

MUX: Development of a Holistic Mobile User Experience Instrument

Soussan Djamasbi
 Worcester Polytechnic Institute
 djamasbi@wpi.edu

E. Vance Wilson
 Worcester Polytechnic Institute
 vwilson@wpi.edu

Abstract

User experience practitioners have benefitted from the availability of usability measures that are short, generalizable, and easy to interpret, such as the System Usability Scale. However, such generalizable instruments fail to address many of the key characteristics of software applications that can be used across a range of mobile devices. Our response is development of MUX, a mobile user experience instrument that provides holistic assessment of specific software-device use scenarios. We find that MUX self-report scales assessing Nuisance, Mobility, and Access can be applied to effectively augment measures of the System Usability Scale or can be used as a standalone instrument for rating and comparing user experiences in mobile computing.

1. Introduction

According to the 2015 Technology Device Ownership Report by the Pew Research Center [21], ownership of mobile computing devices by U.S. adults grew to near-saturation between 2011 and 2015. During this period smartphone ownership approximately doubled, increasing from 35% to 68%, and tablet ownership nearly tripled, growing from 13% to 45%.

People enjoy using mobile devices not only for anywhere/anytime communication, but also as a means to quickly and conveniently connect to news and other information [22]. Indeed, a recent survey reports that 44% of owners sleep next to their smartphone in order to not miss calls, text messages or updates through the night and 29% state that they can't imagine living without their smartphone [20].

As mobile devices proliferate they often are used to replace other technologies, such as televisions and computers. Because mobile devices are made to be portable, they typically have smaller screen sizes and keyboards [18]. As a consequence, designing applications and content that work well for multiple

screen and keyboard sizes has become a challenge for web developers [10][11]. The ability to measure and compare user experience with software applications across a range of mobile devices could provide developers with invaluable insights. However, relatively little research has been completed to date to develop measures specific to user experience with mobile devices, especially research that attempts to integrate with and augment existing industry approaches to assessment of user experience. Our objective in this paper is to develop and initially validate such a mobile user experience instrument.

2. Background

Readers may note that a survey instrument recently has been introduced to assess the usability of mobile applications [17]. This mobile application usability (MAU) instrument captures user reactions in six dimensions: 1) application design, which captures the overall reaction to design, 2) application utility, which captures user perception of the usefulness of the application, 3) user interface graphics, which captures whether users find the interface graphics are designed effectively or not, 4) user interface input, which captures whether people have the impression that they can input data easily, 5) user interface output, which captures whether people feel that content is presented effectively, and 6) user interface structure, which captures whether users think that the application has an overall effective structure.

The MAU instrument was developed following rigorous guidelines. Yet, while it captures important usability aspects of mobile applications, it does not address perceptual characteristics of mobile devices. Research shows that people use mobile devices not only to gain access to communication and information, e.g., by using mobile applications, but also because these devices give them the ability to gain access quickly and conveniently and from virtually anywhere and at any time [20][21][22]. These capabilities are based on the premise that mobile devices can be carried portably from one place to another. Thus, we

propose that the content of a *holistic* mobile user experience instrument must include assessments of the mobile device in addition to the software application that is being used on it. Based on characteristics noted in the literature cited above, these criteria should—at a minimum—include speed of use, convenience, connectivity, mobility, and portability.

Performance of these criteria is likely to be affected by the type of mobile device (e.g., laptop computer, tablet, or smart phone) as well as the feature set that the device implements. For example, the overall experience of using a website is shaped not only by the website’s design but is also influenced by how quickly and conveniently the website can be accessed and what locations it can be accessed from. Even within the range of mobile computing devices (e.g., tablet vs. smart phone), we might expect these attributes will affect the overall user experience of an application. Despite its rigor and other qualities, the MAU instrument cannot adequately represent the *universe of content* [9] that comprises the holistic mobile user experience, as this content intrinsically includes characteristics of the mobile device that is being used.

