New Directions in Third Wave Human-Computer Interaction, Volume 2: Methodologies

A Holistic Approach to Measuring User Engagement

Heather L. O'Brien, Associate Professor
iSchool, University of British Columbia, Vancouver, Canada

Abstract

This chapter provides an overview of methodological approaches and current work in the evaluation of user engagement (UE). Using a series of propositions about the nature of engagement, I review a selection of recent research that utilizes varied methodological approaches to study UE in various human-computer interaction settings. The propositions and the reviewed literature are used to propose a methodological framework to guide decision making and reflection regarding how UE will be evaluated in a given context. The chapter concludes with reflections on broader issues related to how researchers' methodological stances influence the evaluation of UE. Overall, the chapter argues that UE should be measured using a thoughtful mix of qualitative and quantitative methods, considering the particulars of the use context, and balancing established and emerging subjective and objective metrics.

Introduction

Today's digital technologies, including web search engines, e-book readers, mobile apps, library databases, social networking sites (SNS), and Massive Open Online Courses (MOOCs), log vast quantities of data. Users leave behavioral traces whenever they download, click, scroll, like, query, etc. where these actions are often equated with *user engagement*. While indicators of user activity and "stickiness" (i.e. repeated use), these behaviors may not reveal the whole story about how people engage with technology in the moment or over time. Over the past decade, I have focused on user engagement (UE) as a *quality* of user experience, arguing for the need to understand the cognitive, emotional, behavioral, and, increasingly, social, dimensions of people's digital interactions. Beyond what people *do* when they use these applications, *why* do they shop, learn, search, connect or play in the first place? What affordances of technologies foster, deter, or sustain meaningful engagement? Furthermore, what is the benefit of UE: Does it help people think critically, feel connected to others, learn, relieve stress, or manage their health?

A holistic view of user engagement acknowledges the myriad motivational, individual, technological, and contextual factors that shape experience, and recognizes that meaningful *outcomes*, such as learning how to program or changing an unhealthy behavior, are more nuanced than *outputs*, such as pages viewed during a web search or products purchased while shopping online. A truly holistic view, however, is not always pragmatic for designers, product developers or content strategists who operate in competitive and fast paced environments, or those with limited resources to evaluate user experience. Sets of metrics (e.g., daily active users, number of comments or likes) that can be tracked over time and that generate interpretable outputs (e.g., retention rates, amount of downloaded content) may be preferable and more feasible than trying to account for the many factors that precipitate engagement, or to map less tangible effects

O'Brien, H. L. (2018). A holistic approach to measuring user engagement. In M. Filimowicz & V. Tzankova (Eds.), *New Directions in Third Wave Human-Computer Interaction: Volume 2 - Methodologies* (pp. 81-102). Springer, Cham.

of technology interactions, such as wellbeing and learning gains.

Yet, design is not prescriptive and neither is measurement. It is problematic to apply a single set of engagement metrics to *all* digital experiences, given that applications vary in terms of targeted user group, purpose, media type, and content. Consider online communities, for example, and how engagement might differ in citizen science versus micro-worker crowdsourcing communities; users' motivations for participating, the nature of the activities carried out, and the degree of subject knowledge required for completing crowdsourcing tasks mean that we cannot conclude that engagement is the same even within the same digital domain. This is not to say that there is no room for quantifiable, easy-to-apply metrics in the evaluation of UE, but we need to ensure they are effective proxies within the contexts in which we are applying them.

Thus, in the study of user engagement, we need not abandon the methods of "second wave" HCI completely in favour of the third wave. This is not to say we should continue to view HCI interactions as purely cognitive (e.g., humans as "information processors) rather than as situated in particular contexts (Harrison et al., 2007). Rather than dividing methods as second or third wave, it may be more productive to examine the overlap between the modes of inquiry used in each paradigm (Bödker, 2015). Taking a more pragmatist stance, we might first ask what it is we want to know about a phenomenon, and the constraints operating within the context that will shape our inquiry and design processes. We can then employ a variety of relevant methods (objective, subjective) and explore the interface between the findings to achieve "rich, detailed descriptions of specific situations" (Harrison et al., 2007, p. 11) that characterized third wave ways of knowing.

This chapter seeks to provide an overview of methodological approaches and current work in the evaluation of user engagement, by exploring the definition and measurement of UE. Using a series of propositions about the nature of engagement, I will review a selection of recent research that utilizes varied approaches to study engagement. The chapter will conclude with reflections on broader issues related to how researchers' methodological stances influence the evaluation of UE.

An Overview of User Engagement Methodological Approaches

A number of methods have been employed to investigate user engagement (UE) across a range of digital applications[1]:

- Self-reporting via questionnaires, think aloud/think after protocols, interviews;
- Log analysis, or usage patterns derived from behavioral observations, such as number of mouse clicks, scrolling behavior, number of unique or returning users, time spent using an application, etc.;
- Neurophysiology, which uses measures such as heart rate, electrodermal activity (EDA), electroencephalography (EEG), electromyography (EMG), functional magnetic resonance imaging (fMRI), and eye tracking to infer, for example, users' attention, cognitive load and level of arousal based on electrical activity, blood flow, pupil dilation changes, etc. in the brain and body;
- Ecological momentary assessments (EMAs), which prompt people to record

[1] For more in-depth discussion on different UE evaluation approaches, please see Lalmas, O'Brien & Yom-Tov (2014) and Yardley et al. (2016).

their current behaviors or experiences in the moment; and

- Mobile or environmental sensors that record geographical location or behaviors (e.g., step counts, route information) as people move through time and space.

Each of these methods has its benefits and drawbacks. For example, as Yardley et al. (2016) point out, EMAs are interruptive; they purposefully disengage people from their activities and may negatively impact engagement; O'Brien and Lebow (2013) suggest that analytic metrics effectively capture large-scale user behavior patterns but do not account for people's motivations, goals, or emotional responses to their activities.

The purpose of this chapter is not to advocate for/against a particular method, but to make a case that the study of UE must be approached with a well-equipped toolbox, the expertise to utilize its contents effectively, the ability to consider the impact of contextual nuances on evaluation, and ultimately, an openness to the different ways that meaningful engagement is constructed through digital technologies. All methodological approaches have their place and time, and all contexts come with a set of constraints to be negotiated, such as access to users or their data, time and resources to collect and make sense of data, etc. There are also differences in the nature of individual measures in terms of whether they have been substantiated in a particular setting and generalized to others, or are emerging and therefore more exploratory and limited in their validity.

