The Development and Evaluation of a Survey to Measure User Engagement

Heather L. O'Brien
*School of Library, Archival and Information Studies, University of British Columbia, Vancouver, BC V6T 1Z1
Canada. E-mail: hlobrien@interchange.ubc.ca*

Elaine G. Toms
*School of Business Administration, Dalhousie University, Halifax, Nova Scotia, B3H 3J5 Canada.
E-mail: Elaine.Toms@dal.ca*

Facilitating engaging user experiences is essential in the design of interactive systems. To accomplish this, it is necessary to understand the composition of this construct and how to evaluate it. Building on previous work that posited a theory of engagement and identified a core set of attributes that operationalized this construct, we constructed and evaluated a multidimensional scale to measure user engagement. In this paper we describe the development of the scale, as well as two large-scale studies (N = 440 and N = 802) that were undertaken to assess its reliability and validity in online shopping environments. In the first we used Reliability Analysis and Exploratory Factor Analysis to identify six attributes of engagement: Perceived Usability, Aesthetics, Focused Attention, Felt Involvement, Novelty, and Endurability. In the second we tested the validity of and relationships among those attributes using Structural Equation Modeling. The result of this research is a multidimensional scale that may be used to test the engagement of software applications. In addition, findings indicate that attributes of engagement are highly intertwined, a complex interplay of user-system interaction variables. Notably, Perceived Usability played a mediating role in the relationship between Endurability and Novelty, Aesthetics, Felt Involvement, and Focused Attention.

Introduction

The expectation of today's users is that software applications will not only be functional, but also engaging (Overbeeke, Djajadiningrat, Hummels, Wensveen, & Frens, 2003). In technologically pervasive work and marketplaces, users spend little time assessing the relevancy and utility of an application (Lingaard, Fernandes, Dudek, & Brown, 2006). There is an impetus for technology developers to exceed usability and provide an experience. Today's consumers are

making up their minds about what technologies to invest their time, effort, and dollars in based on how they make them feel. Thus, the question is no longer only whether an application is efficient, effective, or satisfying, but how well it is able to engage users and provide them with an experience (Bannon, 2005; Overbeeke et al., 2003). The experience movement is being driven both by the desire to identify system features that facilitate positive user experiences and apply them to the design of technologies (Lingaard, 2004), and to market beyond tangible goods and services, and incorporate experience as a means of differentiating goods and services in a competitive environment (Pine & Gilmour, 1999).

Designing for engaging experiences is an oft-cited goal of interactive system development in many disciplines, yet there are no guidelines to channel designers' efforts to "make things engaging" (Overbeeke et al., 2003). Part of the issue has been that, until recently, there have been disparate ideas about what constitutes engagement. Without a consistent definition of engagement, it is difficult to ascertain that the systems we design and market are, in fact, engaging, or to identify what aspects of the interaction with technology engage or fail to engage users. Without a means of measuring engagement, we are limited in our understanding of this quality of user experience and our evaluations of user-centered technologies.

To resolve this problem, O'Brien & Toms (2008) drew together the threads of multiple research projects to identify the core attributes that constitute engaging experiences, and to propose a conceptual, process-based model of engagement that was founded on Aesthetic, Flow, and Play Theories. This was followed by an exploratory study that tested the model and confirmed the attributes. This multistage process standardized the concept engagement and illuminated its essence. This research follows that work by focusing on the assessment of engagement. Reported here are the results of our research that constructed and tested a means of evaluating engagement with technology. In the process of doing so, we explored and verified attributes of engagement and examined

the relationship among them. The central questions guiding this work were:

1. How can we measure engagement?
2. Would the attributes reported in O'Brien & Toms (2008) be supported in the current research?
3. What is the nature of the relationships amongst these attributes?

We undertook a process of scale development that included the conduct of two large-scale surveys and a rigorous data analysis process. Addressing the aforementioned research questions resulted in a 31-item scale for measuring engagement (Appendix 1).

Literature Review

The Composition of Engagement

Until recently, engagement was articulated loosely as a cognitive (Laurel, 1993), affective (specifically intrinsic motivation) (Jacques, 1996; Jacques, Preece, & Carey, 1995; Jones, n.d.), and behavioral (Kappelman, 1995; Hutchins, Holland, & Norman, 1986) state of interaction with a computer application that "makes the [user] want to be there" (Jones, n.d.). Engaging interactions were thought to involve attention (Chapman, 1997; Webster & Ho, 1997), intrinsic interest (Jacques, 1996; Chapman, 1997; Webster & Ho, 1997), interactivity (Quesenbury, 2003), perceived control and choice (Jacques et al., 1995; Webster & Ho, 1997), functionality (Jacques et al., 1995), and motivation (Makkonen, 1997). This collection of attributes was derived from studies carried out in multiple disciplines (i.e., education, e-commerce, human-computer interaction, etc.) with varied applications, including educational multimedia (Chapman, Selvarajah, & Webster, 1999) and presentation software (Webster & Ho, 1997), print and online reading (Schraw, Flowerday, & Reisetter, 1998; Konradt & Sulz, 2001), museum exhibits (Haywood & Cairns, 2005; Hull &

Reid, 2003), and video games (Reiber, 1996). For a detailed review of this research, see O'Brien & Toms, 2008.

As a group, these studies provide a view of engagement through a wide interdisciplinary lens and offer numerous insights into what it means to be engaged in a variety of information-rich contexts. According to this body of work, engagement is impacted by: media richness through the use of animations and video (Webster & Ahuja, 2006); format (e.g., text, audio, and video) (Chapman, 1997; Jacques et al., 1995; Laarni, Ravaja, Kallinen, & Saari, 2004); interactivity and exploration (Haywood & Cairns, 2005); communication or socialization with others (Haywood & Cairns, 2005; Hull & Reid, 2003); aesthetics and sensory appeal (Haywood & Cairns, 2005; Hull & Reid, 2003; Laarni et al., 2004); intellectual challenge (Douglas & Hargadon, 2000; Mandryk, 2004; Said, 2004); and affective involvement (Laarni et al., 2004; Schraw et al., 1998; Said, 2004).

A Model of User Engagement

Building on this previous research, we posited a model of engagement that examined these attributes in concert and rooted them in a theoretical framework of Aesthetic, Play, and Flow Theories, and then tested the model through an exploratory study that involved interviewing users of four different types of technology: Web searching, shopping, video games, and online learning (O'Brien & Toms, 2008).

Overall, the model hypothesized that engagement is both a process and product (Kappelman, 1995) of interaction; its intensity may change over the course of an interaction (Said, 2004) depending on the combination of users' needs, goals, emotions, actions, and thoughts, or the format (Chapman, 1997; Jacques, 1996), visual presentation, and organization of the computer interface (Quesenbury, 2003). A range of user and system-specific attributes were identified in the review of related research and exploratory study: aesthetics, sensory appeal, focused attention, awareness, challenge,

TABLE 1. Definition of the attributes of engagement.

Attribute	Definition
Aesthetics	Visual beauty or the study of natural and pleasing (or aesthetic) computer-based environments (Jennings, 2000).
Affect	"The emotional investment a user makes in order to be immersed in an environment and sustain their involvement in the environment" (Jennings, 2000); "The user's emotional response to the system" (Stone, Jarrett, Woodroffe, & Minocha, 2005, p. 483).
Focused Attention	The concentration of mental activity; concentrating on one stimulus only and ignoring all others (Matlin, 1994).
Challenge	The amount of effort experienced by the participant in performing an online task.
Control	How "in charge" users feel over their experience with the technology.
Feedback	Response or reaction from the task environment or system that communicates the appropriateness of the users past actions or demonstrates progress toward a specific goal; serves as a basis for future action ("Feedback," Penguin Dictionary of Psychology, 1995); "Information that is sent back to the user about what action has been done or what result has been accomplished" Stone et al., p. 613).
Interest	The "feeling that accompanies or causes special attention to an object or class of objects" ("Interest," M-W Online).
Motivation	Elements that bring about focus or a desire to proceed with an activity (Jennings, 2000).
Novelty	Variety of sudden and unexpected changes (visual or auditory) that cause excitement and joy or alarm (Aboulafia & Bannon, 2004); Features of the interface that that "users find unexpected, surprising, new, and unfamiliar" (Huang, 2003).
Perceived Time	Users' perception of estimated time spent on task.

control, feedback, interest, motivation, novelty, and perceived time. Influences on engagement are perceived usability, interactivity, and environmental/situational variables (e.g., interruptions, task pressures). Table 1 defines the attributes that were considered during the initial stages of this research. The articulation of these attributes provided a foundation for considering the measurement of engagement.

Measuring Engagement

It has been stated that there is sparse empirical research about how engagement should be measured (Chapman, 1997; Jacques et al., 1995). However, over the past 15 years a number of researchers have purported to measure engagement using a variety of data collection techniques. The most common of these has been self-report measures, including Webster and Ho's (1997) seven-item questionnaire with items pertaining to attention, challenge, intrinsic interest, and variety, and Jacques' (1996) 13-item survey containing items that related to attention, perceived time, motivation, needs, control, attitudes, and overall engagement.

Some researchers have used performance indicators, not as measures of engagement, but as correlates of engagement. For example, Konradt and Sultz (2001) employed pre- and post-task measures to examine changes in users' affective and cognitive states over the course of interacting with an educational application. Chapman (1997) studied the relationship between engagement (as measured by Webster and Ho's instrument) and performance metrics, such as time on task (as gathered through log files) and knowledge of system content (as measured by multiple-choice quizzes) as a consequence of media format. One researcher posited that multiple measures (self-report, participant observation, biometrics) are needed to study engagement but did not offer any results of empirical data collection or describe how these measures might be triangulated to say something meaningful about engagement (Champion, 2003).

Performance indicators and physiological metrics have the advantage of being observed in users' biology (e.g., heart rate) or behaviors (e.g., eye gaze, mouse clicks) and collected over the course of an interaction with a system. Yet while metrics such as log analysis, number of eye fixations, heart rate, etc., answer the question of what is taking place during a user's interaction with a system, they do not address the users' cognitive or emotional state, both of which are critical to engagement. For example, Seah and Cairns (2008) found that conducting their study of immersion in video games in an experimental setting resulted in confounds they attributed to gamers' strong preferences for particular games. In addition, we do not have the ability to connect these metrics to engagement at this point in time. For instance, is a user "engaged" when they spend a great deal of time on a single screen of an application, or simply confused about how to navigate away from it? Increased attention, as measured through eye tracking, may signal heightened engagement, but focused attention may be one of many components of engagement, and thus this would give us an incomplete assessment.

Furthermore, collecting physiological data may be obtrusive for some participants and interfere with the experience.

Self-report measures are not as objective as performance and physiological measures, but they do offer a convenient and efficient means of assessing the users' perspective of an experience. As stated, a variety of researchers have developed survey instruments to evaluate engagement (e.g., Webster & Ho, 1997; Jacques, 1996), but these have not been generalized to domains beyond the ones in which they were created and administered. In addition, the items contained in each of these instruments represent a small portion of the attributes of engagement found in our exploratory study (see O'Brien & Toms, 2008). Existing measures of usability (e.g., The Questionnaire for User Interaction Satisfaction: http://www.lap.umd.edu/QUIS/index.html; The System Usability Questionnaire, Brooke, 1996) address the functionality of systems and users' satisfaction in using them, but do not tap into more experiential aspects of use, such as one's interest and motivation to use the application or continue using it, or the visual appearance of the interface, etc. Given the demands of today's technology users, it is essential to develop an instrument that encapsulates more than usability.

As in prior research, we believe that a survey instrument is the most appropriate technique for collecting users' perception of their level of engagement. Unlike prior research our earlier work strongly suggested that engagement was more broadly defined. As a result, we elected to build a scale that encompassed significantly more attributes than previous instruments. Such a scale would serve an added purpose: it could validate or refine the attributes identified in the earlier exploratory study. From a practical viewpoint, we envisioned that the resulting instrument would be an effective tool for comparing technologies (e.g., Website X and Website Y), collecting feedback during the design process of an application, or assessing users' responses to an existing system. The instrument would not only provide an overall evaluation of the user's experience, but would convey users' perceptions of the attributes.