We also note that instruments designed for use by industry practitioners must meet pragmatic limits focusing on brevity, generalizability, and ease of interpretation. One example of a successful usability instrument is the System Usability Scale (SUS) [6]. SUS is a 10-item survey that is widely used in industry to assess user experience of products and services [1][24]. Although designed prior to the era of mobile devices and applications, SUS often is used to assess them (e.g., [10][15]).

Benefits for practitioners who use SUS are (1) the quickness of administering this short instrument, (2) an abundance of existing empirical data that supports generalizable comparison of usability across products, and (3) output scoring that is easy to interpret. SUS produces a single usability score that can range between 0-100. Data collected over decades of SUS administration show that scores above 85 indicate excellent user experience, between 70 and 85 a good user experience, between 50 and 70 an acceptable experience, and below 50 a poor experience[4][5].

A key feature of SUS is that it can capture aspects of experience that are technology independent, hence it can be used to assess various systems from hardware to software applications [23]. However, SUS has no ability to capture certain experiences that are unique to mobile user experience, such as speed of use, convenience, connectivity, mobility, and portability factors that we identified previously. Research shows such factors are important to users, particularly for applications that are designed to be used on multiple types of mobile devices [14]. Paying attention to users’

perception of these attributes could provide important insight for pinpointing sources of influence that shape user reactions to a specific device or across multiple devices.

Our objective in this line of research is to develop and initially assess a mobile user experience instrument with the following characteristics:

- Provides a holistic assessment of mobile user experience by jointly addressing software applications and physical device types, thereby representing the universe of content that surrounds the mobile user experience
- Can effectively augment SUS scores
- Is quick to administer
- Is easy to interpret

In the following sections we outline our research method, report results of our initial assessment of the mobile user experience (MUX) instrument, and discuss our findings, recommendations for future study, and conclusions.

3. Research Method

The MUX instrument was designed, purified, and validated in a three-stage process. In Stage 1, we identified concepts and items to represent the content of user experience that is specific to use of software applications on mobile devices. In Stage 2, we administered these items to subjects and used the results to identify emergent factors and to purify the scales. In Stage 3, we validated the scales with a separate subject population. These stages are described in the following sections.

3.1. Stage 1 Procedure

In Stage 1, we reviewed studies that address user expectations of their mobile user experience [20][21][22] in order to identify and populate conceptual categories representing the universe of content that is specific to this area. Five factors emerged from this review: *Speed of use*, *convenience*, *connectivity*, *mobility*, and *portability*. In addition to these, we included two further categories that were not directly identified in the literature but we felt to be important to the mobile user experience. *Viewability* was conceptualized as the ease of viewing text and graphics on the device. *Interaction* was conceptualized as the ability to interact with the software application using the device for input and navigation.

Subsequently, we developed items to represent each content category following procedures outlined by Gable and Wolf [13] and incorporating the domain-referenced approach developed by Anderson [2]. In

this approach, the target, direction, and intensity of each characteristic under study is explicitly assessed to ensure that resulting measurement items and semantic transformations reflect the underlying category.

For each scale, at least five reflective items were created [3]. In order to anchor the concepts, each item was directed toward use of a specific software and device type. Items were phrased with approximately equal numbers of positive and negative wordings. All items were phrased to elicit a response of agreement or disagreement with the item, measured using a five-position scale with endpoints marked Strongly Agree (1) and Strongly Disagree (5). The complete list of initial content categories and measurement items is shown in Table 1.

3.2. Stage 2 Procedure

The objective of Stage 2 was to pilot-test and refine the initial scale items developed in Stage 1 (see Table 1). Subjects in Stage 2 were 174 students attending undergraduate information systems courses at a large university in the Midwest U.S. Gender distribution of subjects was 94 males (54%) and 80 females (46%), with average age of 21 years. By voluntarily participating in the study or completing an alternative assignment, subjects earned extra course credit.

Students who had registered to participate in the study were notified to begin via an email message that contained participation instructions and a hyperlink to access an online survey study implemented using Qualtrics survey software. The survey was available for completion during a period of one week following notification, and registrants who had not completed the survey after five days were sent a follow-up reminder message via email.