Furthermore, the abstract nature of UE makes the act of measuring it problematic. In many instances, we are *inferring* engagement rather than truly measuring the phenomenon itself. For example, let us consider e-commerce settings where potential shoppers spend varying amounts of time, have different goals ("I know what I am looking for" vs. "I want to browse and see if something catches my eye"), and may or may not make a purchase. How do we understand shoppers' levels of engagement? We can't be inside other people's minds (and sometimes people have great difficulty explaining their own motivations and actions), but patterns of interactions (e.g., dwell time on product information pages, adding products to the shopping cart, purchasing behavior, activities of new versus returning customers) can be used as *proxies* of engagement. Yet different e-businesses may use the same metrics to tell different stories. One business might view returning customers as a positive sign of loyalty or brand reputation, while another business might want to look a little more deeply at the returning customers in terms of their purchasing behavior, or the time between return visits: some returning customers may make periodic visits over a long period of time, while others may make frequent visits within a small window of time. The former return visitors may be the loyal customers, whereas the latter might be making a decision about a single purchase. It is for this reason that we need to look beyond the metrics, and why our definition of UE is fundamental to our methodological choices.

Defining User Engagement

When I began research in the area of UE in 2005, I drew upon the foundational work of Richard Jacques (Jacques 1996; Jacques, Preece & Carey 1995), Jane Webster (Webster & Ho 1997), and Brenda Laurel (1993) among others, published in the mid- to late-1990s. I attempted to unite scholarship across different fields of inquiry (human-computer interaction (HCI), information systems, learning sciences, etc.) to define user engagement (UE) and to distinguish it from related concepts, such as flow, immersion and presence (O'Brien 2008; O'Brien & Toms 2008). Early on, I envisaged a single definition of user engagement that could be applied consistently to aid in evaluation and facilitate communication amongst multidisciplinary stakeholders. I followed in the path of other researchers who focused on identifying attributes of and influences on UE (Jacques 1996; Webster & Ho 1997). Based on a systematic review of the literature and an exploratory interview study with online gamers, searchers, shoppers, and learners, I proposed that:

> Engagement is a quality of user experiences with technology that is characterized by challenge, aesthetic and sensory appeal, feedback, novelty, interactivity, perceived control and time, awareness, motivation, interest, and affect (O'Brien & Toms 2008, p. 949).

The benefit of this definition (and others like it) is that it operationalizes dimensions of user engagement that can be isolated and measured. For example, psychological tests, such as the Stroop task (Stroop, 1935), can reliably measure selective attention, which has been demonstrated to be a component of engagement. In my own work, I operationalized engagement through the User Engagement Scale (UES). I compiled a bank of questions derived from the literature and my initial interview study that corresponded to each of the attributes in my original definition. I then tested a portion of these items (that were shown to have face validity) in two online surveys with hundreds of online shoppers to further reduce the number of questions to a parsimonious and reliable set, and to examine the factor structure, or how items grouped together. The latter analysis demonstrated distinct dimensions of UE as measured by the UES: focused attention, involvement, novelty, usability, aesthetic appeal and endurability, or overall feelings of success and willingness to engage in the future (O'Brien 2008; O'Brien & Toms 2010). Thus the attribute-based definition of engagement led to the construction of questions that together captured user engagement as a multidimensional construct through self-reporting.

Attribute-based definitions, however, can be problematic. We might ask how stable these attributes are over time and across different user groups and types of technologies (O'Brien 2016a), which leads to the question of the universality of any definition. Further, it may not be so much about the presence or absence of attributes, such as motivation, novelty, challenge, etc. but rather their intensity. I put forward a stage-based Process Model of User Engagement where I envisaged that technology users move through a point of engagement, a period of sustained engagement, disengagement and (potentially) re-engagement, and that some attributes may be more salient at particular stages of the process (O'Brien 2008; O'Brien & Toms 2008). I have also noted that different attributes are more compelling than others for people placed in the same situation. In a study I conducted of online news browsing, some participants' engagement was embodied in the physical interactions they had with the news site, while others were cognitive or affectively engaged with the news content (O'Brien 2011). This definition is less rigid than the previous one, yet more conceptual in nature and therefore more challenging to operationalize. However, it creates space to select, adapt and experiment with methodological approaches that reflect a broad perspective of UE and the anticipated outcomes of engagement in a given scenario.

An Interpretive Framework for Studying User Engagement

There are multiple perspectives on user engagement, and these are guided by our epistemological practices as researchers and designers, as well as the outcomes we are trying to facilitate for people through technologies. Despite different approaches to the study of UE, there are commonalities. Recently I documented “unifying propositions” within the literature, with the idea that these could be used to develop a theory of user engagement (O’Brien 2016a). This was not intended to be “Theory with a capital T,” which is constructed over many years and can seem overly abstract for day-to-day application. Rather, the propositions were intended to provide a flexible, interpretive framework for considering the scope and meaning of engagement in a given setting, and, consequently, its evaluation. The propositions are summarized as follows:

- User engagement is a process *and* product of digital interactions.
- User engagement has *affective*, *behavioral*, and *cognitive* aspects.
- User engagement is a quality of UX that is characterized by the *depth* of the actor’s temporal, emotional and cognitive investment in the interaction; users’ level of engagement may range from shallow to deep; this depth continuum occurs within individuals and communities. Shallow participation may be adaptive to human wellbeing and contribute positively to “the digital ecosystem” made up of “touchpoints” that “are ‘woven together’ by social practice” (Bagnara & Pozzi, 2016, p. 64). Thus, engagement is not an all-or-nothing phenomenon characterized by constant, high arousal or concentration, or lengthy interactions; rather, the intensity of experienced engagement can ebb and flow depending on the user or community’s need for (inter)action.
- In addition to the nature of the interaction itself, *external factors*, such as situational constraints, and users’ goals, motivations, and personal qualities (e.g., computer self-efficacy, topic interest or expertise). Therefore, UE is *context dependent*, and context may be discerned at different levels (personal, social, task or situation related) (O’Brien, 2016a, p. 22).

These propositions represent a synthesis of the conceptual understandings of engagement over the past two decades, yet they can also be used to consider the expression, as well as the merits and drawbacks of different methodological approaches. By focusing on theoretical propositions rather than types of measures, my intention is to place user engagement in a more intersectional space where methods are not rigid but adaptable to the kinds of complex research questions we need to pose in HCI. In the following sections, I will use recent examples to illustrate each of these propositions; these examples are not exhaustive, but are intended to show contemporary work in the area. It is important to note that no single study encapsulates all aspects of any one proposition; for instance, not all researchers view UE as behavioral, cognitive and affective and therefore measures related to all of these facets are not included in their work. However, we can examine these studies with respect to the conceptualization of UE and approach to measurement, and the role of contextual constraints and outcomes of interest on the research design.