Developing the User Engagement Scale

Constructing a survey instrument is a longitudinal process (see Figure 1) that begins with the development of items and design of the scale, and leads to the systematic evaluation of the instrument's reliability, validity, and generalizability (Lavie & Tractinsky, 2004; Peterson, 2000). In the following sections we outline our process of scale construction; subsequently, we describe two large-scale studies undertaken to evaluate the reliability and validity of the resulting instrument.

Scale Construction

Engagement is hypothesized to be a multidimensional construct and therefore it was imperative to construct a multidimensional survey instrument. The items were derived from attributes identified in the literature review and exploratory
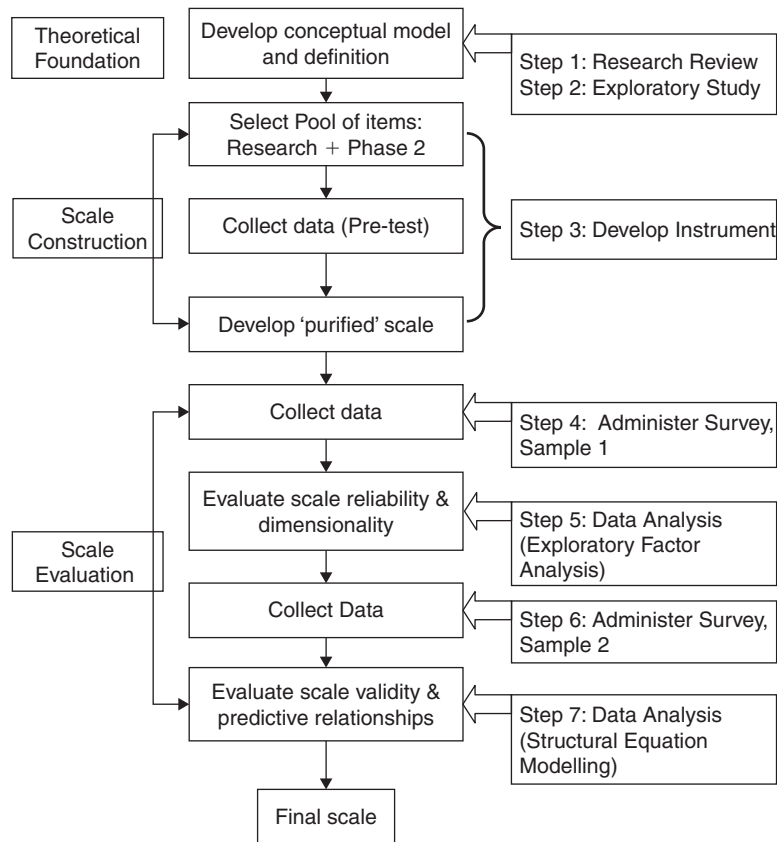
FIG. 1.   Overview of research design (adapted from Peterson, 2000, p. 78; London: Churchill, 1979; Gerbing & Anderson, 1988).

study (O'Brien & Toms, 2008). First we determined if there were well-established, existing scales for each of the attributes and, if so, whether or not they were appropriate to use in the context of this research.[1]

Select (though not exhaustive) results of this analysis are presented in Appendix 2, which also notes the general purpose and format of the items, and the context in which the items/scales were developed (e.g., the Web, or everyday life). Note that some attributes have been the subject of scales (e.g., aesthetics) more than others (e.g., control). Some instruments measure a single attribute, e.g., Park, Choi, and Kim's (2004) aesthetics scale, while others, such Huang's (2003) measure of interactivity, assess multiple attributes. Not all measures examined were appropriate for our use, namely, those that addressed facets of people's personalities rather than their interaction with an object (Litman & Speilberger, 2003; Reio, 1997).

In the course of this analysis, items associated with "intended inquiry" (i.e., intention to use an e-commerce Website to find a product or product information), "intended purchase" (i.e., intention to purchase an item from an e-commerce Website), and "loyalty" (Gefen & Straub, 2001; Choi & Kim, 2001; Webster & Ahuja, 2006) were included. Although not part of the list of attributes to emerge from the exploratory work, both Jones (n.d.) and Jacques (1996) viewed "continued use" as an outcome of engagement and this concept is embedded in reengagement. Reengagement is one of the stages in our process-based model, which signifies that engagement has ebbs and flows during the course of an individual's interaction with a system (O'Brien & Toms, 2008). Appendix 2 also shows these items, which are referred to in the literature as "intention to return," "intended purchase," "loyalty," and "intended inquiry," depending on the context. Existing measures of engagement (Webster & Ho, 1997) were also taken into account for inclusiveness.

The interview transcripts from our exploratory study (O'Brien & Toms, 2008) contained statements about users' perceptions of their engagement with technology. Due to the richness of interviewees' descriptions and the power and authenticity of their words, passages of text from the transcripts were highlighted and examined as potential items for the instrument. This practice is adopted from Kelley (2005).

In the gathering of potential items, we deliberately cast our net wide. It was necessary to ensure that the item pool was broad in order to establish the "conceptual boundaries" of the engagement scale, with the idea that unrelated items would be eliminated during the analysis, but potentially related items could not be added later (Clark & Watson, 1995).

---

[1]In this research the term "scale" refers to the survey instrument; "items" or "questions" are the statements included in the survey that are rated using a Likert scale. Multidimensional scales or instruments endeavor to measure more than one concept or attribute. In such cases, "subscales" are used to delineate the distinct attributes encapsulated in the scale.

TABLE 2. Nature of the assessment.

Item	Attribute	Retain?	Nonaffiliated researcher's rationale
The interface font was legible.	Aesthetics	No	I would expect computer font to be legible—does not get at proportion, type, creative use—addressed the mechanical and practical rather than the aesthetic.
Exploring system features by trial and error was easy.	Feedback	No	Does "trial and error" refer to not finding things where you expected to, or doing things correctly/incorrectly and receiving clear feedback? Ambiguous.

Compiling items. All of the possible items from existing scales ($n = 109$) and the interview transcripts ($n = 350$) were compiled. Given the formidable number of items ($n = 459$), a strategic process was adopted to evaluate this initial list. A nonaffiliated researcher was asked to evaluate each of the items according to: duplication/repetitiveness; potential applicability to human-computer interaction environments and experience measurement outcomes; and potential to be used across a range of computer applications. For example, Witmer and Singer's (1998) "How well do you feel today?" pertains to people psychological state, but not their attitudes toward a Website (see Appendix 2).

The nonaffiliated researcher was given the 459 items, the definitions of the attributes (see Table 1), and the above criteria. For each item she indicated whether the item was representative of its corresponding attribute, as well as whether it should be retained and her rationale. This was organized in a spreadsheet. Table 2 demonstrates the nature of her assessment (Table 2).

Upon completion the nonaffiliated researcher and the first author met to assess the number of agreements ($n = 440$) and disagreements ($n = 19$) and to discuss their respective rationales for retaining or not retaining each item. In situations where an agreement could not be achieved the second author intervened. This process reduced the number of potential items from 459 to 186 that were distributed across 11 subscales: aesthetics, affect, focused attention, challenge, control, engagement, feedback, reengagement, motivation, novelty, and perceived time. These subscales represent the attributes of engagement outlined in Table 1, as well as items for "intention to return," etc. (Appendix 2) (labeled "reengagement" for our purposes) and items that pertained to engagement but did not correspond to an attribute (e.g., "Application x was engaging to use."); these were combined in an "engagement" subscale.

Scale construction. The compiled items were reformatted to make them into statements rather than adjectives or phrases, general rather than specific, appropriate in tone (i.e., both negatively and positively phrased questions), and void of ambiguous terms (e.g., "could," "should,") or vague quantifiers (e.g., "occasionally," "most," or "very") (Peterson, 2000; DeVellis, 2003). The Likert scale, a common way of measuring attitude (Singleton, Straits, & Straits, 1993), was selected for its fit with the data and ability to provide summed ratings. The scale options addressed the intensity of users' attitudes about the application; the final five-point Likert scale was "strongly disagree," "disagree," "neutral," "agree," and "strongly agree"; there was also a "no response/not applicable" category.

Survey instrument. The online survey was constructed using Perseus Software Solutions, an application that allows for the creation of questionnaires and compilation of data. In both Studies 1 and 2, context effects (Peterson, 2000) were negated by randomizing items; everyone who completed the survey viewed the items in a different order.

Domain selection. In our previous work, users of four different technologies were interviewed (O'Brien & Toms, 2008). For the purposes of developing and testing an instrument, only one application, online shopping, was selected in order to increase the statistical power of the findings. Online shopping is "the consumers' adoption of the WWW as a means to purchase" (Shang, Chen, & Shen, 2005), involves directed searching and browsing, and is a domain where "usability and user experience come into strong contact" (Wright & McCarthy, 2005). It is also an environment in which engagement has seldom been explored, yet many proposed attributes of engagement have been highlighted in relation to e-commerce success, including motivation (e.g., Arnold & Reynolds, 2003), intention to return to a product or company Website (Webster & Ahuja, 2006), ease of use, aesthetics, attention, and interest (Mathwick & Rigdon, 2004). Thus online shopping was an appropriate and novel arena in which to evaluate the engagement instrument.

Pretesting. The survey was pretested with three people who were observed responding to the survey. Their reactions and questions were noted during this exercise and verbal comments were gathered after they had completed it. Comments pertaining to the survey's functionality, navigability, and clarity were used to improve the administration and content of the survey. A second pretest was then conducted with three different individuals and further feedback was incorporated into improving the presentation and understanding of the survey. The participants in the pretest were a convenience sample but were not associated with the research. Overall, the two pretests reduced the survey to 123 items (from 186) and this was the final version used in Study 1.

Study 1

The objectives of this study were:

- To assess each item to ensure that the instrument contained only the most parsimonious set of items;

- To evaluate the reliability of the subscales constructed for each attribute; and
- To examine the reliability of the overall instrument.

Methodology

The online survey. The Web-based presentation of the scale consisted of 13 Web pages. All pages of the survey contained instructions at the top of the page and a progress bar along the bottom to give people feedback about their proximity to completion. Page 1 contained introductory and informed consent information; Page 2 consisted of a demographic questionnaire; Pages 3–12 presented the survey items with 10–14 items per page to minimize scrolling; and the Concluding Page thanked respondents for their participation and invited them to enter a prize draw by clicking on a link that brought up a separate window where they could enter their email address. Prize winners were selected using a random number application in Excel on the list of emails. All personal information collected for the prize draw was stored separately from survey data and used only to contact winners.

Recruitment. The survey was posted online for 2 weeks (May 15–31, 2007) and in-person and online recruitment were ongoing during this time. In-person recruitment involved visiting undergraduate classes and placing notices around coffee shops and eateries on/near a university campus. Online recruitment was accomplished through online forums, discussion boards, and listservs using the snowball method ("Sampling," World of Sociology, 2001).

Participants. A total of 440 individuals completed the online survey. There were 305 females (69.3%) and 131 males (29.8%); four individuals (0.9%) did not state their gender. Respondents ranged in age from 18–25 ($n = 98$, 22.3%); 26–35 ($n = 172$, 39.2%); 36–50 ($n = 113$, 25.8%); to 51–70 ($n = 41$, 12.3%). Almost all participants used email (99.5%) and Web browsers (98%) on a daily basis. Of the 426 individuals who stated their occupation, only 104 were students; the remainder were employed in fields such as business, finance, health, education, information technology and management, retail, journalism, engineering, general laborers, etc.