The introductory screen for the survey explained the objectives of the study and rights of participants. After indicating their agreement to participate, subjects were asked to “Select the type of device you used MOST RECENTLY to access [name of the university’s learning management system (LMS)]”. Offered selections were “Smart Phone”, “Tablet”, “Laptop PC”, and “Desktop PC”. Subsequent questions in the survey were phrased to ask about use of the selected device for accessing the LMS, e.g., “Using a Tablet to access [name of the university’s LMS] was handy.” After completing the main portion of the survey containing items shown in Table 1, subjects were asked a set of demographic questions including their age, sex, and identification for extra credit. They then exited the survey.

Results from Stage 2 administration were entered into SPSS Exploratory Factor Analysis (EFA) using Principal Components Analysis for initial extraction

followed by the Varimax orthogonal rotation procedure. The number of factors to be extracted was based on the default of including factors having eigenvalues > 1 —a criterion supported by review of the scree plot for this analysis—and including at least one item loading > 0.5 . Three factors were identified in the EFA. Each emergent factor combined items from multiple initial content categories, however, there was no overlap, i.e., each initial category contributed items to only one of the emergent factors. Based on our interpretation of the item content in each factor, we assigned them the following titles:

- *Nuisance*, comprising negatively-worded items from speed of use, convenience, and connectivity;
- *Mobility*, comprising items from mobility and portability; and
- *Access*, comprising items from viewability and interaction

Items that did not load on the emergent factors, that cross-loaded substantially, or that had relatively low loadings were removed through an iterative purification process [7][8][19]. This process resulted in retention of the items marked in Table 1 within parentheses following the retained item. Five items were retained for each emergent factor. Factor loadings ranged 0.829 to 0.901 for nuisance, 0.828 to 0.896 for mobility, and 0.634 to 0.884 for access. The three emergent factors cumulatively explain 73% of variance reported in the EFA.

3.3. Stage 3 Procedure

The objective of Stage 3 was to validate the MUX instrument. Subjects were 171 students attending undergraduate information systems courses at a large university in the Midwest U.S. Gender distribution of subjects was 97 males (57%) and 80 females (44%), with average age of 22 years. By voluntarily participating in the study, or by completing an alternative assignment, subjects earned extra course credit. None of the subjects in Stage 2 participated in the Stage 3 study.

Using the procedure described in Stage 2, students who registered to participate in the study were notified to begin via an email message providing access to the online survey study implemented using Qualtrics survey software. Administration order of items in the Stage 3 survey was individually randomized for each subject, as recommended by Wilson and Lankton [25]. The survey was available for completion during a period of one week following notification, and registrants who had not completed the survey after five days were sent a follow-up reminder message via email.

