The Development of a Multidimensional Engagement Measure

Emergent Research Forum (ERF)

Stephanie A. Andel
University of South Florida
sandel@usf.edu

Gert-Jan de Vreede
University of South Florida
gdevreede@usf.edu

Vivek Singh
University of South Florida
vivek4@mail.usf.edu

Triparna de Vreede
University of South Florida
tdevreede@usf.edu

Balaji Padmanabhan
University of South Florida
bp@usf.edu

Paul E. Spector
University of South Florida
pspector@usf.edu

Abstract

Over the past decades, user engagement has become central to product success more than ever before. However, while engagement has been studied extensively in various disciplines, these bodies of knowledge are siloed. One indication of this separation is that there is no existing user engagement measure that can be used consistently and reliably across research domains. This emergent research paper aims to progress towards bridging this gap by developing an engagement scale that may be generalized to multiple disciplines. To that effect, we first identified engagement as a three dimensional phenomenon and developed definitions for each dimension. Next, we conducted a series of four studies to develop and validate a measure of user engagement. This resulted in a 16-item measure to assess cognitive, behavioral, and emotional engagement. Future steps include further refinement of the measure, and additional studies to test the generalizability of the scale across disciplines.

Keywords

Cognitive engagement, behavioral engagement, emotional engagement, survey development

Introduction

As a result of the digital explosion in the past couple of decades, human computer interaction has moved beyond providing simple and intuitive user interfaces to creating engaging experiences (Shankar et al., 2016). The proliferation of web based alternatives means that users have unlimited options for the platforms that offer the content of their choice. Interfaces that do not engage these users are dismissed without much thought. Users move on to others that provide the same services and yet retain their interest. Therefore, the hallmark of successful technologies now has an extra requirement of engagement in addition to usefulness and usability. In other words, the very survival of technology-based platforms may be dependent on the degree of engagement they can elicit from the users and the speed with which this engagement is built.

This sentiment regarding engagement appears to be reflected across multiple disciplines. For example, the education domain has an extensive engagement literature that examines how engagement among students can be enhanced in classrooms and academic institutions (Christenson, Reschly, & Wylie, 2012). Organizational psychology considers the engagement of employees within the workplace and the impact of employee engagement on important work-related outcomes such as employee productivity, satisfaction, and turnover intentions (Bailey, Madden, Alves, & Fletcher, 2017). Additional engagement literatures can be found in marketing and e-commerce (i.e., customer engagement; van Doorn et al., 2010) and online user experience domains (O’Brien & Toms, 2008).
Despite the wealth of research on user engagement, these literatures are siloed within their own disciplines. This separation hinders efficient progress regarding what we know about this significant concept and prevents researchers from building upon previous research that has been conducted in other domains. As a result, this disconnect has led to various issues, including inconsistent definitions of engagement (Fredricks & McColskey, 2012) and multiple concurrent domain specific measures of engagement. Many measures of engagement, while robust, are so specific to the domains where they were developed, that they cannot easily be generalized to other fields. Consequently, other unrelated domains are not able to effectively take advantage of advances in engagement research.

In summary, despite the critical need to reliably evaluate engagement, it appears that there are no agreed upon and validated scales that measure user engagement consistently and reliably across disciplines. To establish a scale that can assess engagement levels of all users accurately, it is important to define engagement and identify its measurable and identifiable components that remain stable irrespective of the platform where it is measured.

The purpose of this research in progress paper is to develop a discipline independent engagement scale and to test its reliability and validity on a single domain (i.e., online learning), with the ultimate goal of validating the scale across various additional contexts. We aim to integrate the synergies of engagement literature and allow researchers to seamlessly build upon each other’s work. The next section provides a review of the engagement literature from several disciplines to create a domain independent engagement scale. These disciplines include sales, organizational behavior, education, and human computer interaction. The method section elaborates on the process of evaluation and validation of the engagement scale over three iterations. Finally, the results of the study, its implications, and future directions are discussed.

**Current Definitions and Contributions**

This study aims to unify the engagement literature by providing a clear definition of the term and development of a self-report scale that can be effectively used across disciplines. There are many views on the nature of the engagement concept. For instance, engagement in Internet-based platforms can involve playfulness, sensory integration, first impressions of the platform, and the enjoyment that users experience (e.g., Laurel, 1993; Quesenbery, 2003). Other characteristics of engagement that have been identified are degree of control, level of challenge, alignment to goals, type of presentation, and choice (e.g., Appleton, Christenson, Kim, & Reschly, 2006, Webster & Ho, 1997). Regardless of these variations, researchers agree that engagement is a multidimensional construct. Although engagement can be broken down in various ways, one of the most widely accepted and agreed upon conceptualizations of engagement suggests that the construct is made up of three different components or sub-dimensions, namely, behavioral, cognitive, and emotional engagement (Fredricks & McColskey, 2012).