Procedure. In-person and online recruitment methods directed participants to the online survey via a URL. The introductory Webpage led participants to the demographic questionnaire. Upon completing the demographic items, participants were instructed to recall the last online shopping experience they had and, with this scenario in mind, respond to the items on the following screens. These instructions were repeated at the top of each screen. The final page was a conclusion to the survey.

There were features that facilitated participants' navigation from the introduction, through the engagement items, to the conclusion. A "continue" button was in the bottom right-hand corner and the progress oriented respondents to their location in the survey. Once participants reached the concluding page they could either close the browser or click on the link to enter the prize draw, where they entered their name and email, and then hit a "submit" button.

Data preparation and exploration. To prepare the data for analysis, 37 of the 123 items were reverse-coded and responses with the value "6" ("not applicable" on the Likert scale) were recoded to reflect missing data. Next, an initial exploration of the data involved examining descriptive statistics, including frequencies of valid responses, means and standard deviations, and interitem correlations; these results were used to determine whether items should be retained for further analysis based on the nonresponse rate of each item, or variability of the item's rating (DeVellis, 2003). As a result, questions with means greater than 4 ($n = 12$) or less than or equal to 1 ($n = 0$) were eliminated, with the exception of three items that pertained to reengagement, the potential to return to an application at a later time. One item was kept because it was part of an established scale (DeVellis, 2003, p. 94), while the other two items demonstrated high variance, meaning that responses to these questions should vary according to individual survey respondents, and were retained (DeVellis, 2003, p. 93). The corrected interitem correlations of the items associated with each attribute were examined in detail. Questions that correlated negatively with other items (after reverse scoring) were eliminated ($n = 12$).

Results

The sample of 440 was more than adequate to proceed with data analysis (Nunnally, 1978). The analysis of the results consisted of:

1. Examining the reliability of the subscales, and
2. Performing exploratory factor analysis (EFA) to assess construct validity and the nature of the factors.

Reliability estimates of scale items. DeVellis' (2003, p. 95) guidelines were used in the assessment of reliability: below 0.60: unacceptable; between 0.60 and 0.65: undesirable; between 0.65 and 0.70: minimally acceptable; between 0.70 and 0.80: respectable; between 0.80 and 0.90: very good; much above 0.90: attempt to reduce the number of items. Items with the lowest item-total correlations were eliminated iteratively until optimal alpha values were achieved (between 0.7 and 0.9) with the least number of items. In the course of this procedure, the interest and motivation items were combined and the total scale was reduced from 123 to 49 items.

The reliability of each of the resulting subscales was calculated and the correlations among the subscales were examined. Table 3 shows the means, standard deviations, and alpha values of each subscale, as well as the correlation matrix among the subscales. While some of the correlations were low (<0.4), many were in the moderate (0.4–0.6) range, and some were between 0.7 and 0.8. The latter correlations

TABLE 3. Descriptive statistics and correlations for subscales.

	N	M	SD	α	1	2	3	4	5	6	7	8	9	10
Aesthetics	438	3.48	0.99	0.890	1									
Affect	417	3.69	0.89	0.872	0.554**	—								
Focused Attention	438	2.32	1.04	0.889	0.347**	0.097*	—							
Challenge	411	3.97	0.74	0.823	0.489**	0.759**	−0.100*	—						
Control	406	3.79	0.78	0.736	0.486**	0.757**	0.044	0.752**	—					
Engagement	438	2.9	0.92	0.749	0.486**	0.364**	0.780**	0.164**	0.219**	—				
Feedback	418	3.8	0.62	0.807	0.528**	0.659**	−0.053	0.749**	0.723**	0.154**	—			
Reengagement	428	4.1	0.73	0.837	0.481**	0.672**	0.092	0.787**	0.693**	0.288**	0.616**	—		
Motivation	438	3.7	0.84	0.837	0.590**	0.659**	0.389**	0.695**	0.624**	0.550**	0.547**	0.740**	—	
Novelty	438	3.28	0.90	0.728	0.568**	0.501**	0.467**	0.343**	0.367**	0.611**	0.337**	0.445**	0.580**	—
Perceived Time	438	2.80	0.88	0.766	0.434**	0.295**	0.779**	0.100*	0.169**	0.783**	0.066	0.241**	0.499**	0.576**

suggested that there could be some overlap among the subscales and that not all of these subscales would remain intact after performing factor analysis.

Exploratory factor analysis. EFA was used to examine the construct validity and multidimensionality of the instrument. The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy (KMO = 0.94) indicated that factor analysis should result in distinct, reliable factors (Hutcheson & Sofroniou, 1999) and the Bartlett's Test of Sphericity verified that relationships existed among the items ($\chi^2 = 10562.56$, $df = 1128$, $p < 0.001$).

Principal components extraction was selected to maximize the variance extracted and because an outcome of this analysis was to identify the most parsimonious set of items (Tabachnick & Fidell, 2007). Varimax rotation, the most common of the rotational techniques, was used to simplify the factors. Item loadings were interpreted using Comrey and Lee's (1992, as cited in Tabachnick & Fidell) criteria:

- 0.71 or greater (50% overlapping variance between variable and factor): excellent;
- 0.63 or greater (40% overlapping variance): very good;
- 0.55 or greater (30% overlapping variance): good;
- 0.45 or greater (20% overlapping variance): fair;
- 0.32 (10% overlapping variance): poor.

The cutoff value of 0.45 was selected to be conservative. Six iterations of factor analysis were performed. During each iteration, items that loaded strongly on multiple factors were eliminated (Tabachnick & Fidell, 2007). These findings are presented in Table 4 and show the item loadings for each factor, as well as the total variance explained. There were no significant cross-loadings of any items on more than one factor. At the end of the sixth round, six factors and 33 items remained.

Interpretation of the Factors. As part of EFA, factors were interpreted based on their make-up and labeled accordingly. The six factors, Focused Attention, Perceived Usability, Aesthetics, Endurability, Novelty, and Felt Involvement are described in the following section, according to the amount of variance explained by each factor, the resulting number of items, item loadings, and alpha values. With respect to alpha

values, it was necessary to calculate the reliability of the items that loaded on each of the six factors. First, interitem correlations were examined: no negative interitem correlations were found among the variables that loaded on any of the six factors. Second, reliability of the factors was computed using Cronbach's alpha.

Factor 1: Focused Attention. This factor accounted for 29.73% of the variance and consisted of nine items. These items related to users' perceptions of time passing and their degree of awareness about what was taking place outside of their interaction with the shopping Website. The remaining items pertained to users' ability to become absorbed and lose themselves in the shopping experience. Item loadings on this factor ranged from 0.57 to 0.85. Since the items pertained to absorption, awareness, and perceptions of time passing, this factor was labeled "Focused Attention." The calculated alpha value for the nine Focused Attention items was 0.928. This suggested that some items should be removed, according to Devellis' (2003) recommendations. An examination of the items revealed that there was little variability among three items and thus two were systematically removed over two iterations, resulting in an alpha value of 0.9 for this seven-item factor.

Factor 2: Perceived Usability. This factor was defined by eight items and accounted for 15.63% of the total rotated variance. Item loadings ranged from 0.57 to 0.75. This factor's items pertained to the emotions experienced by respondents when completing their shopping task, i.e., "annoyed," "frustrated," "stimulated," and "discouraged." Related to these emotions were items that tapped into the challenge required to shop on the Website (i.e., "taxing") and their perceptions of the navigation of the e-commerce site as being "confusing." Items also measured whether users felt they could perform the tasks they wanted to through the Website, and their perceived control over the interaction. Overall, these items assessed users' perceived effort in using the Website, their ability to accomplish their shopping tasks, the navigation and organization of the Website, and the emotions evoked by using the Website. Thus, "Perceived Usability" was an appropriate label. The calculated value of alpha for the eight items was

TABLE 4. Principal components factor analysis with Varimax rotation.

Item #	Item	Focused attention	Perceived usability	Aesthetics	Endurability	Novelty	Involvement
Q133	I forgot about my immediate surroundings while shopping on this website.	0.850	0.00	0.00	0.00	0.00	0.00
Q140	I was so involved in my shopping task that I ignored everything around me.	0.838	0.00	0.00	0.00	0.00	0.00
Q243	I lost myself in this shopping experience.	0.831	0.00	0.00	0.00	0.00	0.00
Q145	I was so involved in my shopping task that I lost track of time.	0.821	0.00	0.00	0.00	0.00	0.00
Q130	I blocked out things around me when I was shopping on this website.	0.764	0.00	0.00	0.00	0.00	0.00
Q129	When I was shopping, I lost track of this world around me.	0.727	0.00	0.00	0.00	0.00	0.00
Q240	The time I spent shopping just slipped away.	0.695	0.00	0.00	0.00	0.00	0.00
Q97	I was absorbed in my shopping task.	0.583	0.00	0.00	0.00	0.00	0.00
Q232	During this shopping experience I let myself go.	0.573	0.00	0.00	0.00	0.00	0.00
Q64	I felt frustrated while visiting this shopping website.	0.00	0.754	0.00	0.00	0.00	0.00
Q271	I found this shopping website confusing to use.	0.00	0.732	0.00	0.00	0.00	0.00
Q53	I felt annoyed while visiting this shopping website.	0.00	0.716	0.00	0.00	0.00	0.00
Q69	I felt discouraged while shopping on this website.	0.00	0.685	0.00	0.00	0.00	0.00
Q173	Using this shopping website was mentally taxing.	0.00	0.677	0.00	0.00	0.00	0.00
Q177a	This shopping experience was demanding.	0.00	0.666	0.00	0.00	0.00	0.00
Q192	I felt in control of my shopping experience.	0.00	0.617	0.00	0.00	0.00	0.00
Q208	I could not do some of the things I needed to do on this shopping website.	0.00	0.571	0.00	0.00	0.00	0.00
Q33	This shopping website is attractive.	0.00	0.00	0.805	0.00	0.00	0.00
Q28	This shopping website was aesthetically appealing.	0.00	0.00	0.799	0.00	0.00	0.00
Q6	I liked the graphics and images used on this shopping website.	0.00	0.00	0.784	0.00	0.00	0.00
Q446	This shopping website appealed to my visual senses.	0.00	0.00	0.738	0.00	0.00	0.00
Q27	The screen layout of this shopping website was visually pleasing.	0.00	0.00	0.709	0.00	0.00	0.00
Q405	Shopping on this website was worthwhile.	0.00	0.00	0.00	0.748	0.00	0.00
Q408	I consider my shopping experience a success.	0.00	0.00	0.00	0.713	0.00	0.00
Q209	This shopping experience did not work out the way I had planned.	0.00	0.00	0.00	0.666	0.00	0.00
Q407	My shopping experience was rewarding.	0.00	0.00	0.00	0.661	0.00	0.00
Q296	I would recommend shopping on this website to my friends and family.	0.00	0.00	0.00	0.564	0.00	0.00
Q416	I continued to shop on this website out of curiosity.	0.00	0.00	0.00	0.00	0.650	0.00
Q419	The content of the shopping website incited my curiosity.	0.00	0.00	0.00	0.00	0.538	0.00
Q364	I felt interested in my shopping task.	0.00	0.00	0.00	0.00	0.518	0.00
Q237	I was really drawn into my shopping task.	0.00	0.00	0.00	0.00	0.00	0.750
Q228	I felt involved in this shopping task.	0.00	0.00	0.00	0.00	0.00	0.640
Q61	This shopping experience was fun.	0.00	0.00	0.00	0.00	0.00	0.508
Amount of variance explained		*9.811*	*5.160*	*2.431*	*1.209*	*1.147*	*1.026*
Percentage of variance explained		*29.732*	*15.637*	*7.368*	*3.662*	*3.475*	*3.108*

0.884. Since this fell into the "very good" range, all items were retained.

Factor 3: Aesthetics. This factor consisted of five items and accounted for 7.36% of the variance. Items loadings ranged from 0.71 to 0.80. This set of items pertained to specific features of the interface, such as the screen layout and graphics/images, and to respondents' overall aesthetic impressions of the Website's attractiveness and sensory appeal. Since the items all related to the visual appearance of the interface, "Aesthetics" was a fitting name for this factor. Cronbach's alpha for the items comprising this factor was 0.89. This was a "very good" value and there was no need to remove any of the items.