Table 1. Initial and Final Content Categories with Measurement Items*

Speed of use content	<p>I think using a [device] to access [software] would help me accomplish tasks quickly.</p> <p>I felt using a [device] to access [software] would slow me down. (<i>Nuisance**</i>)</p> <p>Using a [device] to access [software] helped me work efficiently.</p> <p>Using a [device] to access [software] would keep me from accomplishing tasks efficiently.</p> <p>I would characterize using a [device] to access [software] as fast in operation.</p> <p>I would characterize using a [device] to access [software] as slow in operation.</p>
Convenience content	<p>Using [device] to access [software] made it convenient to access information.</p> <p>Using [device] to access [software] was handy.</p> <p>Using [device] to access [software] was inconvenient. (<i>Nuisance**</i>)</p> <p>Using [device] to access [software] was not handy.</p> <p>It was very convenient to use a [device] to access [software].</p> <p>Completing tasks using a [device] to access [software] inconvenienced me. (<i>Nuisance**</i>)</p>
Connectivity content	<p>Using a [device] to access [software] helped me stay connected no matter where I am.</p> <p>Using a [device] to access [software] allowed me to maintain my connections.</p> <p>Using a [device] to access [software] let me to be continuously in touch.</p> <p>Using a [device] to access [software] gave me the feeling that I was part of what was going on.</p> <p>Using a [device] to access [software] made me feel isolated. (<i>Nuisance**</i>)</p> <p>Using a [device] to access [software] made me feel disconnected. (<i>Nuisance**</i>)</p>
Mobility content	<p>I was be able to use a [device] to access [software] almost everywhere I went.</p> <p>I thought a [device] could only be used to access [software] in a small number of locations.</p> <p>Using a [device] to access [software] would improve my ability to be mobile. (<i>Mobility**</i>)</p> <p>I would be able to use a [device] to access [software] on the go. (<i>Mobility**</i>)</p> <p>I believe using a [device] to access [software] would hinder my mobility.</p> <p>Using [device] to access [software] would tie me down.</p>
Portability content	<p>I would characterize a [device] used to access [software] as compact.</p> <p>I would characterize a [device] used to access [software] as bulky.</p> <p>I think a [device] used to access [software] would be easy to carry with me. (<i>Mobility**</i>)</p> <p>I feel a [device] used to access [software] would be very portable. (<i>Mobility**</i>)</p> <p>I would be able to take a [device] used to access [software] with me almost everywhere I go. (<i>Mobility**</i>)</p>
Viewability scale items	<p>I had difficulty using a [device] to read text when accessing [software].</p> <p>It was easy to view things using a [device] to access [software].</p> <p>A [device] provided a good view of information when accessing [software]. (<i>Access**</i>)</p> <p>Using a [device] to access [software] could strain my eyes.</p> <p>I had difficulty using a [device] to view images when accessing [software].</p> <p>I had no trouble viewing text when using a [device] to access [software]. (<i>Access**</i>)</p>
Interaction content	<p>When using a [device] to access [software] I have difficulty clicking on links or buttons</p> <p>Entering text is difficult when using a [device] to access [software].</p> <p>Navigating between screens using a [device] to access [software] was cumbersome</p> <p>Clicking on links or buttons was easy to accomplish using a [device] to access [software]. (<i>Access**</i>)</p> <p>I have no problem entering text when using a [device] to access [software]. (<i>Access**</i>)</p> <p>Using a [device] to access [software] makes it easy to navigate between screens. (<i>Access**</i>)</p>

* Five-position response endpoints were marked Strongly Agree (1) and Strongly Disagree (5); [software] was replaced in the survey by the name of the university's learning management system (LMS); [device] was replaced in the survey by the device that the subject selected from a list to describe the device had last used to access the university's LMS (smartphone, tablet, laptop computer, desktop computer)

** Retained under the noted factor title following exploratory factor analysis of Stage 2 data

Data from Stage 3 were entered into two CFA models using AMOS 22 software. CFA of the MUX three-factor model produced a factor structure in which all measurement items load significantly on the anticipated latent factor (see Table 2). The three-factor model produced fit statistics that meet established threshold standards (Hair et al., 2009): GFI = 0.912, AGFI = 0.866, NFI = 0.938, and RMSEA = 0.060. A second CFA was run on a model which assesses effects of all 15 items on a single factor. All goodness of fit measures were lower for this one-factor model than for the three-factor model, and in no case do the one-factor model measures meet threshold standards (GFI > 0.9, AGFI > 0.8, NFI > 0.9, RMSEA < 0.8).

3.4. Construct Validation

Convergent validity in the three-factor was assessed by calculating Cronbach's alpha and composite reliability (CR) of each MUX factor (see Table 3). Both Cronbach's alpha and CR were 0.86 or higher for every factor, exceeding the .70 criteria proposed by Hair et al. [16].

Table 2. MUX CFA Loadings (Stage 3 Data)

Measure	Loading
Nuisance1	0.788
Nuisance2	0.829
Nuisance3	0.735
Nuisance4	0.727
Nuisance5	0.816
Mobility1	0.879
Mobility2	0.821
Mobility3	0.703
Mobility4	0.736
Mobility5	0.650
Access1	0.686
Access2	0.768
Access3	0.703
Access4	0.731
Access5	0.475

Discriminant validity was assessed through analysis of average variance extracted (AVE). The AVE for each latent factor is greater than .50, and the square root of AVE is higher than any correlation of that latent factor with any other factor, thereby meeting criteria proposed by Fornell and Larcker [12]. Based on these results, the three-factor MUX measurement model demonstrates satisfactory construct validity.

