choice and interest in reader engagement. *J Exp Educ, 72*, 93–114.
12. Fredricks, J. A., Blumenfeld, P. C., and Paris, A. H. 2004. School engagement: Potential of the concept, state of the evidence. *Rev Educ Res, 74*, 59–109.
13. Fredricks, J. A., McColskey, W., Meli, J., Montrosse, B., Mordica, J. and Mooney, K. 2011. *Measuring student engagement in upper elementary through high school: A description of 21 instruments.* Greensboro, NC. http://ies.ed.gov/ncee/edlabs/regions/southeast/pdf/rel_2011098.pdf.

14. Freund, L., Kopak, R. & O'Brien, H.L. (in press). The effects of textual environment on reading comprehension: implications for searching as learning. *J Inform Sci*.
15. Gwizdka, J. 2014. Characterizing relevance with eye-tracking measures. In *Proceedings of the Information Interaction in Context Symposium* (pp. 58-67). ACM.
16. Grafsgaard, J.F. (2014). Multimodal Affect Modeling in Task-Oriented Tutorial Dialogue. Unpublished Doctoral Dissertation, North Carolina State University. Raleigh, North Carolina.
17. Hu, P.J-H. and Hui, W. 2012. Examining the role of learning engagement in technology-mediated learning and its effect on learning effectiveness and satisfaction. *Deci Support Syst, 53*, 782-792.
18. Jacques, R.D. 1996. *The nature of engagement and its role in hypermedia evaluation and design*. Unupblished doctoral dissertation, South Bank University, London.
19. Jacques, R., Preece, J. and Carey, T. 1995. Engagement as a design concept for multimedia. *Cdn J. Educ. Comm.* 24,1, 49-59.
20. Kelly, D. 2009. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval, 3*(1-2), 1-224.
21. Kintsch, W. 1998. *Comprehension: A Paradigm for Cognition*. Cambridge University Press, New York, NY.
22. Kintsch, W. 2009. Learning and constructivism. In Tobias, Sigmund and Duffy, Thomas M. (Eds), *Constructivist Instruction: Success or Failure?* (pp. 223-241). Routledge, New York, NY.
23. Kopak, R. and Chiang, C 2007. Annotating and linking in the Open Journal Systems. *First Monday*, 12(10). http://firstmonday.org/article/view/1961/1838
24. Lalmas, M., O'Brien, H., and Yom-Tov, E. 2014. *Measuring User Engagement*. Morgan Claypool Pub.
25. Liu, Z. 2006. Reading behavior in the digital environment: changes in reading behavior over the past ten years. *J Doc, 61*(6), 700-712.
26. Marshall, C. 2009. *Reading and Writing the Electronic Book*. Synthesis Lectures on Information Concepts, Retrieval, and Services 1:1, pp. 185. Morgan Claypool Pub.
27. McCay-Peet, L., Lalmas, M., and Navalpakkam, V. 2012. On saliency, affect and focused attention. In *Proceedings of Conference on Human Factors in Computing Systems* (pp. 541-550). ACM.
28. Maynard S. and McKnight, C. 2001. Children's comprehension of electronic books: An empirical study. *New Rev Children's Lit. Libr., 7*, 29-53.
29. McNamara, D.S. and Shapiro, A.M. 2005. Multimedia and hypermedia solutions for promoting metacognitive engagement, coherence, and learning. *J Educ Comput Res, 33*(1), 1-29.
30. McNamara, D. S. and Magliano, J. P. 2009. Towards a comprehensive model of comprehension. In B. Ross (Ed), *The Psychology of Learning and Motivation, vol. 51*, (pp. 297-284). Elsevier Science, New York, NY.
31. O'Brien, H.L. Translating Theory into Methodological Practice. In H.L. O'Brien and P. Cairns (Eds). *Why Engagement Matters: Cross-Disciplinary Perspectives and Innovations on User Engagement with Digital Media*. Springer, forthcoming.
32. O'Brien, H.L. 2011. Exploring engagement in online news interaction. In *Proceedings of the Annual Meeting of the Association for Information Science and Technology*, New Orleans, LA.
33. O'Brien, H.L., and Toms, E.G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology . *JASIS&T*, 59(6), 938-955.
34. O'Brien, H.L. and Toms, E.G. 2013. Measuring engagement in search systems using the User Engagement Scale (UES). *Inform Process Manag, 49*(5), 1092-1107.
35. O'Brien, H.L. and Toms, E.G. 2010. The development and evaluation of a survey to measure user engagement in e-commerce environments. *JASIS&T, 61*(1), 50-69.
36. Pearson, J., Buchanan, G. and Thimbleby, H. 2013. *Designing for Digital Reading*. Synthesis Lectures on Information Concepts, Retrieval, and Services 5(4), 1-135.
37. Royer, J. M., Greene, B. A. and Sinatra, G. M. 1987. The sentence verification technique. *J Reading*, *30*(5), 414-422.
38. Schutte, N. S. and Malouff, J. M. 2007. Dimensions of reading motivation: Development of an adult reading motivation scale. *Read Psychol*, *28*(5), 469-489.
39. Shannon S. and Winterman B. 2012. Student comprehension of primary literature is aided by companion assignments emphasizing pattern recognition and information literacy. *Issues in Science and Technology Librarianship* http://istl.library.ucsb.edu/12-winter/refereed3.html
40. Sung, Y-T., Chang, K-E., Hou, H-T., and Chen, P-F. 2010. Designing an electronic guidebook for learning engagement in a museum of history. *Comput Hum Behav, 26,* 74-83.
41. Vakkari, P/ (2001). A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *J Doc*, 57 (1), 44 – 60.
42. van Dijk, T.A. 1980. Story comprehension: An introduction. *Poetics, 9*(1), 1-21.
43. Webster, J. and Ahuja, J. S. 2006. Enhancing the design of web navigation systems: The influence of user disorientation on engagement and performance. *MISQ, 30*(3), 661-678.
44. Webster, J. and Ho, H. 1997. Audience engagement in multimedia presentations. *ACM SIGMIS Database*, *28*(2), 63-77.
45. Whitman, L. 2013. *The Effectiveness of Interactivity in Multimedia Software Tutorials*. Unpublished Doctoral Dissertation, North Carolina State University, Raleigh, NC.
46. Wiebe, E. N., Lamb, A., Hardy, M., and Sharek, D. 2014. Measuring engagement in video game-based environments: Investigation of the User Engagement Scale. *Comput Hum Behav*, *32*, 123-132.
47. Wilson, M. J., and Wilson, M. L. (2013). A comparison of techniques for measuring sensemaking and learning within participant generated summaries. *JASIS&T*, *64*(2), 291-306.