Factor 4: Endurability. This factor was defined by five items and accounted for 3.66% of the total variance. Item loadings were from 0.56 to 0.72. Items assessed respondents' likelihood to recommend the shopping Website to others, as well as to perceive shopping experience as "successful," "rewarding," "worthwhile," and working out as planned. Overall these items measured respondents' willingness to return to the shopping Website and to recommend the Website to others, as well as their overall evaluations of the experience. "Endurability," the likelihood to remember things that we have enjoyed and a desire to do again an activity that has been fun (Read, MacFarlane, & Casey, 2002), was an appropriate name for this factor. Cronbach's alpha for this factor was 0.843. This was considered a "very good" outcome.

Removing an item would have resulted in a slight higher alpha value (0.853), but it was retained because there was no statistical need to increase alpha; retention of five items, rather than four, was more conservative at this exploratory stage of scale evaluation.

Factor 5: Novelty. The fifth factor comprised 3.47% of the variance and consisted of three items with loadings between 0.518 to 0.650. Each of these spoke to the curiosity evoked by or participants' interest in the shopping task. Stimulating respondents' curiosity indicated that the shopping Website or experience contained surprising, unexpected, or new information at various points in time. Thus, this factor was called "Novelty." The alpha value for this factor was 0.73. All three items were kept.

Factor 6: Felt Involvement. The sixth factor consisted of three items with loadings that ranged from 0.50 to 0.75 and comprised 3.108% of the variance. These items pertained to respondents feeling of being drawn into and involved in the shopping task and their overall assessment of the experience as "fun." Involvement is a "need-based cognitive (or belief) state of psychological identification with some object" that is based on an individual's salient needs and perception that the object will satisfy those needs (Kappelman, 1995, p. 66). This label was adopted for this factor because perceptions of involvement and fun in engagement are based on the level of importance, significance, or relevance (Kappelman, 1995) given to an experience by the user. The calculated alpha for this factor was 0.723. This was a respectable value and all three items that constituted this factor were retained.

Correlation analysis. Table 5 shows the descriptive statistics and correlations for the variables retained after the steps of factor analysis and reliability testing. The correlations among the factors were all found to be significant at the $p = 0.01$ level. This demonstrated that each factor was distinct, with no significant overlap and no additional factors present. The 15 calculated associations were in the low to moderate range:

- Focused Attention was minimally associated with Perceived Usability ($r^2 = 0.18$) and Aesthetics ($r^2 = 0.19$), and moderately correlated with Novelty ($r^2 = 0.47$) and Felt Involvement ($r^2 = 0.37$).

- Perceived Usability was moderately correlated with Aesthetics ($r^2 = 0.41$), Endurability ($r^2 = 0.58$), Novelty ($r^2 = 0.3677$), and Felt Involvement ($r^2 = 0.47$).
- Aesthetics was moderately correlated with Endurability ($r^2 = 0.41$), Novelty ($r^2 = 0.40$), and Felt Involvement ($r^2 = 0.564$).
- Novelty and Felt Involvement were moderately correlated with each other ($r^2 = 0.51$), and with Endurability ($r^2 = 0.43$ and $r^2 = 0.51$, respectively).

Summary

This analysis evaluated the reliability of the engagement scale using the responses of 440 general shoppers to the 123 items. In the reliability analysis, the redundant items were removed leaving 49 items representing 10 internally consistent attribute-based subscales. Next, EFA reduced the 49-item instrument to 33 items dispersed over six factors through six iterations. The internal consistency of the resulting factor structure was examined. Prior to EFA, 10 internally consistent subscales existed. After EFA, only two of these subscales, aesthetics and novelty, were identified as distinct factors. Items on the other subscales loaded on four factors that were interpreted based on the item loadings and labeled as Focused Attention, Perceived Usability, Endurability, and Felt Involvement. Lastly, the reliability of the resulting six factors was examined using Cronbach's alpha and two additional items were removed. Overall, the factors were interpreted and labeled: Focused Attention (factor 1), Perceived Usability (factor 2), Aesthetics (factor 3), Endurability, (factor 4), Novelty (factor 5), and Felt Involvement (factor 6). With a parsimonious, reliable scale, we next sought to examine the validity of the instrument in Study 2.

Study 2: Evaluating the Validity of the Engagement Scale

The proposed Engagement Scale was reduced and evaluated for reliability in Study 1; the next test in the development process was to assess discriminant validity; in other words, would we obtain similar results with another sample? At the same time, we added to our analysis a test of the relationships among the attributes, a test not always performed in scale development and evaluation. In this case, the scale represented multiple dimensions, and an outstanding research question concerning all of these attributes concerned the

TABLE 5. Descriptive statistics, internal consistency values, and intercorrelations for factors.

Factors	M	SD	α	1	2	3	4	5
1. Focused Attention	1.89	0.56	0.9	1.0000				
2. Perceived Usability	3.14	0.33	0.844	0.182**	—			
3. Aesthetics	3.53	0.68	0.89	0.199**	0.410**	—		
4. Endurability	3.84	0.70	0.843	0.138**	0.589**	0.412**	—	
5. Novelty	3.39	0.75	0.730	0.476**	0.367**	0.406**	0.432**	—
6. Felt Involvement	3.51	0.70	0.723	0.372**	0.464**	0.477**	0.515**	0.514**

Note. Listwise N for correlations = 394; α Cronbach's index of internal consistency for revised scale.

**Correlation is significant at the 0.01 level (2-tailed).

relationships among them. Typically, in the past, each of the attributes was considered and tested as independent dimensions, but we speculated that relationships existed among them. For example, usability, aesthetics, etc., are typically looked at together, yet in isolation from other aspects of interactive experiences. As a consequence, we first tested the validity of the scale by confirming the factor structure with the results of Study 1, and then tested the relationships among the factors represented in the scale.

To accomplish both of these objectives, we used Structural Equation Modeling (SEM), which combines Confirmatory Factor Analysis (CFA), which assesses whether or not the factor structure of data in Study 1 is comparable to Study 2, and Path Analysis (PA), which examines the predictive relationships among the resulting factors. As prescribed in SEM, it is imperative to construct hypotheses prior to collecting and analyzing the data. The following section describes the hypotheses generated for SEM, which are based on prior research.

First, we predicted that the current study would yield the same results as our previous study regarding the six-factor structure of the engagement scale.

H1: The Engagement Scale is comprised of six distinct factors: Focused Attention, Perceived Usability, Aesthetics, Endurability, Novelty, and Felt Involvement

In addition to confirming and thus replicating the results from Study 1, we proposed that a series of relationships existed among the six factors. These relationships are described in Hypotheses Two through Six.

Aesthetics. Existing definitions of engagement have suggested that engaging systems "catch and captivate user interests," "draw people in," and "encourage interactions" (Quesenbury, 2003, p. 86). Jennings (2000) related these aspects of focused attention and involvement to aesthetic experiences. In addition, Jacques et al. (1995) found that multimedia users demonstrated strong preferences for visually based multimedia; others (Lingaard et al., 2006; Shenkman & Jonsson, 2000) have found that aesthetic elements (i.e., illustrations) contributed to Web users' first impressions of a Website. We predicted that these initial perceptions would determine the level of Focused Attention users would invest and the involvement they would achieve. As a result, we hypothesize that:

H2A: Aesthetics predicts Focused Attention
H2B: Aesthetics predicts Felt Involvement

Novelty. Pace (2004) suggested that novelty in online content has the potential to sustain users' attention, specifically when novelty is introduced through links and content that are pertinent to users' goals. Pace emphasized the "congruence" between interest, novelty, and searchers' goals in directed attention. This is concurrent with the view that there is an "optimal level" of arousal; diverging from this level may have negative consequences for users' experiences (Hebb,

1972). The relationship between novelty and focused attention may be approached from the state-trait curiosity model (Boyle, 1983). Boyle stated that external and internal stimuli, as well as individual differences, combine to form a cognitive appraisal of a situation, which results in behavior. An illustration of this would be a Web searcher who sees a hyperlink (external stimuli) to further information on a topic of personal interest (internal stimuli) and, depending on their sense of time or urgency (individual difference), determines whether or not they will click on the link (behavior). In the act of cognitive appraisal, stimuli are attended to and processed. Based on the research linking novelty with attention and interest, we hypothesized that:

H3A: Novelty predicts Focused Attention
H3B: Novelty predicts Felt Involvement

Focused Attention and Perceived Usability. In this work, the Focused Attention factor is defined by items pertaining to focused attention, awareness, and perceptions of time. These facets represent some, though not all, characteristics of Flow. Since prior research explored Flow and aspects of usability (e.g., Pace, 2004), this is a useful starting point for examining the relationship between Focused Attention and Perceived Usability. Pace (2004) found that the match between perceived challenge and searchers' skill was directly related to participants' ability to achieve a state of flow. Furthermore, Pace found that poor system usability, exemplified by lengthy response times, disorganized content, inconsistent navigation cues, cluttered page layout, etc., may interrupt flow by directing users' attention away from the salient aspects of the experience. Thus:

H4: Focused Attention predicts Perceived Usability

Felt Involvement and Perceived Usability. The Felt Involvement factor contained items about how much fun users' were having during the interaction and how drawn in they were able to become. Mathwick and Ridgon (2004) found that the quality of users' interactions depended on the degree of challenge presented by an online search task, the skills they possessed to meet those challenges, and their perceptions of control over the interaction. Pace (2004) stated that poor usability could impede users' enjoyment. Based on this, it would seem that Perceived Usability would predict Felt Involvement. However, our instrument measures postexperience Perceived Usability. Therefore, if the user has experienced Felt Involvement, it is because the usability of the system did not interrupt or prevent them from enjoying themselves; in this situation, judgments of the Perceived Usability will be influenced by the level of involvement achieved. Thus:

H5: Felt Involvement predicts Perceived Usability

Endurability. Endurability is the assessment of users' perception of success with a task, and their willingness to use an application in future or recommend it to others. The Endurability factor is really the summation of the experience and is therefore positioned as the outcome variable in the model.
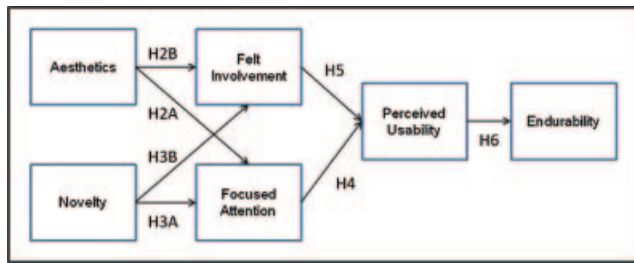
FIG. 2.   Proposed SEM engagement model showing hypothesized relationships.

Users' evaluations of the Aesthetics and Novelty of an application will influence their degree of Focused Attention and Felt Involvement, and subsequently their Perceived Usability of the system. All of these factors will predict users' lasting impressions of the experience and their willingness to engage with the application at another point in time. The final hypothesis is that:

H6: Perceived Usability predicts Endurability

From these predictions, a fully mediated model emerges where Perceived Usability mediates the relationship between the Endurability of the experience and the other four factors. In other words, Aesthetics, Novelty, Felt Involvement, and Focused Attention are filtered by users' perceptions of the system's usability; this mediation leads to users' appraisal of the overall experience (see Figure 2).

Objectives

The objectives of Study 2 were: 1. To validate the six-factor structure of the Engagement Scale that resulted from Study 1; 2. To test the predictive relationships hypothesized among the six factors of Focused Attention, Perceived Usability, Aesthetics, Endurability, Novelty, and Felt Involvement. To address these objectives, we undertook another large-scale survey in the e-commerce environment, this time focusing on customers of a specific online retailer rather than general shoppers. We used the 31-item instrument that resulted from Study 1.