Table 3. AVE, Reliability, and Correlations between MUX Scales using Stage 3 Data*.

Factor	AVE	Alpha	CR	Mobility	Nuisance	Access
Mobility	0.67	0.92	0.91	0.82		
Nuisance	0.70	0.93	0.92	-0.18	0.84	
Access	0.57	0.86	0.86	-0.13	0.33	0.75

* Square roots of average variances extracted (AVEs) are shown in bold on the diagonals

3.5. Augmenting MUX with SUS Measures

MUX was designed to holistically assess aspects of user experience in mobile settings. Because this context did not yet exist when SUS was developed, it is not reasonable to expect that SUS will measure all these aspects, despite its other practical benefits. SUS has been used for decades for general-purpose assessment of products and services, and a large portion of the instrument's value is the development of a substantial inventory of SUS scores and distribution data over this period [4]. Thus, our principal objective in this study is to augment SUS. To accomplish this objective, it is important to understand how MUX and SUS relate to one another.

In order to address this question, we conducted EFA on Stage 3 data including SUS items in addition to MUX items. We chose Principal Components Analysis for initial extraction followed by the Varimax orthogonal rotation procedure. The number of factors to be extracted was based on the default of including factors having eigenvalues > 1 and including at least one item loading > 0.5.

As shown in Table 4, the three MUX factors load separately in the same manner as with the Stage 2 data reported earlier. Although SUS is typically considered to be a unidimensional measure, its measurement items loaded on three factors. SUS items 1, 3, 7, and 9 loaded on the new Factor 4, which emphasizes ease of use. All negatively-phrased SUS items (i.e., 2, 4, 6, 8, and 10) loaded on Factor 1 with the MUX Nuisance measurement items. SUS item 5 loaded on Factor 3 with the MUX Access measurement items. A follow-up EFA with just the SUS items finds the positively- and negatively-phrased items form separate factors, with the negatively-phrased-item factor accounting for the bulk of explained variance (42% vs. 17%). These findings suggest that SUS may not be unidimensional in at least some mobile use contexts, despite routine reports to the contrary in other environments (e.g., [4][6]).

Table 4. EFA of MUX Scales Augmenting SUS

Item / Factor (% of total variance explained by this factor)	1 (33%)	2 (15%)	3 (12%)	4 (6%)
Nuisance1: I felt using a [device] to access [software] would slow me down.	0.820	-0.067	0.216	0.112
Nuisance2: Using [device] to access [software] was inconvenient.	0.850	0.172	0.176	0.002
Nuisance3: Completing tasks using a [device] to access [software] inconvenienced me.	0.827	-0.091	0.190	0.196
Nuisance4: Using a [device] to access [software] made me feel isolated.	0.834	-0.012	0.186	0.171
Nuisance5: Using a [device] to access [software] made me feel disconnected.	0.816	0.140	0.194	0.001
Mobility1: Using a [device] to access [software] would improve my ability to be mobile.	-0.060	0.879	0.002	0.032
Mobility2: I would be able to use a [device] to access [software] on the go.	-0.042	0.828	-0.120	0.130
Mobility3: I think a [device] used to access [software] would be easy to carry with me.	-0.095	0.886	-0.018	-0.030
Mobility4: I feel a [device] used to access [software] would be very portable.	-0.109	0.860	-0.098	0.020
Mobility5: I would be able to take a [device] used to access [software] with me almost everywhere I go.	-0.116	0.830	0.002	-0.093
Access1: A [device] provided a good view of information when accessing [software].	0.081	-0.124	0.791	0.153
Access2: I had no trouble viewing text when using a [device] to access [software].	0.155	-0.096	0.852	0.153
Access3: Clicking on links or buttons was easy to accomplish using a [device] to access [software].	0.041	-0.152	0.829	0.027
Access4: I have no problem entering text when using a [device] to access [software].	0.133	0.061	0.684	0.366
Access5: Using a [device] to access [software] makes it easy to navigate between screens.	0.226	-0.032	0.673	-0.102
SUS01: I think that I would like to use a [device] to access [software] frequently.	-0.121	0.116	0.022	0.675
SUS02: I found using a [device] to access [software] unnecessarily complex.	-0.616	-0.136	0.025	0.280
SUS03: I thought a [device] was easy to use to access [software].	-0.244	-0.070	0.159	0.789
SUS04: I think that I would need the support of a technical person to be able to use a [device] to access [software].	-0.768	-0.157	-0.061	0.154
SUS05: I found the various functions in using a [device] to access [software] were well integrated.	-0.074	0.133	0.581	0.298
SUS06: I thought there was too much inconsistency in using a [device] to access [software].	-0.719	-0.147	0.189	0.211
SUS07: I would imagine that most people would learn to use a [device] to access [software] very quickly.	-0.172	0.019	0.420	0.634
SUS08: I found using a [device] to access [software] very cumbersome to use.	-0.557	-0.228	-0.047	0.014
SUS09: I felt very confident using a [device] to access [software].	-0.332	-0.045	0.239	0.709
SUS10: I needed to learn a lot of things before I could get going with using a [device] to access [software].	-0.817	-0.156	0.057	0.261