**Cognitive Engagement**

Definitions of cognitive engagement vary, ranging from beliefs about the value or importance of a task at hand, learning goals, and the use of cognitive strategies to complete a task, to learning self-regulation techniques and going beyond the basic requirements of a task (Fredricks & McColskey, 2012). For instance, the School Engagement Measure assesses cognitive engagement as the use of cognitive strategies with items such as, “When I read a book, I ask myself questions to make sure I understand what it is about” (Fredricks et al., 2005). Others define cognitive engagement as complete absorption and focus on a task (e.g., May, Gilson, & Harter, 2004; “Performing my job is so absorbing that I forgot about everything else”), which is what others might consider as “flow” (Csikszentmihalyi, 1990). In the current study, we define cognitive engagement as the extent to which individuals are willing to extend their intellectual efforts beyond the minimum required to complete the task.

**Emotional Engagement**

As with cognitive engagement, the definitions of emotional engagement also vary across the engagement literature. For example, within the academic space, researchers measure this emotional dimension in
terms of positive affective reactions to the academic experience, including emotions such as happy, interested, and excited (e.g., Fredricks et al., 2005; “I am interested in the work at my school”). Others measure this construct as the degree to which one feels a sense of belongingness to their school (e.g., Finn, 1989; “I feel proud of being a part of my school”). Within the organizational psychology literature, researchers such as May et al. (2004) operationalize emotional engagement as an emotional attachment to the workplace experience (e.g., “My own feelings are affected by how well I perform my job”). Further, other researchers focus upon emotional disengagement, defined as the extent of negative emotions that result from the task (e.g., “When we work on something in class, I feel discouraged”; Skinner, Kindermann, & Furrer, 2009). In the current study, we define emotional engagement as the extent to which individuals experience a positive reaction and attachment towards the task.

Behavioral Engagement

Finally, behavioral engagement is also commonly assessed within the various engagement literatures. For instance, many researchers conceptualize behavioral engagement as basic compliance with a task, and examine things such as participation, concentration, effort, and adherence to rules and instructions (Fredricks & McColskie, 2012). For instance, items such as “I pay attention in class” (Fredricks et al., 2005) and, “When I’m in class, I listen very carefully” (Skinner et al., 2009) are commonplace within the academic literature. Within the marketing domain, customer engagement is described as a customer’s behavioral manifestation of motivation toward a specific brand or product (Van Doorn et al., 2010), while organizational psychology considers behavioral engagement to be discretionary and externally manifested effort exerted toward one’s job tasks (Macey & Schneider, 2008). In the current study, we define behavioral engagement as the extent to which individuals are persistent, exert effort, participate, and are involved with the task at hand.

Method

Scale development requires a specific series of steps in order to construct a reliable and valid scale. Specifically, as outlined by MacKenzie, Podsakoff, and Podsakoff (2011), proper scale development requires the following ten steps: 1) form a conceptual definition of the construct, 2) generate items, 3) assess content validity, 4) specify the measurement model, 5) collect data to pre-test, 6) scale refinement, 7) gather new sample and re-examine scale properties, 8) assess scale validity (repeat 6-8 as needed), 9) cross-validate the scale with samples from different populations, and 10) develop scale norms. This paper reports on the first eight steps of this process. The conceptual definition (Step 1) can be found in the background of this submission.

After creating the conceptual definition (Step 1) and reviewing the literature, a series of items were generated by the researchers to measure each sub-dimension of engagement (Step 2). In the first iteration, emotional and cognitive engagement sub-dimensions had six items each and behavioral engagement had seven items. Once the items were developed, the scales were evaluated by a team of six researchers to assess the content validity of the items by ensuring that they were adequately reflective of the engagement sub-dimensions (Step 3). These researchers were from the disciplines of psychology, management, and information systems. Each had prior experience in engagement research. The feedback from the team led to adjustments to some of the items to ensure that they were clearly worded.

Next, a series of data collection efforts were conducted through Amazon’s Mechanical Turk (MTurk). MTurk was deemed appropriate for this study, as previous research has shown that it is a feasible and cost-effective way to gather psychometrically valid and reliable data (e.g., Buhrmester, Kwang, & Gosling, 2011). In the first round of data collection, 133 individuals were recruited from Amazon’s Mechanical Turk (MTurk) to watch a brief 2-minute video and complete a survey with the engagement items (Steps 4 & 5). Based on the results and the resulting adjustment to the scale, a second round of data was collected from a sample of 120 MTurk workers using the same task. The scale was further refined and a third sample of data was collected from 120 MTurk workers using the same task as the previous two times (Steps 6 & 7). Finally, a fourth sample of data was collected with 115 individuals being recruited from MTurk to complete the survey items, which validated the final 16-item engagement scale (Step 8).
Results

Exploratory factor analysis (EFA) was conducted on the first sample to assess item loadings and to empirically derive the factor structure of the data (Steps 4 and 5). Upon inspection of the eigenvalues and scree plot, it was found that three factors explained most of the variance. Specifically, 58% of the total variance was explained, with factor 1 explaining 33%, factor 2 explaining 18%, and factor 3 explaining 7%. In alignment with Tabachnick and Fidell (2007), only items that loaded onto a primary factor of at least .50 and no other factor at .32 or more were retained (Step 6). This resulted in seven, eight, and eight items to measure cognitive, emotional, and behavioral engagement, respectively.