Methodology

The online survey.   As in Study 1, the scale was administered using a modified version of Perseus Software Solutions. The Web-based survey consisted of nine Web pages: The introductory page explained the purpose of the survey, the length of time it would take to complete the survey, and their rights as a participant in the study, e.g., ability to withdraw at any time or to decline answering any question. Proceeding to the next screen was viewed as "informed consent." Pages 2–7 contained the engagement scale items. There were approximately six randomized questions per screen. As in Study 1, all items were rated using a five-point Likert scale ranging from strongly disagree (1) to strongly agree (5) with a

TABLE 6.   Participant demographics, Study 2.

	N	%
Sex		
Male	208	26
Female	562	70.1
Not Specified	32	3.9
Age		
14–20	27	3.37
21–30	109	13.6
31–40	176	21.97
41–50	195	24.34
51–60	154	19.3
61–83	123	15.36
Not specified	18	2.2
Residence		
Urban	487	60.7
Rural	200	24.9
Not specified	115	14.4
Education**		
High School	376	46.9
Technical College	76	9.5
Community College	216	26.9
Undergraduate University	283	35.3
Graduate University	100	12.5
Professional	60	7.5
Other	70	8.7
Occupation*		
Art, Culture, Recreation, Sport	30	3.8
Business, Finance, Administration	126	15.7
Health	60	7.5
Management	68	8.5
Natural and Applied Sciences	55	6.9
Processing, Manufacturing	19	2.4
Primary Industry	4	0.5
Sales and Services	61	7.7
Social Science, Education, Government, Religion	102	12.7
Trades, Transport, Equipment	15	1.9
Other/ Not Specified	262	32.7

*Based on National Occupational Classification: Major Category.
**This question was "check all that apply" so numbers do not add up to 100%.

sixth option for "not applicable." Page 8 contained demographic questions. The Final Page thanked participants for their participation and reiterated the contact information of the researcher.

Participants.   The online survey was active for 3 weeks (between June 11 and July 4, 2007) and was administered in cooperation with a major online book retailer, who sent an email to 10,000 customers who had purchased an item from the company's Website within the last 3 months. A total of 802 customers completed the online survey. Although the response rate was less than 10%, the number of responses was adequate to proceed with data analysis. Table 6 summarizes the participant demographics.

Procedure.   The survey was prepared and mounted on a secure server. Participants clicked on the URL in the recruitment notice and were taken to the introductory page of the survey. A progress bar located at the bottom of the screen

informed participants about the page they were currently on in relation to the total number in the survey. Respondents navigated the content of the survey by clicking on the "Next" button at the bottom left-hand corner of the screen.

Data preparation. Some items were reverse-coded (where applicable) and values of 6 on the Likert scale (indicating the question was "not applicable") were recoded to reflect missing data. The frequencies of valid responses were examined to identify questions with low response rates. Since missing responses ranged from 1 (1.2%) to 50 (6.3%) of the total responses, no questions were excluded from further analysis. Results were analyzed using SPSS and LISREL 8.8 data analysis tools. SEM was used to perform CFA and PA in order to validate the six-factor structure of the scale and the hypothesized relationships among the factors (see Figure 2).

Results

Before proceeding to SEM, we assessed the reliability of the six subscales representing each engagement factor using Cronbach's alpha. This was done in order to evaluate the consistency between Study 1 and Study 2 with respect to the items that made up each factor. As in Study 1, 0.7 to 0.9 was considered "respectable" to "very good" (DeVellis, 2003). Table 7 shows the alpha values for the engagement attributes.

All of the calculations, except for Novelty, were acceptable at above 0.7. The reason for this low value may be that this factor contained a small number of items; this is one aspect, along with interitem correlations and dimensionality, that may affect alpha levels (Cortina, 1993). Steps were taken to determine how to proceed. First, the interitem correlations of the three Novelty items were examined and were

all significant at the $p < 0.001$ level. Second, Principal Components Analysis (PCA) was performed to ensure that the Novelty items formed a unidimensional construct and did not overlap significantly on other factors or stand alone. This exploration revealed that the Novelty items formed their own factor with factor loadings ranging from 0.57–0.75.

PCA was then run with all 31 items to ensure that a single factor did not emerge. Correlations for the measures used in the study were calculated (Table 8). The factors were moderately associated with each other and all correlations were significant at the 0.01 level. The correlation between Focused Attention and Perceived Usability was 0.245, and between Perceived Usability and Novelty was 0.330. The remaining factors had associations between 0.4 and 0.5 (one-third of the correlations calculated), between 0.5 and 0.6 (20% of the correlations calculated), or between 0.6 and 0.7 (20% of the correlations calculated). The highest correlation was between Focused Attention and Felt Involvement (0.703). Overall, there was internal consistency among the six factors and no additional factors present.

The results of this initial stage of analysis confirmed that the scale was multidimensional and that the items loading on each factor were consistent with the findings of the previous study. These findings indicated common method variance was not an issue (Kline, Sulsky, & Rever-Moriyama, 2000) and further hypothesis testing was appropriate. In other words, we were able to confidently proceed to SEM analysis.

SEM: calculating the model of "best fit." LISREL software (v. 8.8) was used to perform a series of nested model comparisons in order to examine the relationships among the attributes of engagement. SEM involves determining the "goodness of fit" between the observed and theoretical factor structures. To increase the robustness of the results the 802 cases were randomly split into two separate files using SPSS. One file contained 75% ($n = 478$) of the cases and was used to test the proposed model. The other contained 25% (n = 143) of the cases and was used to confirm the final model. In each case, the following indices of fit were calculated: root mean square error of approximation (RMSEA), goodness of fit (GFI), adjusted goodness of fit (AGFI), normed fit (NFI), comparative fit (CFI), and parsimonious fit (PFI).

Calculated with 75% of the cases, the predicted fully mediated model was not found to be satisfactory according to
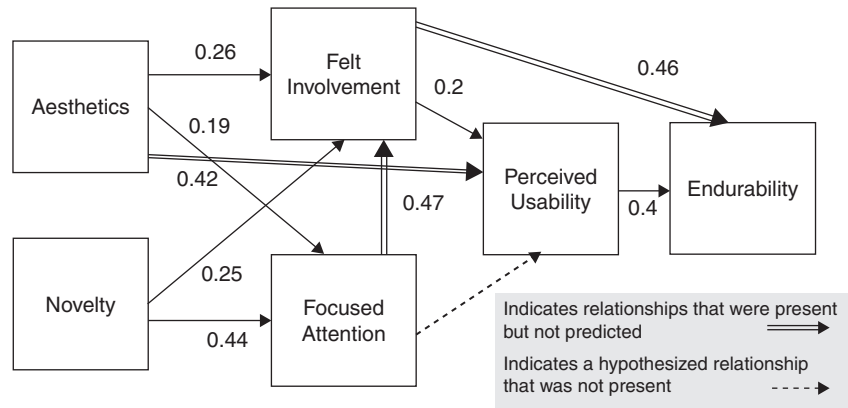
TABLE 7. Descriptive statistics and reliability estimates of factors.

Factor	Number of items	Cronbach's alpha value
Focused Attention	7	0.921
Perceived Usability	8	0.916
Aesthetics	5	0.898
Endurability	5	0.866
Novelty	3	0.588
Involvement	3	0.707

TABLE 8. Intercorrelations for factors.

Measures	M	SD	α	Attention	Usability	Aesthetics	Endurability	Novelty
Attention	2.69	0.783	0.92	1.00				
Usability	4.00	0.667	0.91	0.245**	1.00			
Aesthetics	3.62	0.617	0.89	0.409**	0.530**	1.00		
Endurability	4.05	0.599	0.86	0.422**	0.689**	0.504**	1.00	
Novelty	3.26	0.623	0.58	0.541**	0.330**	0.462**	0.463**	1.00
Involvement	3.47	0.636	0.70	0.703**	0.434**	0.548**	0.623**	0.625**

Note. Listwise N for correlations = 617.

α Cronbach's index of internal consistency for revised scale.

**Correlation is significant at the 0.01 level (2-tailed).

FIG. 3. Final structural equation model.

the fit indices ($\chi^2 = 456.45$, $df = 7$; $p < 0.00$; CFI = 0.79; RMSEA = 0.323). A better fitting model of the data involved adding three additional pathways (between Aesthetics and Perceived Usability, between Focused Attention and Felt Involvement, and between Felt Involvement and Endurability) and removing one pathway (between Focused Attention and Perceived Usability).

The resulting model ($\chi^2 = 13.82$, $df = 5$; $p < 0.016$; CFI = 1.00; RMSEA = 0.096) produced GFI and NFI values of 0.99. The test of close fit showed that the RMSEA ($p = 0.27$) was not significant, indicating that this model was a good fit. The RMSEA values should be below 0.10 in order to be considered a good fit for the data; ours was calculated to be 0.096 (Kelloway, 1998).

The path model was then tested with the remaining 25% of the cases ($\chi^2 = 13.57$, $df = 5$; $p = 0.018$; CFI = 0.99; RMSEA = 0.111). The results were similar to those obtained using 478 cases. The GFI (0.97) and NFI (0.98) were both at acceptable levels; RMSEA was slightly above the optimal value of 0.10, but the p-value of RMSEA ($p = 0.069$) was not significant, indicating adequate fit for the model.

Path analysis. The final structural equation model (see Figure 3) contained three additional pathways (indicated with double lines):

- Aesthetics predicts Perceived Usability,
- Focused Attention predicts Felt Involvement, and
- Felt Involvement predicts Endurability.

The predicted relationship between Focused Attention and Perceived Usability was not supported by the data. Figure 3 includes the standardized parameter estimates in the path model. As demonstrated, Aesthetics and Novelty predicted Focused Attention ($\beta = 0.69$, $p < 0.01$) and Felt Involvement ($\beta = 0.38$, $p < 0.01$), Felt Involvement predicted Perceived Usability ($\beta = 0.69$, $p < 0.01$), and Perceived Usability and Felt Involvement predicted Endurability ($\beta = 0.47$, $p < 0.01$).

Maximum likelihood estimates indicated that Perceived Usability was a product of Aesthetics and Felt Involvement ($r^2 = 0.31$) and this relationship explained 31% of the error variance. Focused Attention was mediated by Novelty and

Aesthetics ($r^2 = 0.31$) and explained 41% of the error variance. Felt Involvement was predicted by Focused Attention, Novelty, and Aesthetics ($r^2 = 0.62$) and explained 15% of the error variance. Endurability was the outcome variable and was predicted by Perceived Usability and Felt Involvement ($r^2 = 0.53$) and explained 17% of the variance.

Summary

The resulting path model (Figure 3) contained more relationships than predicted in the proposed model (Figure 2). Overall, Endurability was predicted by the other five factors. Aesthetics predicted Perceived Usability, Focused Attention, and Felt Involvement; Novelty predicted Focused Attention and Felt Involvement. The fact that the actual model was a modification of the predicted model indicates that this model must be retested in future with another independent sample.

Discussion

The attributes identified through previous research and an exploratory study (O'Brien & Toms, 2008) formed the foundation for constructing a multidimensional scale to measure engaging user experiences with technology. In this paper we have described our process of survey development and two studies to evaluate the instrument's reliability and validity in the online shopping environment. The outcome of this research is a reliable and valid scale comprised of six distinct factors: Focused Attention, Perceived Usability, Endurability, Novelty, Aesthetics, and Felt Involvement.