* Extraction via Principal Components Analysis; rotation via Varimax with Kaiser Normalization; KMO measure = 0.876; bolding indicates loading values above 0.5

To further inform our understanding of the relationship between MUX and SUS, we performed a Pearson correlation on the four measures (see Table 5). We find SUS to be highly correlated with MUX Nuisance (inversely) and moderately correlated with MUX Mobility.

Table 5. Correlation between MUX Scales and SUS using Stage 3 Data*.

Factor	Mobility	Nuisance	Access	SUS
Mobility	1			
Nuisance	0.094	1		
Access	0.141	-0.348*	1	
SUS	0.271*	-0.527*	.043	1

* Correlation significant at 0.01 level (two-tailed)

3.6. Discriminant Analysis

As a final analysis in Stage 3, we conducted discriminant analysis using the MUX instrument to categorize the device types that the Stage 3 subjects used when completing MUX instrument. We then contrasted this categorization with separate analyses, with MUX augmented by SUS and with SUS only. Our rationale for this action was to assess the relevance and specificity of MUX to the context of mobile user experience which necessarily emphasizes the type of device that is being used. SUS was included in the assessment as an industry-standard benchmark for general usability of computer software and hardware.

We used SPSS discriminant analysis with simultaneous entry of factors and prior probabilities set with all groups equal (see Table 6). MUX Mobility and Access scale group means varied significantly among the three device types we assessed: Smartphones/tablets (these were combined due to low numbers in each separate category), laptop computers, and desktop computers. No differences were observed among device types in the MUX Nuisance scale or the SUS scale.

A discriminant analysis model containing the three MUX scales correctly categorized 63.2% of cases in the Stage 3 data. SUS alone correctly categorized 24.0% of the cases, and MUX augmented with SUS correctly categorized 62.6% of the cases. Based on the ability to discriminate categorical differences among computing devices, these findings suggest that MUX has potential for finer-grained distinctions, e.g., between smart phones or tablets that implement different feature sets.

Table 6. Discriminant Analysis Group Statistics*

Group Means (Standard Deviations) for the Factors				
	MUX Nuis.	MUX Mobil.	MUX Access	SUS
Smartphone / Tablet (n = 26)	2.95 (1.02)	1.85 (0.54)	2.42 (0.76)	56.2 (11.5)
Laptop Computer (n = 111)	2.95 (0.81)	2.39 (0.72)	2.14 (0.57)	54.3 (9.9)
Desktop Comp. (n = 34)	2.88 (1.05)	3.85 (1.11)	1.87 (0.91)	54.1 (11.8)
Test for Equality of the Group Means				
	MUX Nuis.	MUX Mobil.	MUX Access	SUS
Wilks' Lambda / $F_{2/168}$	0.999 / 0.08	0.590 / 58.30	0.946 / 4.76	0.695 / 0.37
Significance	p = 0.93	p < 0.0001	p = 0.010	p = 0.70

* MUX Nuisance, Mobility, and Access values are entered as the average response to items within each scale; SUS values are entered as calculated SUS scores [6]; original coding direction is maintained for items in all scales (see Table 1).