User Engagement as Process and Product of Interaction

User engagement is a process and product of people’s interactions with digital environments. In other words, UE can be analyzed *during* and *after* human-computer interactions. It has been more common to measure engagement as an outcome of interaction. Self-report measures administered after a user has completed a computer-mediated task are common practice. It has been more challenging to evaluate engagement as a process. While neurophysiological equipment has been becoming more accessible to HCI researchers, the dilemma of syncing and making sense of the large amounts of data generated (Lebow & O'Brien 2012) still remains. Further, commercially available neurophysiological equipment, while less invasive and more affordable than before, may affect data quality. Andujar et al. (2017), for example, critiqued wearable Brain-Computer Interface (BCI) devices:

> [The BCI headset] requires the user to stand still, restrict muscle movement as much as they can, have shorter or no hair, and be at a room at a decent cold temperature to avoid sweat. These requirements do not allow the user to perform a real-world task in a natural manner and may affect how they feel about BCIs negatively. An ergonomic issue is that these BCIs are not adjustable for all the different types of head shapes and sizes (p. 104).

However, recent work also highlights strides made to collect, analyze, and interpret process-based data using neurophysiological methods, increasing their potential to be used outside of controlled laboratory settings. The following studies show the complexities of collecting and interpreting process data, and also the role of the human body in the measurement of UE, where engagement is the result of "doing things in world" (Bagnara & Pozzi 2016; Harrison et al., 2007, p. 7)

Anzalone et al. (2015) performed two studies to explore robots' ability to foster social intelligence in young adults and children. In the first study with young adults, the researchers were interested in the robots' ability to elicit non-verbal communication behaviors through a task where participants attempted to teach a robot the colours of objects. In study two, the aim was to induce joint attention between children and the robot on an object; the purpose was to compare children with autism spectrum disorder and those experiencing typical development. The authors used sensors to detect and log participants' gestures, eye gaze, and body and head positions in the interaction space, and inferred UE with the robots based on nonverbal behaviors. Anzalone et al. discussed the potential of their findings to monitor humans' non-verbal behaviors for the purposes of increasing attention focus and responsiveness in human-robot interactions to "strengthen the engagement, regulate the rhythm of interaction, and arouse in people the perception of social intelligence" (p. 474).

Li et al. (2016) were also interested in engagement and the "rhythms" of interaction between humans and robotic systems. They developed and tested a cyber-physical stroke rehabilitation system (CP-SRS) to enhance stroke survivors' motor abilities when performing rehabilitation exercises. The CP-SRS combined an assistive robotic system and gamification, and monitored engagement using data collected from on-screen content changes, eye movements, facial expressions, electromyography (EMG), and electroencephalogram (EEG). Some of the measures were more successful than others in capturing engagement. For example, although the exercises were created to require different levels of cognitive load, this did not result in differences in cognitive engagement. In addition, there was a lack of correspondence between the facial recognition data and self-reports used to measure emotional engagement; they did, however, find that including an accuracy requirement in the exercises increased the amount of attention and effort needed to complete the tasks.

Li et al. (2016), Anzalone et al. (2015) and other researchers, are confronting the difficult measurement challenge of capturing and making sense of data gathered through numerous channels to identify (non) engaged states. As in Li et al.'s (2016) study, part of this sense-making may involve combining process and product measures, and looking for correspondence between these. Another recent example is the work of Jensen et al. (2016). They used a strategy/simulation game and a training video to teach students about the impact of cognitive biases on decision-making, comparing engagement and responses to failure by mode of interaction. The performed a study at two different sites with over 150 students using eye tracking, self-report and physiological methods; heart rate and EDA (specifically skin conductance levels) were measured only at one site. The self-report measures included established questionnaires for cognitive absorption (CA) (Agarwal & Karahanna 2000), positive and negative affect (PANAS) (Watson, Clark & Tellegen 1988), and personality (ten-item personality inventory or TIPI) (Gosling, Rentfrow, & Swann, 2003). Results corroborating the various measures were mixed. Although there were no physiological differences between the game and training video conditions, the researchers observed that skin conductance levels decreased for those in the video condition and remained stable for those playing the game compared to participants' own baselines. Only one dimension of cognitive absorption (temporal disassociation) differed between the video and game conditions, but PANAS scores showed that positive affect declined from pre- to post-task for the game group, which would be expected since these participants experienced failing to win the game while those in the video condition watched failure occur but did not experience it personally.

Martey et al. (2014) also explored UE in a gaming environment using multiple measures: self-reported presence, EDA, and mouse clicks, mouse rests, and attention (time spent looking at the screen based on screen captures). The same two-dimensional (2D) puzzle-based gaming environment was used in two studies but with different manipulations: in study one, people were/were not able to customize an avatar, and in study two the art (simple versus detailed) and narrative (rich versus light detail) varied. Across both studies, they found consistency in and correspondence across the various self-report measures. However, the relationships between physiological, behavioral and self-report data were not as significant. This is similar to the findings of O'Brien and Lebow (2013), who found strong correlations between self-reported engagement, cognitive absorption and usability, but not between self-reported UE and search behaviors (browsing time, reading time, links and pages visited), or heart rate, EDA and EMG.

User Engagement as Affective, Behavioral and Cognitive

My perspective is that user engagement has affective, behavioral, and cognitive aspects. However, examining the multifaceted nature of UE effectively in a single research study may not be feasible or desirable. In these cases it is prudent to be clear about how UE is being defined and why in a given scenario.

Nguyen (2015) chose to examine only behavioral engagement in the context of crowd-based open collaborations, justifying this decision by reasoning that "to the crowdsourcers, online users' engagement is significant only when they actually contribute something to the crowdsourcing events" (p. 4). In doing so, Nguyen articulated that engagement is voluntary, "on-task," and observable in the context of online collaborative initiatives. In addition, he speaks of engagement as process-based and associates it with active, effortful participation but not the quality of the contributions: "[engagement] represents an effort, not a work outcome. That is, it does not matter whether users' contributions are outstanding or mediocre. The emphasis is on the fact that they actually make an attempt to do something" (p. 28). Nguyen's framing of user engagement allowed him to operationalize it as the "intensity, sustainability, and variety of tangible effort online users voluntarily devote to what is requested in an open collaboration initiative" (p. 4) and to develop measures accordingly. *Intensity* referred to the effort required to make various contributions, where contributions were weighted according to effort required to complete specific tasks; *sustainability* examined participants' contribution patterns over the duration of the initiative, and *variety* accounted for the diversity of participants' activities.

While Nguyen (2015) focused solely on behavioral engagement, Andujar et al. (2017) sought to measure affective engagement in their study with high school students. They used a wearable brain computer interface (BCI) to collect EEG data as students performed a simple task with one of two different programming environments over a ten-minute period of time, with changes in engagement averaged every two minutes. The engagement patterns captured with the EEG data were not statistically different between the two programming environments. However, they noted differences in terms of whether engagement was demonstrably different or similar amongst participants using the same programming tool.