*The Composition of the Engagement Survey:
From Subscale to Factors*

Initially we proposed a range of attributes: aesthetics, focused attention, awareness, challenge, control, feedback, interest, motivation, novelty, perceived time, continued use, etc., in the compilation of subscales. Preliminary data analysis in Study 1 demonstrated that these subscales were internally consistent prior to performing EFA. Yet EFA yielded six factors, rather than 10 subscales, where only novelty and aesthetics translated directly into factors. The remaining

four factors contained items from the remaining eight sub-scales. For example, the Focused Attention factor included items from the engagement and perceived time subscales, as well as attention; the Perceived Usability factor consisted of items from the affect, challenge, control, and feedback sub-scales; and items from the affect and motivation subscales contributed to the Felt Involvement and Endurability factors. Felt Involvement also consisted of some engagement items; Endurability had items from the reengagement subscale. The composition of these factors was supported by the results of Study 2.

The ways in which these different subscales contributed to the creation of broader factors is worthy of exploration. Affective items that dealt with personal feelings and motivations were dispersed throughout the Perceived Usability, Felt Involvement, and Endurability factors. This may be because emotions influenced both impressions of applications and evaluations of their use (Lindgaard, 2004). According to Forlizzi and Battarbee (2004, p. 163), "emotion affects how we plan to interact with products, how we actually interact with products, and the perceptions and outcomes that surround those interactions. Emotion serves as a resource for understanding and communicating about what we experience." The presence of affective and motivation items in more than one factor was not incomprehensible. It corresponds with user experience frameworks, such as McCarthy and Wright (2004), which embed emotion throughout all stages of experience and the current view that computer-user interactions involve more than behavior and cognition (Nahl, 2007). Our findings with regard to affect support a more holistic representation of user engagement that indicates affect should be incorporated into interaction design and measurement.

Relationships Among the Factors

We examined the relationships among the six factors using PA. The model of best fit (Figure 3) contained three additional and one less path than the hypothesized model (Figure 2). Here we turn our attention to further elaboration of these results.

Aesthetics and Novelty predicted the other factors in the model. This pairing suggests that developers make "creative use" of visual and sensory features as a means of attracting users' attention (Jacques, 1996, p. 93). However, Aesthetics was not only related to Focused Attention and Felt Involvement as hypothesized, but directly to Perceived Usability. This relationship suggests that aesthetic judgments are not based solely on users' first impressions; the perceived usability of a system is intertwined with its visual presentation. All of the pathways that originate with Aesthetics support the idea that Aesthetics is experienced as more than visual appeal, and is also core to creating meaning and satisfying the need for wholeness (Beardsley, 1982) as embodied in users' lasting impressions of the Endurability of the experience. This may seem contradictory to Lingaard et al.'s (2006) findings about the snap judgments made about an interface regarding its appearance. However, it should be remembered that participants in our studies shared their reflections about completing an online shopping task, while Lingaard et al. focused on users' intention to use rather than actual use.

Endurability was an outcome variable of the model and pertained to shoppers' holistic evaluations of the experience. This is not unlike the Information System Success (ISS) model (DeLone & McLean, 1992) that related users' perception of system quality and content quality to their actual use and satisfaction. In our model, Endurability, the sum of how rewarding, successful, and worthwhile the interaction was, and how likely participants would be to use the system again or recommend it to others, corroborates the influence of user and system variables on overall perceptions of experience.

The hypothesis not supported by the data (indicated with a dotted line in Figure 3) was that Focused Attention would be mediated through Perceived Usability. Instead, Felt Involvement mediated the relationship between Focused Attention and Perceived Usability and, indirectly, Endurability. Felt Involvement had a direct relationship with Perceived Usability, and also with Endurability. While Perceived Usability was the "hub" in the model as hypothesized, Felt Involvement emerged as a more defining factor than predicted. This raises the question of whether personal interest, involvement, and fun are as powerful as usability in influencing users' appraisals of an experience. This supports related work that advocates creating positive affective experiences and views emotion as an essential component in system development (e.g., Laurel, 1993; Nahl, 2007; Pace, 2004).

Limitations

A limitation of this work lies in its use of an online survey as a method of data collection. In Study 1, the survey was lengthy, consisting of 123 items, and in Study 2 our response rate was low (8.02% or 802 respondents out of a sample of 10,000). Typically response rates of 30% or more are considered adequate (Cooke, Heath, & Thompson, 2000). However, to minimize the effects of participant fatigue all questions in both administrations of the survey were randomized (i.e., responses to items were not biased based on when they appeared during the process of completing the survey). Overall, the number of responses obtained more than satisfied the parameters for conducting EFA and SEM analyses.

The nature of survey research means that biases may result from differences between those who responded versus those who did not. However, the demographics of the surveyed groups indicated that we had a heterogeneous sample of various ages, occupations (e.g., not all students), and geographic areas. Although there were more female than male respondents ($\approx$70%:30%) in both studies, gender effects were examined in the data but were not found.

Future Work

This work focused on the development of a scale to measure user engagement and evaluated its reliability and validity. The next steps in the process are to assess generalizability and criterion validity (Peterson, 2000).

Generalizability. Our next goal is to examine the generalizability of the survey instrument in domains and applications outside of online shopping. Given that the same six factors emerged from the data collected with both general shoppers (Study 1) and customers of a particular retailer (Study 2), it is conceivable that the instrument could be generalized to other environments, such as the Web, digital libraries, or task-specific applications (e.g., bibliographic citation or qualitative data analysis software). Although constructs such as flow (e.g., Webster, Trevino, & Ryan, 1993; Finneran & Zhang, 2003), play (Mathwick & Rigdon, 2004), and interactivity (Liu, 2003) have been studied in these contexts, engagement has not. Another issue in the generalizability of the scale is that previous research indicated that format influences engagement (Jacques et al., 1995). In our previous work we found subtle differences among gamers, shoppers, learners, and searchers regarding the manifestation of some of the engagement attributes, particularly among those designed for individual versus collaborative use (O'Brien & Toms, 2008). Another area of generalizability is in individual versus group-based tools, given that social interaction may add another level of complexity and engagement to the use of technology. Lastly, we administered the Engagement Scale to respondents who recalled a past shopping experience. Although their retrospective accounts reflected the memorability and endurability of the experience, the recency of the shopping experience (immediately after versus within the last 3 months) may have influenced the responses. It would be interesting to compare different samples that vary in the immediacy of their responses.

Criterion validity. Another aspect of future work is exploring the Engagement Scale in the context of other types of validity, such as criterion validity, which examines the relationship and coherence among different types of metrics. Possible methods for demonstrating criterion validity of the Engagement Scale include eye tracking data, usability metrics, and biometrics. Eye tracking metrics (i.e., gaze positions, fixation number, fixation duration, repeat fixations, search patterns, blink rates, and blink durations) have been associated with attention, affect, interest, and novelty (e.g., Vick & Ikehara, 2002) and have been used to unobtrusively compare different interface designs and users' ability to locate interface features (Bojko, 2006). Performance and usability metrics could be employed with respect to engagement, but not in the traditional sense since the Perceived Usability factor of the scale incorporated effectiveness and satisfaction, but not efficiency. This makes sense since an engaged user may not be concerned about how much time they are spending using an application. However, it does mean that some usability metrics, including time spent performing a task or examining a screen, may not be applicable.

Lastly, biometrics (e.g., galvanic skin response, heart rate, electromyography of the jaw, respiration rate, and respiration amplitude) may be another route for assessing the criterion validity of the engagement instrument. Research has attempted to relate physical and subjective measures of fun in video games (Mandryk, 2004) using biometrics. Given the obtrusiveness of biometrics, and the fact that conducting usability, physiological, and eye tracking studies can be involved and require special, costly equipment, demonstrating the criterion validity of the scale may offer the design and research communities a more cost-effective, quick, and accessible means of evaluating user engagement with technologies and be used in various stages of interface design.

Another aspect of future work is to address its relationship with other scales (e.g., Technology Acceptance Model, Cognitive Absorption, etc.) in order to ascertain whether the Engagement Scale correlates with other measures. This will enable us to examine theoretical/empirical redundancies and differences (Clark & Watson, 1995) between the Engagement Scale and other metrics.

Conclusion

In this research we developed and evaluated the reliability and validity of a scale developed to measure user engagement in online shopping environments. The result is a survey instrument comprised of six factors: Perceived Usability, Aesthetics, Novelty, Felt Involvement, Focused Attention, and Endurability. In addition to identifying and confirming these six factors, we articulated the relationships among them using PA. Perceived Usability and Felt Involvement were integral components of the path model that mediated the relationships between Aesthetics, Novelty, Focused Attention with the outcome variable, Endurability. Overall, the six factors identified were all interconnected, demonstrating that there is a need during the design process to consider "the whole user experience" rather than "a single dimension" (Quesenbury, 2003, p. 89).

These results also have implications for evaluating the holistic nature of user experience. In the past, aesthetics, novelty, usability, etc., have typically been examined in isolation or in select sets. For example, Tractinsky (1997; Tractinsky, Katz, & Ikar, 2000), as well as Shenkman & Jonsson (2000), brought awareness to how the aesthetic qualities of interfaces influenced users' judgments of the usability and the selection of information content. We postulate that multiple factors of experience must be examined concurrently and are related to each other, and conclude that engagement, a quality of users experience with technology, is a multidimensional construct. Moving beyond measurement, the co-presence of multiple factors during experience will, in future, influence design guidelines.

The survey instrument itself is a brief, easily administered, and statistically verified tool that may be used by software designers to assess their applications or by researchers for academic purposes. For example, one could use it to determine users' ratings of a Website's aesthetics or novelty. The instrument was derived from the concepts of actual users and previous metrics and, as a result, is more comprehensive than previous measures and expands the scope of how engagement has been characterized in the past. It consolidates

disparate attributes (i.e., attention, challenge, curiosity, intrinsic interest, and control) (Webster & Ho, 1997; Chapman et al., 1999); curiosity, feedback, and challenge (Skelley et al., 1994), aesthetics (Chapman, 1997), and interaction, system format, and the presentation of the content (Jacques et al., 1995) into six factors that encompass the complex interaction between people and technology. The scale is a window into users' appraisals of the application and these perceptions are highly valuable in a marketplace, where there is a high degree of competition for users' time, interest, and money. Yet, its importance goes beyond the marketing of technology and into the fundamentals of what we are designing for and measuring in our work to enhance users' interactions with technology.

The next wave of research is shifting beyond human-computer interaction to human-computer experience; this represents the holistic and diverse relationships that people have with technology encountered in daily life. The scope of user experience—what one could be designing for and measuring—is broad. This work is a foundation upon which we can build a dialog about engagement, perform further testing of the model and scale, and examine additional methods for measuring the attributes and their relationships.

References

Aboulafia, A., & Bannon, L.J. (2004). Understanding affect in design: An outline conceptual framework. Theoretical Issues in Ergonomic Science, 5(1), 4–15.

Adams, D.A., Nelson, R.R., & Todd, P.A. (1992). Perceived usefulness, ease of use, and usage of information technology: A replication. MIS Quarterly, 16(2), 227–247.

Arnold, M.J., & Reynolds, K.E. (2003). Hedonic shopping motivations. Journal of Retailing, 79, 77–95.

Babin, B.J., Darden, W.R., & Griffin, M. (1994). Work and/or fun: Measuring hedonic and utilitarian shopping value. The Journal of Consumer Research, 20(4), 644–656.

Bannon, L. (2005). A human-centred perspective on interaction design. In A. Pirhonen, H. Isomäki, C. Roast, & P. Saariluoma (Eds.), Future Interaction Design (pp. 9–30). Berlin, Germany: Springer.

Beardsely, M. (1982). The Aesthetic Point of View: Selected Essays. Ithaca, NY: Cornell University Press.

Bojko, A. (2006). Using eye tracking to compate web page designs: A case study. Journal of Usability Studies, 3(1), 112–120.

Boyle, G.J. (1983). Critical review of state-trait curiosity test development. Motivation and Emotion, 7(4), 377–397.

Brooke, J. (1996). SUS: A 'quick and dirty' usability scale. In P.W. Jordan, B. Thomas, I.L. McClelland, & B. Weerdmeester (Eds.), Usability Evaluation in Industry. Philadelphia: Taylor and Francis.