4. Discussion

Overall we are satisfied with initial results in developing and validating the MUX instrument, especially with abilities it has demonstrated in accounting for the preponderance of variance in EFA and in clearly distinguishing categorical differences among computing devices.

Our initial conceptualization that a wide-ranging universe of content would be necessary to evaluate the holistic mobile user experience proved to be mistaken. Instead of the seven content areas we initially proposed, only three dimensions proved to be important in practice. Nonetheless, the pattern of content representation is reassuring to us. While each final dimension drew from multiple content areas, none of these overlapped, i.e., two dimensions drawing from the same content area, and all content areas contributed items to the final instrument.

4.1. Implications for Research

It is an axiom that researchers benefit from having access to effective tools, and our initial assessment of MUX suggests that this instrument can fill a need that is largely unmet. Our approach of holistically measuring specific combinations of mobile software applications and devices provides a unique perspective in evaluating mobile user experiences.

However, we acknowledge that further research will be necessary to clarify the overall contribution of MUX. It will be important, for example, to learn how scores on the MUX dimensions relate to user perceptions, beliefs, and actions. It also will be useful to learn whether MUX can be effectively applied as a second-order construct, thereby integrating Nuisance, Mobility, and Access evaluations into a single score comparable to that produced by SUS. Finally, it will be important for future researchers to explore the relationship of SUS to the MUX dimensions, especially the Nuisance dimension which showed a high level of inverse correlation with SUS and shared substantial factor loadings with SUS in the present study.

4.2. Implications for Practice

Development of MUX was motivated to a large extent by our observation that there is an unmet need for an instrument that can provide *holistic* assessment of mobile user experiences, i.e., by jointly considering the mobile device in addition to the software application being used on it. The practical aspects of this need further motivated our desire to develop an instrument that is short, generalizable, and easy to interpret—characteristics that have helped prior instruments such as SUS to succeed in practice [6].

We argue that MUX fulfills these criteria. While MUX is lengthier than the 10-item SUS, 15 items is still a relatively *short* survey, and we discuss below how the MUX instrument might be further reduced if need arises. Because MUX content is oriented toward common interactions with and perceptions of mobile applications and devices, we anticipate it will be *generalizable* to a wide range of settings. In addition, MUX dimensions (Nuisance, Mobility, and Access) are clear concepts that practitioners should find to be *easy to interpret*. We anticipate these qualities will assist adoption of MUX by user experience practitioners.

4.2.1 Potential for Instrument Reduction. A key feature of SUS is the brevity of this 10-item instrument. Each of the current MUX factor scales contains five measurement items, for a total of 15.

Because each of these factors is measured through reflective items, which are theorized to covary in unison, this number could be reduced further. Bagozzi ([3] p. 271) writes, “With two or more reflective factors, as few as two measures of each factor are required to avoid ambiguity, although three or more measures per factor would be better”, suggesting it would be possible to operationalize a MUX instrument with as few as six measures if this was required by the circumstances

4.2.2 Integrating MUX with SUS. The findings imply two different routes for practical deployment of MUX. First, MUX in total (or just the MUX mobility and access scales) can be deployed to augment SUS, which does not effectively assess aspects of mobile user experience relating to mobility and access. Special attention should be given in this case to assessment of unidimensionality in SUS. We found that negatively- and positively-framed measurement items loaded on separate factors, suggesting our subjects considered “nuisance” aspects of mobile user experience to be distinct from ease-of-use aspects.

A second route is to deploy MUX as a replacement for SUS. This may be especially appropriate in situations where products and services are relatively simple and routinely easy to use.

4.4. Limitations

This research is limited in several ways that may be overcome through additional study. First, purification of initial scales and validation of final scales was based on the responses of undergraduate business students. While these participants form a substantial body of mobile device users, they are not necessarily representative of all users. Future confirmatory research using a broad sample of users can help increase the generalizability of the results.