Andujar et al.'s (2017) emphasis on affect and changes within individual learners relates to related work by O'Brien, Freund and Kopak (2016) where students' comprehension of materials in a digital reading environment was evaluated. Students' engagement was categorized as high, medium, or low based on their scores related to UES questions about involvement, novelty and endurability. The researchers tested comprehension using true and false items to assess recall and the sentence verification technique (SVT) to explore participants' understanding of main ideas presented in the texts. The low and high engagement groups performed significantly better on the tests than the medium group; their average scores were similar for the true and false questions, but the least engaged participants had the highest mean SVT scores. In other words, participants who experienced low and high engagement achieved the same learning outcomes, but the high engagement group had a more positive experience, and the researchers concluded that, while not necessary for learning to occur, the ability to engage with the readings helped this group do "the hard work of learning" (p. 73). Thus in learning and other applications – where it is not the destination but the journey – it may be crucial to attend to the affective qualities of the interaction, such as motivation and interest, and affective measures may help the researcher make sense of non-significant behavioral or performance differences between experimental conditions or study outcomes.

Lohse, Boyd and Hodges (2016) looked specifically at the role of engagement in enhancing motor learning in a game-based environment, but were also interested in affective aspects of the experience. Forty university students were assigned to a rich game condition, where they practiced a novel motor skill, or a sterile condition, where they engaged in the same game stripped of its aesthetic features. Findings were compared across conditions and participants who engaged in low (200 trials in one day) or high

(400 trials over two days) doses of practice; retention was tested one week after the practice session. Modified versions of the User Engagement Scale (O'Brien & Toms 2010) and the Intrinsic Motivation Inventory (McAuley, Duncan, & Tammen, 1989) were used to measure engagement and intrinsic motivation, respectively, and behavioral metrics (e.g., catching an object, trial time) assessed game performance. The researchers concluded that the more engaging game environment improved motor learning. Although the game and sterile groups performed similarly during the practice phase of the study, the game group's motor skill retention was higher when they played the games a week later. This was not due to differences in participants' intrinsic motivation, but in overall engagement with the more aesthetic version. This study's emphasis on retention and the time lapse between gaming sessions also calls attention to *long-term engagement*.

In another study, Leiker et al. (2016) used the same game conditions to test motor skill retention and transfer, but added cognitive components to the study design. Audio probes were presented randomly during game play to tax participants' attention and EEG was used to monitor brain activity during the first session only. The researchers failed to detect learning or engagement effects across the sterile and game groups, with both groups achieving similar performance outcomes in the second session. However, they did observe an association between EEG and self-reported engagement, where engagement was related to how much attention participants were required to expend in both game environments, i.e., the auditory probes increased cognitive complexity. The authors suggested that this was strong evidence that engagement is related to cognitive changes and is more than just an affective experience.

Some research has attempted to examine multiple aspects of UE in the same study. Returning to Li et al.'s (2016) work to design and test the CP-SRS, different types of engagement were defined, and distinct measures were developed to reflect these types. Since the system was designed to assist people to complete physical exercises, motor engagement was important, and operationalized as "active and effortful motion" (p. 3). However, they were also interested in perceptive engagement or "sensory concentration," as measured by eye gaze and cursor and content change positions, since patients were supposed to focus on and interact with a video game to complete the exercises. The researchers also measured cognitive engagement and emotional engagement using neurophysiological methods –EEG and facial expressions – to derive patients' degree of concentration and positive/negative emotions during their interactions. They then performed a series of experiments to test and validate their chosen measures for each of four types of engagement. What is noteworthy here is the operationalization of engagement in this particular study as it related to the broader goal of developing an effective rehabilitative system.

User Engagement as Depth of User Investment

User engagement is not an all-or-nothing phenomenon: people can experience different degrees of engagement as they engage, disengage and re-engage with technology. In fact, the ebb and flow of engagement is essential, as fatigue would be eminent with constant high engagement or could lead to addictive behaviors. In addition to being personally beneficial to disengage periodically, it is not necessarily harmful to have different levels of engagement operating within user communities.

In their study of two, two-year citizen science crowdsourcing projects, Ponciano and Brasileiro (2014) found five distinct engagement profiles of the sites' thousands of volunteers based on their performance of over one billion tasks; volunteers were defined as those who made an "ongoing contribution" to the projects, contributing on more than one day (p. 253). The profiles – hardworking, spasmodic, persistent, lasting, and moderate – were created by clustering participants' data according to number of activities performed, duration of involvement in the project, and level of activity within the contribution period. For example, "spasmodic" volunteers contributed for a brief period of time, but their irregular contributions are punctuated by bursts of intense activity, whereas "moderate" volunteers were steady, achieving immediate scores on all engagement metrics.

Ponciano and Brasileiro used the Process Model of User Engagement (O'Brien 2008; O'Brien & Toms 2008) as a framework for developing their metrics: the amount of time the volunteer could potentially be part of the project (based on the projects' duration), days the volunteer remained linked in the project, number of active days, time spent contributing on an active day, and days elapsed between two active days. These behaviors allowed the researchers to trace when volunteers became engaged, moved through periods of sustained or active participation, disengaged and re-engaged. By analysing and clustering volunteers' previous behaviors, the authors reasoned that different engagement strategies could be employed to appropriately trigger (increase, decrease, or sustain) activity levels. Their study demonstrates that engagement can look different for different people, yet still result in meaningful contributions: brief but highly active engagement ("hardworkers") can be as beneficial to project outcomes as longer, less active participation ("persisters").

Viewing engagement as a continuum where people operate at different levels of intensity also means that some attributes of UE will mean different things to different people at different times. In my original interview study with technology users I drew upon McCarthy and Wrights (2005) "Threads of Experience" to develop and explain the Process Model of User Engagement (O'Brien 2008; O'Brien & Toms 2008). I plotted the engagement process along McCarthy and Wright's compositional thread and highlighted attributes salient during the point of engagement, period of sustained engagement, disengagement and re-engagement stages based on their connections to the sensual, emotional, and spatiotemporal threads. For example, the aesthetic appeal of the computer-mediated environment and the novelty of the content presented were sensual aspects at the point of engagement, while both negative (uncertainty, boredom, frustration, guilt) and positive (success, accomplishment) feelings were emotional components of disengagement (O'Brien & Toms 2008, p. 948). Fluctuations in intensity may come about due to changes in the technology user, e.g., changing needs or motivations as one uses a digital application over time, or they may be a consequence of interactions between the individual and the technology. Unfortunately, it is difficult to discern when and why engagement changes, and this is further complicated by the rich contexts in which digital interactions unfold.

The Context of User Engagement

Context is an important variable in user engagement, yet extremely challenging to contend with. We can consider context on many levels to the point that we can ask, "what is *not* context" (Absar, O'Brien & Webster 2014)? Returning to Bagnara and Pozzi's

(2016) not of digital eco-systems, we can look at contextual variables as ways in which to better understand UE in a given scenario, and to acknowledge cross-contextual differences in engagement. However, the messiness of context can also threaten our ability to evaluate UE.