Champion, E. (2003). Applying game design theory to virtual heritage environments. In Proceedings of the First International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia (pp. 273–274). Melbourne, Australia. New York: ACM Press.

Chapman, P. (1997). Models of Engagement: Intrinsically Motivated Interaction with Multimedia Learning Software. Unpublished master's thesis, University of Waterloo, Waterloo, Ontario, Canada.

Chapman, P., Selvarajah, S., & Webster, J. (1999). Engagement in multimedia training systems. In Proceedings of the 32nd Hawaii International Conference on System Sciences (HICSS-32) (Vol/Track 1, pp. 1–9). Los Alamitos, CA: IEEE Press. Retrieved September 17, 2009, from http://www.ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=772808&isnumber=16781

Childers, T.L., Carr, C.L., Peck, J., & Carson, S. (2001). Hedonic and utilitarian motivations for online retail shopping behavior. Journal of Retailing, 77, 511–535.

Choi, D., & Kim, J. (2004). Why people continue to play online games: In search of critical design factors to increase customer loyalty to online contests. Cyber Psychology and Behavior, 7(1), 11–24.

Clip Culture: Internet video. (2006). The Economist, 379(8475), 73.

Cook, C., Heath, F., & Thompson, R. (2000). A meta-analysis of response rates in Web- or Internet-based surveys. Educational & Psychological Measurement, 60(6), 821–836.

Cortina, J.M. (1993). What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology, 28(1), 98–104.

Davis, F.D., Bagozzi, R.P., & Warshaw, P.R. (1989). User acceptance of computer technology: A comparison of two theoretical models. Management Science, 35(8), 982–1003.

DeLone, W.H., & McLean, E.R. (1992). Information systems success: The quest for the dependent variable. Information Systems Research, 3(1), 60–95.

DeVellis, R.F. (2003). Scale development: Theory and applications (2nd ed. Vol. 26). Thousand Oaks, CA: Sage.

Douglas, Y., & Hargadon, A. (2000). The pleasure principle: Immersion, engagement, flow. In F. Shipman (Ed.), Proceedings of the 11th ACM Conference on Hypertext and Hypermedia (Hypertext 2000) (pp. 153–160). New York: ACM Press.

Feedback. (1995). The Penguin dictionary of psychology. Retrieved 29 April 2005, from xreferplus. http://www.xreferplus.com/entry/150004Finneran Zhang

Forlizzi, J., & Battarbee, K. (2004, August 1–4). Understanding experience in interactive systems. In D. Benyon, P. Moody, D. Gruen, & I. McAra-McWilliam (Eds.), Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS 2004) (pp. 261–268). New York: ACM Press.

Gefen, D., & Straub, D. (2000). The relative importance of perceived ease of use in IS adoption: A study of e-commerce adoption. Journal of the Association of Information Systems, 1(8), 1–20.

Ghani, J.A., Supnick, R., & Rooney, P. (1991, December 16–18). The experience of flow in computer mediated and in face to face groups. Paper presented at the International Conference on Information Systems, New York, NY.

Guay, F., Vallerand, R.J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). Motivation and Emotion, 24(3), 175–213.

Hargens, L.L. (2001). Sampling procedures. In Encyclopedia of Sociology, (2nd ed., Vol. 4, pp. 2444–2449). New York: Macmillan Reference.

Haywood, N., & Cairns, P. (2005). Engagement with an interactive museum exhibit. In T. McEwan, J. Gulliksen, & D. Benyon (Eds.), People and Computers XIX — The Bigger Picture, Proceedings of HCI 2005. London: Springer.

Hebb, D.O. (1972). Textbook of psychology, 3rd edition. Philadelphia: WB Saunders.

Huang, M. (2003). Designing website attributes to induce experiential encounters. Computers in Human Behavior, 19, 425–442.

Hull, R., & Reid, J. (2003). Designing engaging experiences with children and artistis. In M.A. Blythe, A.F. Monk, K. Overbeeke & P.C. Wright (Eds.), Funology: From usability to enjoyment (pp. 179–187). Amsterdam: Kluwer Academic Publishers.

Hutcheson, G.D., Sofroniou, N. (1999). The multivariate social Scientist: Introductory statistics using generalized linear models. London: Sage.

Hutchins E.L., Holland J.D., & Norman D.A. (1986). Direct manipulation interfaces. In D.A. Norman, S.W. Draper (Eds.), User centered system design: New perspectives on human-computer interaction. Mahwah, NJ: Lawrence Erlbaum Associates.

Jacques, R.D. (1996). The nature of engagement and its role in hypermedia evaluation and design. Unpublished doctoral dissertation, South Bank University, London.

Jacques, R., Preece, J., & Carey, T. (1995). Engagement as a design concept for multimedia. Canadian Journal of Educational Communication, 24(1), 49–59.

Jennings, M. (2000). Theory and models for creating engaging and immersive ecommerce websites. In Proceedings of the 2000 ACM SIGCPR Conference on Computer Personnel Research (pp. 77–85). New York: ACM Press.

Jones, M. (n.d.). Creating engagement in computer-based learning environments. ITFORUM Paper 30. Retrieved 5 March, 2005, from http://itech1.coe.uga.edu/itforum/paper30/paper30.html

Kappelman, L.A. (1995). Measuring user involvement: A diffusion of innovation perspective. Database Advances, 26(2/3), 65–86.

Kline, T.J.B., Sulsky, L.M., & Rever-Moriyama, S.D. (2000). Common method variance and specification errors: A practical approach to detection. The Journal of Psychology, 134(4), 401–421.

Konradt, U., & Sulz, K. (2001). The experience of flow in interacting with a hypermedia learning environment. Journal of Educational Multimedia and Hypermedia, 10(1), 69–84.

Kelloway, E.K. (1998). Using LISREL for structural equation modeling: A researcher's guide. Thousand Oaks, CA: Sage Publications.

Kelley, E. (2005). Context and leadership in the remote environment. Doctoral dissertation. Saint Mary's Univeristy, Halifax, Nova Scotia, Canada. [Publication No. AAT NR03071]. Retrieved October 6, 2009, from http://www.proquest.com/en-US/catalogs/databases/detail/dai.shtml

Laarni N.S., Ravaja J., Kallinen, K., & Saari, T. (2004). Transcendent experience in the use of computer-based media. In Proceedings Third Nordic conference on Human-computer interaction (NordiCHI '04) (pp. 409–412). New York: ACM Press.

Laurel, B. (1993). Computers as Theatre. Reading, MA: Addison-Wesley.

Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of pervieved visual aesthetics of web sites. International Journal of Human-Computer Studies, 60, 269–298.

Lindgaard, G. (2004). Adventurers versus nit-pickers on affective computing. Interacting with Computers, 16, 723–728.

Lindgaard, G., Fernandes, G., Dudek, C., & Brown, J. (2006). Attention web designers: You have 50 milliseconds to make a good first impression! Behaviour & Information Technology, 25, 115–126.

Litman, J.A., & Spielberger, C.D. (2003). Measuring epistemic curiosity and its diverse and specific components. Journal of Personality Assessment, 90(1), 75–86.

Liu, Y. (2003). Developing a scale to measure the interactivity of Web sites. Journal of Advertising Research, 43(2), 207–216.

Luszczynska, A., Diehl, M., Gutierrez-Dona, B., Kuusinen, P., & Schwarzer, R. (2004). Measuring one component of dispositional self-regulation: Attention control in goal pursuit. Personality and Individual Differences, 37, 555–566.

Madden, M. (2007). Online Video. Washington, DC: Pew Internet & American Life Project.

Makkonen, P. (1997). Does collaborative hypertext support better engagement in learning the basics of informatics? SIGCSE Bulletin, 29(3), 130–132.

Mandryk, R.L. (2004, April 24–29). Objectively Evaluating Entertainment Technology. In Proceedings of ACM CHI 2004 Conference on Human Factors in Computing Systems, Extended Abstracts in Computing Systems (CHI '04) (pp. 1057–1058). New York: ACM Press.

Mathwick, C., Malhotra, N., & Rigdon, E. (2001). Experiential value: Conceptualization, measurement and application in the catalog and internet shopping environment. Journal of Retailing, 77, 39–56.

Mathwick, C., & Rigdon, E. (2004). Play, flow, and the online search experience. Journal of Consumer Research, 31(2), 324–332.

Matlin, M.W. (1994). Cognition (3rd ed.). Orlando, FL: Harcourt Brace.

Nahl, D. (2007). The centrality of affect in information behavior. In D. Nahl & D. Bilal (Eds.), Information and emotion: The emergent affective paradigm in information behavior research and theory (pp. 3–38). Medford, NJ: Information Today.

Novak, T.P., Hoffman, D.L., & Yung, Y.F. (2000). Measuring the customer experience in online environments: A structural modeling approach. Marketing Science, 19(1), 22–42.

Nunnally, J.C. (1978). Pyschometric theory (2nd ed). New York: McGraw Hill.

O'Brien, H.L., & Toms, E.G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. Journal of the American Society for Information Science and Technology, 59(6), 938–955.

Overbeeke, K., Djajadiningrat, T., Hummels, C., Wensveen, S., & Frens, J. (2003). Let's make things engaging. In M.A. Blythe, A.F. Monk, K. Overbeeke, & P.C. Wright (Eds.), Funology: From usability to enjoyment (pp. 7–17). Amsterdam: Kluwer Academic Press.

Pace, S. (2004). A grounded theory of the flow experiences of Web users. International Journal of Human-Computer Studies, 60, 327–363.

Park, S.-E., Choi, D., & Kim, J. (2004). Critical factors for the aesthetic fidelity of web pages: Empirical studies with professional web designers and users. Interacting with Computers, 16, 351–376.

Peterson, R.A. (2000). Constructing effective questionnaires. Thousand Oaks, CA: Sage.

Pine B.J. II, & Gilmore, J.H. (1999). The experience economy. Boston, MA: Harvard Business School Press.

Quesenbery, W. (2003). Dimensions of usability. In M. Albers & B. Mazur (Eds.), Content and complexity: Information design in technical communications (pp. 81–102). Mahwah, NJ: Lawrence Erlbaum.

Reio, T.G.J. (1997). Effects of curiosity on socialization-related learning and job performance in adults. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Falls Church, VA.

Rieber, L.P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. Educational Technology Research & Development, 44(2), 45–58.

Said, N.S. (2004). An engaging multimedia design model. In Proceedings of the 2004 Conference in Interaction Design and Children: Building a Community (pp. 169–172). New York: ACM Press.

Schmidt, K.E., Bauerly, M., Liu, Y., & Sridharan, S. (2003). Web page aesthetics and performance: A survey and an experimental study. Paper presented at the International Conference on Industrial Engineering, Las Vegas, NV.

Schraw, G., Flowerday, T., & Reisetter, M.F. (1998). The role of choice in reader engagement. Journal of Educational Psychology, 90(4), 705–714.

Seah, M.-L., & Cairns, P. (2008). From immersion to addiction in video games. In England, D., & Beale, R. (Eds.), Proceedings of Human Computer Interaction Conference, Vol. 1 (HCI 2008) (pp. 55–63). Bedfordshire, UK: British Computer Society. Retrieved October 8, 2009, from http://www.users.cs.york.ac.uk/~pcairns/papers/Seah_Cairns_HCI2008.pdf

Shang, R.-A., Chen, Y.-C., & Shen, L. (2005). Extrinsic versus intrinsic motivations for consumers to shop on-line. Information Management, 42(3), 401–413.

Shenkman, B.N., & Jonsson, F.U. (2000). Aesthetics and preferences of web pages. Behaviour and Information Technology, 19(5), 367–377.

Shernoff, D.J., Csikszentmihalyi, M., Schneider, B., & Shernoff, E.S. (2003). Student engagement in high school classrooms from the perspective of flow theory. School Psychology Quarterly, 18(2), 158–176.