In order to assess the properties of the new items, a second sample was collected (Step 7). An EFA was run (Step 6), and as before, three factors explained most of the variance. Specifically, 58% of the total variance was explained, with factor 1 explaining 42%, factor 2 explaining 10%, and factor 3 explaining 6%. The item pool was again refined based on the guidelines by Tabachnick and Fidell (2007). Further, the construct definition of cognitive engagement was re-assessed, which led to the creation of eight new items.

Consequently, a third sample was collected (Step 7). An EFA was run, and three factors again explained most of the variance. Specifically, 60% of the total variance was explained, with factor 1 explaining 38%, factor 2 explaining 15%, and factor 3 explaining 7%. Twelve items for emotional, eight for behavioral, and four for cognitive) met the criteria outlined by Tabachnick and Fidell (2007), and therefore were retained.

Finally, a fourth sample was collected, and confirmatory factor analyses (CFA) were run to validate the factor structure of the 22-item engagement scale (Step 8). This model (i.e., Model 1) demonstrated poor fit (SB-$\chi^2(206) = 599.88$, $p < .05$, CFI=.85, TLI=.83, SRMR=.10, RMSEA=.098). Upon inspection of correlations among the items, it was found that the correlation between two behavioral engagement items was high enough to warrant potential multicollinearity ($r = .83$). Therefore, the item with the lower factor loading was removed. Further, Garson (2015) suggests that item loadings should be at least .70. Thus, six items that did not meet these criteria were removed with the exception of one item, removal of which would have caused the cognitive engagement latent variable to be underidentified. This model (i.e., Model 2) demonstrated better fit (SB-$\chi^2(101) = 245.29$, $p < .05$, CFI=.91, TLI=.90, SRMR=.07, RMSEA=.085). A chi-square difference test showed that Model 2 fit significantly better than Model 1, $\Delta \chi^2$SB(105) = 264.59, $p < .001$. Cronbach’s alpha for the final cognitive, behavioral, and emotional engagement subscales were .79, .94, and .92, respectively. See Table 1 for the final engagement scale items.

<table>
<thead>
<tr>
<th>Item</th>
<th>Cognitive</th>
<th>Behavioral</th>
<th>Emotional</th>
</tr>
</thead>
<tbody>
<tr>
<td>I asked myself questions to check if I understood the contents of the video</td>
<td>0.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I had a learning goal prior to watching this video</td>
<td>0.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watching this video met some of my learning goals</td>
<td>0.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I listened to the contents of the video as I was expected to</td>
<td>0.74</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>I made an effort to watch the video in its entirety</td>
<td></td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>I dutifully followed the instructions for this activity</td>
<td>0.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I was being attentive to what was being said in the video</td>
<td>0.78</td>
<td></td>
<td>0.91</td>
</tr>
<tr>
<td>I was actively involved in watching the video</td>
<td></td>
<td></td>
<td>0.85</td>
</tr>
<tr>
<td>I diligently watched the video</td>
<td>0.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I was dutifully paying attention to this video</td>
<td>0.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>It made me happy to watch this video</td>
<td></td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>I care about the contents of this video</td>
<td>0.71</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>I liked what was being said on this video</td>
<td>0.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>This video had a positive impact on my mood</td>
<td>0.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I enjoyed watching this online video</td>
<td>0.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>It was interesting to watch this video</td>
<td>0.84</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Final Items with Factor Loadings

*Italicized text may be replaced when using scale in other contexts and domains.

Conclusions and Future Steps

The objective of this study is to bridge the various isolated engagement literatures by developing a measure of engagement that can be easily modified for use across disciplines. Future steps include further
refinement of the final scale (e.g., adding additional cognitive engagement items, and replacing the cognitive engagement item that does not meet Garrison’s (2015) criteria). Additional studies will also be conducted to assess convergent and criterion-related validity of the scale by examining correlations between each subscale and various outcomes of interest (e.g., retention, satisfaction, motivation). Finally, as the ultimate goal of this research is to develop a scale that spans contexts, we will further validate the scale within other contexts and populations, such as online education, marketing, and workplace employees. Ultimately, we are hopeful that this new scale will integrate the synergies of the engagement literature and allow researchers to seamlessly build upon each other’s work.

Acknowledgments

This research was partly supported by OnQ and NSF Grant 1655085.

References