Cross-contextual differences are understood through recognition that different values and interests guide behavioral interactions with technologies. Based on the notion that online communities form to bring people with shared values and interests together, Zhang et al. (2017) developed a typology of community identity based on the language used in online community spaces. Their typology categorized community's interests as distinct versus generic, and dynamic versus stable over time, and was used to examine four unique Reddit communities. Word level measures within each community were explored. For example, the word "kitchen" was specific and frequently used in the *Cooking* community, while "Easter" was highly volatile (i.e., used irregularly) in the *BabyBumps* community. These word level measures were investigated with respect to typical community engagement measures, including community size and activity level, as well as retention, acculturation (the use of community-specific language by frequent versus rare commenters) and content affinity (volatility of language used by active and non-active users). Zhang et al.'s linguistic approach provides a more nuanced view of online community members' interests than examining behavioral data alone, or failing to account for unique characteristics of particular communities that pertain to their endurability for specific members. For instance, *BabyBumps* may be useful for expectant mothers but less applicable once their babies are born, whereas participating in *Cooking* might reflect an ongoing and long-term hobby.

While Zhang et al. (2017) captured contextual differences in online communities linguistically, Aristeidou (2016) applied a design-based research (DBR) methodology to investigate two online citizen science communities: Weatherit and Inquiring Rock Hunters (IRH). Specifically, the author used online focus groups, interviews, log files, questionnaires, researcher notes, and participants' reflections on the researcher's findings to understand Weatherit and IRH's members' engagement, motivation, and learning, and how these are influenced by the design of these online communities. Aristeidou (2016) found different aspects influencing participation and motivation in the two communities. The type of software used and propensity for inquiry within both projects induced engagement and disengagement for members. However, mentoring and collaboration were key in IRH, whereas Weatherit's success was dependent upon acts of community creation and sustainability (e.g., active community e-moderators). The distinction between Weatherit and IRH is important as it reinforces that the same metrics may not be useful for the investigation of all citizen science projects due to the influence of contextual factors. In this case, users' motivations and levels of expertise for participating, the ability of the projects' software to support communication, mentoring, and collaboration activities as needed by members, and even members' perceptions of the purpose of the community (science versus hobby) all distinguished what made for an engaging citizen science community.

Mathur, Lane and Kawsar (2016) performed an interesting study of smartphone users that moved results from the lab to the field, and showed how these distinct contexts can be used to inform each other. In the first phase, they asked ten people to perform fourteen tasks using their smartphones while wearing an EEG headset; the tasks were intended to be of different types and durations to produce variable EEG ratings. They also gathered self-reports using select items from the UES pertaining to focused attention and involvement (O'Brien & Toms 2010). By examining the strength of self-reported engagement in relation to the EEG ratings, the authors built a proxy inference model for EEG Benchmark Engagement Scores (BES). Subsequently, an additional ten people were recruited to wear an EEG headset and have their personal smartphone use logged over a twenty-four hour period (wearing the headset as much as possible). The researchers examined the EEG and log data in concert and generated algorithmic models based on the applications used. The outcome of phase one was a Random Forests classification model that identified "high" versus "low" engagement based on application usage features. In the second phase of the study, 130 smartphone users from twelve countries were followed over a four-month period. Mobile use was logged according to events (e.g., call, screen and application data) and sensory interactions when the screen was turned on and at fifteen-minute intervals; the idea here was to examine usage behaviors over time and in ecologically valid settings. Based on data obtained from over 250,000 usage sessions, the researchers extracted five types of context features to conduct predictive modeling of high and low engagement. Mathur et al. (2016) demonstrated how the controlled environment of the lab can be used to better understand mobile users "in the wild," and how the use of neurophysiological signals can be used to make sense of everyday life mobile use patterns.

However, the intricacies of context in digital interactions also leads to difficulties in translating findings from one study or setting to another, and may mask significant and non-significant results. In Martey et al.'s (2014) study of the 2D puzzle game discussed earlier, the authors considered explanations for why their results were not as anticipated. They speculated whether customization was operationalized effectively in study one (or whether customization was even an important element of engagement in educational games) and whether the incorporation of learning content was incongruent with the goals of the game in study two. Interestingly, the EDA data they collected indicated that arousal levels were highest in the most complex (rich narrative/detailed art) and least complex (light narrative, simple art) conditions. This parallels the findings of Freund, Kopak and O'Brien (2016) in their study of engagement and comprehension in digital reading environments. Comprehension gains were highest for the group who interacted with texts in the simplest reading environment –black font on a white background; the group assigned to the most complex environment, which contained paratextual cues ("in context") and interactive Reading Tools, had the second highest gains. The results of Martey et al. and Freund et al. speak to how computer-mediated environments with minimalist design may be better for attention focus; when minimalism is not possible or desired, creating an immersive environment through aesthetic, interactive and narrative elements may direct attention to key components.

Digital interactions – even when observed in the lab with greater control – are "messy" (Harrison et al. 2007, p. 12), and researchers may need to rely on prior research and other data sources to make sense of the findings. Martey et al. (2014) were particularly perplexed about the non-significant differences between the rich and light narrative groups, given the prominence placed on narrative engagement in the literature, leading them to question how they designed their narrative conditions. Freund et al. (2016) examined behavioral measures collected during the study using screen capture software, specifically total reading time and time spent on different components of the texts, e.g., introduction and conclusion. They reasoned that the interactive Reading Tools affected the comprehension scores of those in the plain and in context conditions differently because these two groups formed two distinct mental models: “one based on print reading, in which Reading Tools are used to facilitate in-depth reading and engagement with texts [i.e., plain], and the other based on web reading, in which Reading Tools are used primarily to facilitate navigation and marking trails through content [i.e., in context]” (p. 90).

This messiness is extended in field-based research. Flores (2016) used an iterative, design-based approach to examine the health of emerging adults (EAs) as they transitioned to university. She conducted three pilot studies using interviews, online diaries, questionnaires, virtual scenarios and participatory design exercises to design her main study. The main study explored engagement with one of the three fidelities: virtual, 2D and paper prototypes, and was conducted in stages with two in-person design sessions separated by a two-week online collaboration period. The author collected observations based on the in-person and online interactions, and this presented opportunities and challenges. The rate of attrition over the main study made conclusions involving questionnaire data difficult to reach. Qualitative exploration of participants’ utterances during the design and collaborative processes showed evidence of engagement but we cannot know why the participants who dropped out of the study disengaged, or what aspects of the experience might have been different for them.

A unifying framework for evaluating user engagement

By focusing on theoretical propositions of user engagement and the range of approaches taken by researchers in the aforementioned studies, we arrive at a methodological framework to guide the evaluation of user engagement (Figure 1).