Singleton, R.A., Straits, B.C., & Straits, M.M. (1993). Approaches to social science research. New York: Oxford University Press.

Skelly, T.C., Fries, K., Linnett, B., Nass, C., & Reeves, B. (1994, April 24–28). Seductive interfaces: Satisfying a mass audience. In C. Plaisant (Ed.), Conference Companion on Human Factors in Computing Systems (pp. 359–360). New York: ACM Press.

Smith, C.A., & Ellsworth, P.C. (1985). Patterns of cognitive appraisal in emotion. Journal of Personality and Social Psychology, 48(4), 813–838.

Stone, D., Jarrett, C., Woodroffe, M., & Minocha, S. (2005). User interface design and evaluation. London: Morgan Kaufman Publishers.

Tabachnick, B.G., & Fidell, L.S. (2001). Using multivariate statistics (fourth ed.). Needham Heights, MA: Allyn & Bacon.

Tractinsky, N. (1997). Aesthetics and apparent usability: Empirically assessing cultural and methodological issues. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems (pp. 115–122). New York: ACM Press.

Tractinsky, N., Katz, A.S., & Ikar, D. (2000). What is beautiful is usable. Interacting with Computers, 13, 127–145.

Unger, L.S., & Kernan, J.B. (1983). On the meaning of leisure: An investigation of some determinants of the subjective experience. The Journal of Consumer Research, 9(4), 381–392.

Vick, R.M., & Ikehara, C.S. (2002). Methodological issues of real time data acquisition from multiple sources of physiological data. In Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS-36). Los Alamitos, CA: IEEE. Retrieved September 17, 2009, from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1174295&isnumber=26341

Walker, G.H., Stanton, N.A., & Young, M.S. (2008). Feedback and driver situation awareness (SA): A comparison of SA measures and contexts. Transportation Research Part F: Traffic Psychology and Behaviour, 11(4), 282–299.

Watson, D., Clark, L.A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. Journal of Personality and Social Psychology, 54, 1063–1070.

Webster, J., & Ahuja, J.S. (2006). Enhancing the design of web navigation systems: the influence of user disorientation on engagement and performance. MIS Quarterly, 30(3), 661–678.

Webster, J., & Ho, H. (1997). Audience engagement in multimedia presentations. The DATA BASE for Advances in Information Systems, 28(2), 63–77.

Webster, J., Trevino, L.K., & Ryan, L. (1993). The dimensionality and correlates of flow in human-computer interactions. Computers in Human Behavior, 9, 411–426.

Witmer, B.G., & Singer, M.J. (1998). Measuring presence in virtual environments: A presence questionnaire. Presence, 7(3), 225–240.

Wright, P.C., & McCarthy, J. (2005). The value of the novel in designing for experience. In A. Pirhonen, H. Isomäki, C. Roast, & P. Saariluoma (Eds.), Future interaction design (pp. 9–30). Berlin: Springer.

Zhang, P., & von Dran, G.M. (2000). Satisfiers and dissatisfiers: A two-factor model for website design and evaluation. Journal of the American Society for Information Science, 51(14), 1253–1268.

Appendix 1.   The User Engagement Scale

1. I lost myself in this shopping experience.
2. I was so involved in my shopping task that I lost track of time.
3. I blocked out things around me when I was shopping on this website.
4. When I was shopping, I lost track of the world around me.
5. The time I spent shopping just slipped away.
6. I was absorbed in my shopping task.
7. During this shopping experience I let myself go.
8. I was really drawn into my shopping task.
9. I felt involved in this shopping task.
10. This shopping experience was fun.
11. I continued to shop on this website out of curiosity.
12. The content of the shopping website incited my curiosity.
13. I felt interested in my shopping task.
14. Shopping on this website was worthwhile.
15. I consider my shopping experience a success.
16. This shopping experience did not work out the way I had planned.*
17. My shopping experience was rewarding.
18. I would recommend shopping on this website to my friends and family.
19. This shopping website is attractive.
20. This shopping website was aesthetically appealing.
21. I liked the graphics and images used on this shopping website.
22. This shopping website appealed to my visual senses.
23. The screen layout of this shopping website was visually pleasing.
24. I felt frustrated while visiting this shopping website.*
25. I found this shopping website confusing to use.*
26. I felt annoyed while visiting this shopping website.*
27. I felt discouraged while shopping on this website.*
28. Using this shopping website was mentally taxing.*
29. This shopping experience was demanding.*
30. I felt in control of my shopping experience.
31. I could not do some of the things I needed to do on this shopping website.*

The scale was administered using a five-point scale with "strongly disagree" and "strongly agree" at the respective endpoints. Items identified with an asterisk (*) indicate items that were reverse-coded.

Appendix 2. Select Analysis of Existing Scales Examined During Scale Development Process

Attribute	Source	Context	Purpose	Examples
Aesthetics	Park, Choi & Kim (2004)	Web aesthetics	To explore aesthetics of dimensions with respect to emotion.	Used representative words such as "bright," "popular," "classical," etc. with design experts.
	Pace (2004)	Web aesthetics	To measure users' appraisals of the aesthetic qualities of websites	"Cluttered information displays are off-putting."
	Mathwick, Malhotra & Rigdon (2001)			"I like the way XYZ's Internet site looks."
	Huang (2003)	Web	To examined users' perceptions of a website	Adjectives: "nice vs. awful"; part of an interactivity instrument
Affect	Witmer & Singer, 1998	Sense of Presence in virtual reality environments	To assess people's general psychological state	"How well do you feel today?"
	Positive and Negative Affect Schedule (PANAS) (Watson, Clark & Tellegen, 1988)	Clinical settings		Using adjectives such as "guilty," "nervous," and "alert" to ask people to rate their emotional state for specific time periods (e.g. today, this moment) on a Likert Scale.
	Babin, Dardin, & Griffin (1994)	Physical & online shopping	To measure shopping enjoyment and motivation	"Compared to other things I could have done, the time spent shopping was truly enjoyable."
	Childers (2001)			"Shopping with TAS would make me feel good."
	Davis, Bagozzi, & Warshaw (1992)	Computer systems, including the Web and email	To assess affective experiences with computer systems	"The actual process of using WWW is: (pleasant/unpleasant)."
	Ghani, Supnick, & Rooney (1991)			"Was deeply engrossed in activity/ Not deeply engrossed."
Focused Attention	Mathwick & Rigdon (2004); Mathwick, Malhorta, & Ridgon (2001)	Online shopping	To evaluate the degree of focused attention on a computer task	"I get so involved that I forget everything else."
	Pace (2004)	Web		"When using the Web, I tend to block out conversations with others around me."
	Schmidt, Bauerly, Liu, & Sridharan (2003)	Web	To investigate design elements that draw or distract users' attention	Lists of features and functions of systems: e.g. pop-up windows
	Luszczynska et al. (2004)	Everyday life	To investigated focused attention as a component of self-regulation, defined as the ability to stay focused on a particular task or stimuli	"If I am distracted from an activity, I don't have any problem coming back to the topic quickly."
	Choi & Kim (2004); Ghani, Supnick, & Rooney (1991); Huang (2003); Konradt & Sulz (2001); Novak, Hoffman, & Yung (2000); Pace (2004); Sarkela et al. (2004); Shernoff, Csikszentmihalyi, Schneider, & Shernoff (2003); Smith & Ellsworth (1985); Webster & Ahuja (2006); Webster & Ho (1997); Webster, Trevino, & Ryan (1993); Witmer & Singer (1998)	Web, online shopping, classroom, multimedia and presentation software	To determine how focused/distracted users are when performing computer tasks; the items for focused attention were all part of broader assessments of flow, navigational disorientation, engagement, presence, interactivity, etc.	"I concentrate fully/I do not concentrate fully" (Novak, Hoffman, & Yung, 2000). "This presentation medium holds my attention" (Webster & Ho, 1997).
	Govern & Marsch (2001); Novak, Hoffman, & Yung (2000); Sarkela et al (2004); Witmer & Singer (1998)	Web, Virtual environment, Everyday life	To understand users' perceptions of their external environments while performing an activity	"How aware were you of events occuring in the real world around you?" (Witmer & Singer, 1998).

(Continued)

Attribute	Source	Context	Purpose	Examples
Challenge	Mandryk (2004)	Affective response to video games	To assess challenge as an aspect of gamers' assessment of fun	"This condition was challenging."
	Webster & Ho (1997)	Engagement with presentation software	To evaluate the level of challenge associated with	"This presentation medium encourages me to think."
	Unger & Kernan (1983)	Individual's approach to leisure	To evaluate challenge as an aspect of personality	"I feel I have been thoroughly tested."
	Smith & Ellsworth (1985)	Cognitive appraisal in everyday life	To evaluate different dimensions and levels of feedback: effort, exertion, prediction, uncertainty, and understanding.	"When you were feeling happy, how much effort (mental or physical) did you feel this situation required you to expend?"
Control	Smith & Ellsworth (1985)	Cognitive appraisal in everyday life	To address the "situational," "self," and "other" elements of perceived control	"To what extent did you feel that you had the ability to influence what was happening in this situation?"
	Choi & Kim (2004); Liu (2003)	Web interactivity	To assess the role of perceived control in users' assessment of website interactivity	"While surfing the website, I had absolutely no control over what I can do on the site (reverse scaled)" (Liu, 2003).
Feedback	Choi & Kim (2004); Novak, Hoffman, & Yung (2000); Schmidt, Bauerly, Liu, & Sridharan (2003); Zhang & von Drann (2000); Liu, (2003); Chin et al.,	Video games; websites	To examine feedback in the context of other construct, including flow, interactivity, aesthetics, and website design and evaluation	"When I use the Web there is very little waiting time between my action and the computer's response" (Novak et al., 2000).
Motivation	Babin, et al. (1994); Mathwick, et al. (2001)	Physical & online shopping	To measure shopping motivations	"I shop from the XYZ's for the pure enjoyment of it" (Mathwick et al., 2001).
	Unger & Kernan (1983); Webster, Trevino, & Ryan (1993); Shernoff et al. (2003); Webster & Ho (1997); Webster & Ahuja (2006); Guay, Vallerand, & Blanchard (2000)	Web, classrooms, general leisure, presentation software	To assess motivation; with the exception of Guay et al., (2000), these items pertain to intrinsic motivation (and not extrinsic motivation).	"There may be good reasons to do this activity, but personally I don't see any" (Guay et al., 2000).
Novelty	Litman & Speilberger (2003); Reio (1997)	Personality make-up	To measures general curiosity as a component of personality	"New ideas excite my imagination" (litman & Speilberger, 2003).
	Webster & Ho (1997); Webster, et al. (1993); Webster & Ahuja (2006)	Web, presentation software	To assess novelty as a component of engagement, web navigation, flow, etc.	"Using X excited my curiosity" (Webster & Ho, 1997).
	Mathwick & Rigdon (2004); Novak, Hoffman, & Yung (2000); Huang (2003); Zhang & von Drann (2000)	Online shopping; Web		"Presence/absence of novel (new) information" (Zhang & von Drann, 2000).
	Unger & Kernan (1983)	Leisure	To assess novelty as a element of leisure activities	"It offers novel experiences."
Perceived time	Pace (2004); Witmer & Singer (1998); Novak et al. (2000)	Web, Video Games	To assess the users' perceptions of time spent using an application	"Time seems to go by very quickly when I use the Web" (Novak, Hoffman, & Yung, 2000); "Do you ever become so involved in doing something that you lose all track of time?" (Witmer & Singer, 1998).
Intended inquiry; loyalty; intended purchase, intention to return	Gefen & Straub, 2001; Choi & Kim, 2001; Webster & Ahuja, 2006; Davis, Bagozzi, & Warsaw, 1989; Adams, Nelson & Todd, 1992	Academic software, video game, Web, online shopping	To garner users' perceptions of how effective the application was to use, and whether they would re-engage with it at a future time	"I would intend to use this web site" (Webster & Ahuja, 2006).