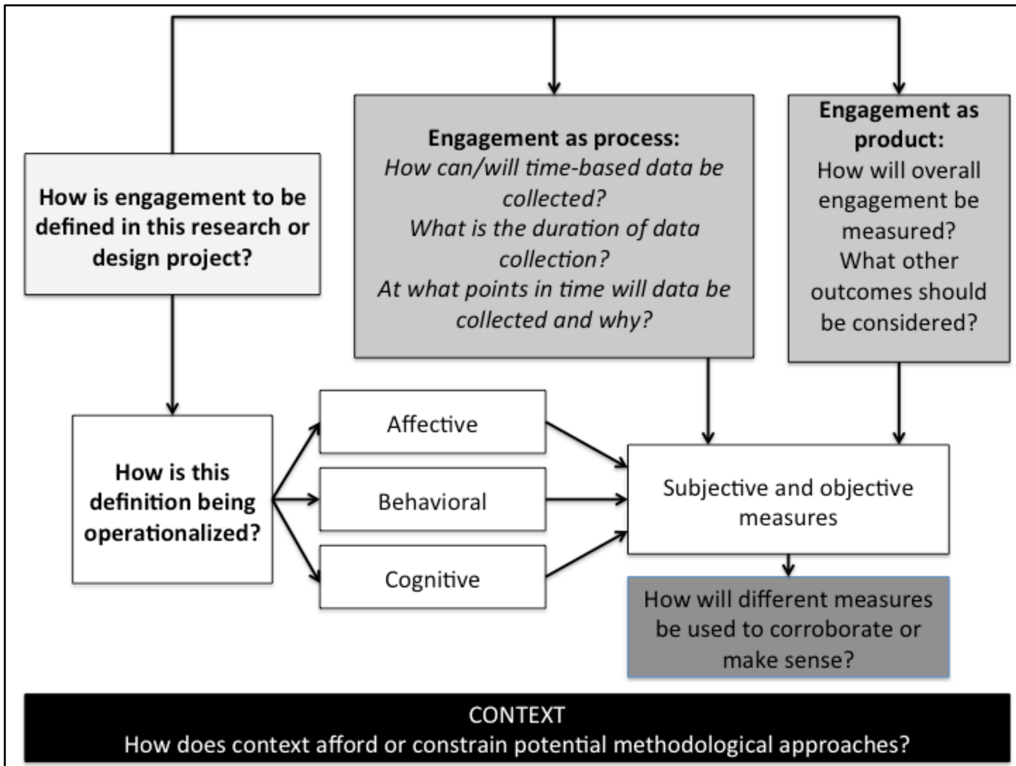

Figure 1. Methodological framework for evaluating user engagement

Firstly, how engagement is to be defined and operationalized is central to selecting an approach. What measures are most appropriate can only be determined once a researcher or designer has determined whether they are interested in UE as a process or product of interaction (or both), and whether they view UE has being affective, cognitive, behavioral or some combination. These decisions can then inform the identification of fitting objective and subjective measures that reflect the researchers' lens on UE. The use of multiple measures – and the ability to effectively use them in conjunction with each other – is essential for issues of validity and sense making. From the outset of any inquiry, the researcher or designer must consider how multiple data sources can be used for the purposes of corroboration, explanation, etc. and when/how they will be analyzed in concert to provide a rich picture of engagement in a particular setting. Finally, all design decisions must be informed by an examination of the context in which the designer/researcher and stakeholders are operating: what affordances and constraints of time, resources, or access to meaningful data are at work? This framework is not prescriptive, but focuses on guiding questions at key decision making junctures. In doing so, it helps to ground the researchers' methodological approach in their philosophy of UE and acknowledges the limitations and benefits of different types of measures and the settings in which we are working.

Further Considerations in the Evaluation of User Engagement

In addition to considering how UE is defined, operationalized and measured, a holistic examination of engagement must also take into account broader paradigmatic issues that operate upon researchers and designers. In the following section, I will briefly explore three considerations related to how we characterize the *quality* of the methodological approach, namely how different measures are evaluated on their own and in concert with other measures, the nature of objectivity, and making space for multiple perspectives.

The role of individual and collective measures in the study of user engagement

It is difficult to draw conclusions about engagement based on a single metric, and multiple measures are more robust. Even in cases where we are using validated measures, we need reassurance that it is behaving as we would expect it to in a new context. I have spent considerable time examining the reliability and validity of the UES through the use of other self-report, behavioral and physiological measures (O'Brien & Cairns 2015; O'Brien 2016b). As previously discussed, some researchers have had mixed results when it comes to corroborating different kinds of measures in the same study. One reason for this is that, while self-report questionnaires, neurophysiological signals, and behavioral metrics all produce quantitative data, the types of variables being measured, time series, and analytic procedures and approaches vary considerably amongst these different types of measures. It is therefore essential not to view the corroboration of metrics as one-to-one mappings where a questionnaire score is "equated" with a physiological measure, or where a behavioral metric is taken to "mean" engagement occurred; the same behavior in another setting may indicate something very different.

Meza-Kubo et al. (2016) took an additive approach to understanding the relationship between self-reports, expert observations, and EEG. Over two studies, the researchers developed, trained and evaluated a neural network using EEG signals to recognize pleasant and unpleasant emotions. The older adult participants were asked to interact with a cognitive wellness system that used a "Snakes and Ladders" like game. In their discussion of the results, the researchers pointed out well-known pitfalls of all three of the methods used, namely the bias of expert observers, the self-inflation of self-report data, and the potential of neurophysiological equipment to capture "noisy" data. They speculated that the combination of self-reports and EEG methods would increase the accuracy of the results obtained by the neural network to 70%, and that the addition of qualitative participant observation data would further increase this to 80%. In this example, the researchers explored what the self-report and the observations *added* to the accuracy of the neural network, rather than the strength of associations between measures. Therefore, we need to explore the robustness of individual metrics and the relationships between different measures both in concert and in parallel over time.

The "goodness" of a particular method

Third wave HCI advocates for different ways of knowing, and the integration of multiple perspectives and methodologies to advance the ability to understand and design for user experience (Harrison, et al 2007). Objective methods, such as eye tracking and behavioral measures, may be seen as more concrete and closer to "truth" than subjective methods, such as interviews and questionnaires, that rely on people to describe or rate their experiences. However, even objective methods are subject to some interpretation on the part of the researcher that affects the conclusions reached. For instance, physiological researchers must determine how data will be filtered and sampled and how to handle "noise" (Lebow & O'Brien 2012).

As de Guinea et al. (2012) have demonstrated, subjective self-reports are as effective as other measures when it comes to evaluating engagement. They administered three established self-report measures of cognitive load, engagement, and arousal and collected electroencephalographic (EEG) and electrocardiogram (EKG) data in an experimental study. The Multi-trait Multimethod Matrix (MTMM) statistical technique was used to assess the reliability and construct validity of the various measures and account for common-method variance (a threat to validity). Participants' neurophysiological baselines were captured before they completed a computer-based task. Their findings provided support for self-report measures. While the neurophysiological data was found to demonstrate less measurement error, the self-report data had greater content and construct validity. Self-report measures may capture more dimensions of a construct and "neurophysiological measurement may be subject to interactions with other physiological elements" (p. 568).

Part of the reason why subjective measures may be dismissed is due to the way in which they are adopted, adapted and employed in research studies. O'Brien and McCay-Peet (2017) looked specifically at this issue in the domain of interactive information retrieval. They highlighted that when questions are removed or modified from the original questionnaire, the reliability and validity of the measure is nullified; there are also issues in how self-report data is analyzed and reported[2]. In addition, these unfortunate practices in the use of questionnaires prevent researchers from evaluating the effectiveness of metrics over time and across contexts of use. In a review of how the UES has been employed since its publication, I was able to conclude that the questionnaire has demonstrated reliability and utility in diverse research scenarios. However, since few researchers have used it in its entirety and reporting practices were not always clear, I could not adequately assess the validity of the tool with confidence (O'Brien 2016b). This comes back to the point that the methodological toolbox must be equipped, but we must also understand the intricacies of the measures we are applying – whether questionnaires or EDA.

Room of multiple perspectives and approaches

Harrison et al. (2007) note that dominant paradigms affect what can be understood about HCI, including user engagement, stating that the dominant views will determine what questions should be asked and how they will be answered methodologically. The inclusion of objective and subjective measures can and should co-exist in the evaluation of engagement, and researchers can articulate their respective frames, recognizing the benefit of multidisciplinary approaches: "whatever our personal stance to research, multiple paradigms allow the field as a whole to develop a more complete understanding of the nature of interaction and good practices around design and evaluation" (Harrison et al. 2007, p. 13). If we recognize that "good work" occurs in each paradigm then notions of "validity" take on different meanings according to different perspectives (Harrison et al. 2007, p. 14), and we must concede that there is no one best metric or method to capture engagement, but there is value in doing whatever we do well.

---

[2] For a more in-depth discussion of issues inherent in questionnaire development, selection, adaptation, analysis and reporting, please see O'Brien and McKay-Peet (2017).

As the discussion of context in this chapter has demonstrated, findings may not be interpretable through statistically supported hypotheses, but may require interpretation of the literature or of other information gathered as part of the study, including the observations of the researchers and the words of the participants. Whether in laboratory environments or field settings, digital interactions are highly complex processes that are influenced by personal, social, and system related factors and may be open to multiple explanations.

Moving to a more nuanced approach to the study of engagement and embracing subjectivity does not necessarily mean abandoning experimental work. Bödker (2015) argues that lab-based research is pragmatic: "we often need to work more directly with technical experiments and participatory prototyping, for simple reasons of time, complexity, and the fundamental openendedness of the design space" (p. 26). However, as Mathur et al. (2016) demonstrated in their study, we need not think of the lab and field as disconnected spaces, but work to move between these locations of inquiry more fluidly. There is merit in testing effects in a controlled environment and critiquing their generalizability to the real world, or bringing clarity to an aspect of engagement in the lab without the "messiness" of an ever-shifting context, building more complex research designs iteratively.

Conclusion

This chapter has presented an overview of methodological approaches for evaluating user engagement, and has explored recent studies through propositions of an overarching framework of user engagement. The propositions ask researchers to consider:

- How do *I* define UE?
- Am I trying to capture engagement as a *process or product* of interaction?
- What is the anticipated or ideal *depth of the users' investment* in the digital experience?
- What *contextual variables* are prominent? How might these enhance or detract from UE in this setting?

In the selection of studies reviewed, how engagement is conceptualized and the methodological approaches is intertwined. The studies ranged from experimental to field-based, sometimes moving between the two contexts, and utilized qualitative (e.g., observations, interviews) and quantitative (e.g., physiological monitoring, self-report questionnaires) modes of data collection individually and collaboratively. The findings demonstrate the benefits and drawbacks of the methodological approaches and the particular perspective of engagement taken by the researchers. This idea was expanded upon through the discussion of broader measurement issues, specifically what methodological approaches are privileged according to objectivity versus subjectivity, their ability to produce corroborative results, and notions of validity.

In conclusion, this chapter argues for measuring user engagement using a thoughtful mix of qualitative and quantitative methods, considering the particulars of the use context, and balancing established and emerging subjective and objective metrics; in this way we are bridging second and third wave approaches rather than dismissing what we have learned about HCI from the past. It is not the methods of first and second wave HCI that are problematic, but the lack of flexibility with which they were employed.

This is reinforced by Harrison et al., who note, "[b]ecause of its emphasis on multiple perspectives, the third paradigm does not espouse a single, correct set of methods or approaches to answer these questions. Instead, we see a variety of approaches that are embedded in a similar epistemological substrate (Harrison et al., 2007, p. 8). As we move forward, notions of successful interactions have moved beyond efficiency, effectiveness and satisfaction, and humanistic and social science ways of knowing and being have entered the scene. Rather than a "plug and play" approach to measuring engagement, researchers and practitioners must define what engagement means in a given scenario, and appreciate the limitations and complexities of the methodological tools available to them and the contexts in which they operate. We must also strive to consider and learn from multiple perspectives if we hope to achieve a holistic understanding of what engagement is and how it impacts the users of digital technologies and HCI more broadly.

Acknowledgements
I wish to thank the Social Science and Humanities Council of Canada (SSHRC), the UBC Hampton Research Fund, and the Networks of Centres of Excellent Graphics, Animation and New Media Project (NCE GRAND) for generous funding support over the past several years. I would also like to acknowledge the colleagues – including and especially graduate students at the School of Library, Archival and Information Studies – who have worked with me over the years on various research projects for their insights, efforts, dedication and enthusiasm.

References

Absar R, O'Brien H, Webster ET (2014) Exploring social context in mobile information behavior. In: Proceedings of the American Society for Information Science and Technology, Seattle, WA, 31 October- 5 November 2014

Andujar M, Morreale P, Jimenez Y, Jimenez L, Gilbert JE (2017) Evaluation of user's affective engagement while interacting with educational technologies: A pilot study. In: Advances in Neuroergonomics and Cognitive Engineering, Springer, Heidelberg, p. 97-105

Anzalone SM, Boucenna S, Ivaldi S, Chetouani M (2015) Evaluating the engagement with social robots. INT J SOC ROBOT 7(4): 465-478

Agarwal R, Karahanna E (2000) Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. MIS QUART 24(4): 665-694

Aristeidou M (2016) Citizen inquiry: Engaging citizens in online communities of scientific inquiries. Dissertation, Institute of Educational Technology

Bagnara S, Pozzi S (2016) The third wave of human computer interaction: From interfaces to digital ecologies. DigitCult 1(1): 59-65

Bødker S (2015) Third-wave HCI, 10 years later---participation and sharing. Interactions 22(5): 24-31

Bødker S (2006) When second wave HCI meets third wave challenges. In: Proceedings of the 4th Nordic conference on Human-computer interaction: Changing roles, Oslo, Norway, 14-18 October 2006

de Guinea AO, Titah R, Léger PM, Micheneau T (2012) Neurophysiological correlates of information systems commonly used self-reported measures: A multitrait multimethod study. In: Hawaii International Conference on System Science (HICSS), IEEE, Maui, HI, 4-7 January 2012, p. 562-571

Flores FDN (2016) Engaging emerging adults in the design process of technology driven health interventions. Dissertation, The University of Wisconsin-Madison

Freund L, Kopak R, O'Brien H (2016) The effects of textual environment on reading comprehension: Implications for searching as learning. J INF SCI 42(1): 79-93

Gosling SD, Rentfrow PJ, Swann WB (2003) A very brief measure of the Big-Five personality domains. J RES PERS 37(6): 504-528

Harrison S, Tatar D, Sengers P (2007, April) The three paradigms of HCI. In: Alt. Chi. Session at SIGCHI Conference on Human Factors in Computing Systems, San Jose, CA, 28 April - 3 May, 2007

Jacques RD (1996) The nature of engagement and its role in hypermedia evaluation and design. Dissertation, South Bank University, London

Jacques R, Preece J, Carey T (1995) Engagement as a design concept for multimedia. CAN J EDUC COMMUN 24(1): 49–59

Jensen ML, Lee YH, Piercy CW, Dunbar NE, Elizondo J, Bessarabova E, Twyman NW, Burgoon JK, Valacich JS, Adame B, Miller C (2016) Exploring failure and engagement in a complex digital training game: A multi-method examination. AIS TRANS HUM COMPUT INT 8(1): 1-20

Laurel B (1993) Computers as theatre. Addison-Wesley.

Lalmas M, O'Brien HL, Yom-Tov E (2014) Measuring user engagement. SYN LECT INFORM CONCEPT RETR SERV 6(4): 1-132

Lebow M, O'Brien HL (2012) Is there a role for physiological methods in the evaluation of human-information interaction? http://dx.doi.org/10.14288/1.0107446. Accessed 15 May 2017

Leiker AM, Miller M, Brewer L, Nelson M, Siow M, Lohse K (2016) The relationship between engagement and neurophysiological measures of attention in motion-controlled video games: A randomized controlled trial. JMIR SERIOUS GAMES, 4(1): e4 doi: 10.2196/games.5460

Li C, Rusak Z, Horvath I, Kooijman A, Ji L (2016) Implementation and validation of engagement monitoring in an engagement enhancing rehabilitation system. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 99: doi: 10.1109/TNSRE.2016.2591183

Lohse KR, Boyd LA, Hodges NJ (2016) Engaging environments enhance motor skill learning in a computer gaming task. J MOTOR BEHAV 48(2): 172-182

Martey RM, Kenski K, Folkestad J, Feldman L, Gordis E, Shaw A, Stromer-Galley J, Clegg B, Zhang H, Kaufman N, Rabkin AN (2014) Measuring game engagement: multiple methods and construct complexity. SIMULAT GAMING 45(4-5): 528-547

Mathur A, Lane ND, Kawsar F (2016) September. Engagement-aware computing: modelling user engagement from mobile contexts. In: Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 12-16 September 2016, p. 622-633

McAuley E, Duncan T, Tammen VV (1989) Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. RES

Q EXERCISE SPORT 60(1): 48-58

McCarthy J, Wright P (2004) Technology as experience. Cambridge, MA: MIT Press.

Meza-Kubo V, Morán AL, Carrillo I, Galindo G, García-Canseco E (2016) Assessing the user experience of older adults using a neural network trained to recognize emotions from brain signals. J BIOMED INFORM 62: 202-209

Nguyen C (2015) Understanding user engagement in the open collaboration model of crowdsourcing. Dissertation, University of Nebraska at Omaha

O'Brien HL (2011) Exploring user engagement in online news interactions. In: Proceedings of the Association for Information Science and Technology, New Orleans, LA, 9-12 October 2011

O'Brien HL (2016a) Theoretical perspectives on user engagement. In: O'Brien HL, Cairns P (eds) Why Engagement Matters: Cross Disciplinary Perspectives on User Engagement, Springer International Publishing, p. 1-26

O'Brien HL (2016b) Translating theory into methodological practice. In: O'Brien HL, Cairns P (eds) Why Engagement Matters: Cross Disciplinary Perspectives on User Engagement, Springer International Publishing, p. 27-52

O'Brien HL, Cairns P (2015) An empirical evaluation of the User Engagement Scale (UES) in online news environments. INFORM PROCESS MANAG 51(4): 413-427

O'Brien HL, Freund L, Kopak R (2016) Investigating the role of user engagement in digital reading environments. In: Proceedings of the ACM on Conference on Human Information Interaction and Retrieval, Carrboro, NC, 13-17 March 2017, p. 71-80.

O'Brien HL, Lebow M (2013) Mixed-methods approach to measuring user experience in online news interactions. J ASSOC INF SCI TECH 64(8): 1543-1556

O'Brien HL, McKay-Peet L (2017) Asking good questions: Questionnaire design and analysis in interactive information retrieval research. In: Proceedings of the Conference on Conference Human Information Interaction and Retrieval, Oslo, Norway, 7-12 March, p. 27-36

O'Brien HL, Toms EG (2008) What is user engagement? A conceptual framework for defining user engagement with technology. J ASSOC INF SCI TECH 59(6): 938-955

O'Brien HL, Toms EG (2010) The development and evaluation of a survey to measure user engagement. J ASSOC INF SCI TECH 61(1): 50-69

Ponciano L, Brasileiro F (2015) Finding volunteers' engagement profiles in human computation for citizen science projects. arXiv preprint arXiv:1501.02134. Accessed 15

May 2017

Stroop JR (1935) Studies of interference in serial verbal reactions. J EXP PSYCHOL 18(6): 643-662

Watson D, Clark LA, Tellegen A (1988) Development and validation of brief measures of positive and negative affect: the PANAS scales. J PERS SOC PSYCHOL 54(6): 1063-1070

Webster J, Ho H (1997) Audience engagement in multimedia presentations. ACM SIGMIS Database 28(2): 63-77

Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, Merchant GC, Naughton F, Blandford A (2016) Understanding and promoting effective engagement with digital behavior change interventions. AM J PREV MED 51(5): 833-842

Zhang J, Hamilton WL, Danescu-Niculescu-Mizil C, Jurafsky D, Leskovec J (2017) Community identity and user engagement in a multi-community landscape. In: Proceedings of the Interactional Conference on Web and Social Media, Montreal, QC, 15-18 May 2017