Second, although our decision to assess use of a university LMS is relevant and appropriate to the student population we studied, this research design choice may limit generalizability to other situations. We recommend that additional research in diverse settings will be necessary before assuming results of the present study are widely generalizable.

Third, a further aspect of our choice to study an LMS is that the SUS scoring for this application (approximately 55 in Stage 3) falls in the low-acceptable range of user experience. We see no reason that this factor obstructs the major conclusions of our research or the validity of the MUX instrument. However, we anticipate the possibility that the MUX Nuisance factor could be less prominent in studies

focusing on software that provide superior user experience to that of the LMS we assessed.

5. Conclusion

The short, generalizable, and easily-interpreted nature of the System Usability Scale has made it one of the most popular survey instruments in industry user experience research [6]. A more recently developed instrument, the MAU [17], offers a rigorously-developed set of scales for assessing characteristics mobile applications. However, neither instrument meets the need for holistic measurement in which specific mobile applications and mobile device combinations are assessed. Although SUS can provide technology independent assessment of usability [6], our results indicate SUS is not sufficiently sensitive to several important characteristics of the mobile context. The MAU is inherently constrained by its exclusive focus on software to be unable to account for distinctions among mobile devices.

The MUX instrument addresses these shortcomings. Our findings show that MUX can augment SUS to provide a more comprehensive picture of mobile user experience. Alternatively, MUX can be used in place of SUS; this may be particularly appropriate when mobility and accessibility are hypothesized to be the primary drivers of user experience.

6. References

- [1] Albert, W., & Tullis, T. (2013). Measuring the user experience: collecting, analyzing, and presenting usability metrics. Morgan Kaufmann.
- [2] Anderson, L. W. (1981). Assessing affective characteristics in the schools. Boston, MA: Allyn and Bacon.
- [3] Bagozzi, R. P. (2011). Measurement and Meaning in Information Systems and Organizational Research: Methodological and Philosophical Foundations. *MIS Quarterly*, 35(2), 261-292.
- [4] Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, 24(6), 574-594.
- [5] Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS Scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3), 114-123.
- [6] Brooke, J. (1996). SUS: A quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194), 4-7.
- [7] Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16, 64-73.
- [8] Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological assessment*, 7(3), 309-319.
- [9] Cronbach, L. J., & Thorndike, R. L. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd edition). Washington, DC: American Council on Education.
- [10] Djasasbi, S., Gomez, W., Kardzhaliyski, G., Liu, T., and Oglesby, F., (2013) "App-Like Mobile Optimization and User Experience" SIGHCI 2013 Proceedings. Paper 12. <http://aisel.aisnet.org/sighci2013/12>
- [11] Djasasbi, S., McAuliffe, D., Gomez, W., Kardzhaliyski, G., Liu, W., & Oglesby, F. (2014). Designing for success: Creating business value with mobile user experience (UX). In *HCI in Business* (pp. 299-306). Springer International Publishing.
- [12] Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of marketing research*, 382-388.
- [13] Gable, R. K., & Wolf, M. B. (1993). *Instrument development in the affective domain* (2nd edition). Boston, MA: Kluwer Academic Publishing.
- [14] Gomez, W., Kardzhaliyski, G., Liu, W., & Oglesby, F. (2012), Improving Mobile Web Experience at Dyn Inc. Major Qualifying Project, Department of Management, Worcester Polytechnic Institute, Worcester, MA
- [15] Grain, H., & Schaper, L. K. (2013, June). Designing a diabetes mobile application with social network support. In *Health Informatics: Digital Health Service Delivery-The Future is Now!: Selected Papers from the 21st Australian National Health Informatics Conference (HIC 2013)* (Vol. 188, p. 58). IOS Press.
- [16] Hair, J., Black, W., Babin, B., & Anderson, R. (2009). *Multivariate data analysis* (7th edition). Upper Saddle River, NJ: Pearson Prentice Hall.
- [17] Hoehle, H., & Venkatesh, V. (2015). Mobile application usability: conceptualization and instrument development. *MIS Quarterly*, 39(2), 435-472.
- [18] Lee, D., Moon, J., Kim, Y. J., & Mun, Y. Y. (2015). Antecedents and consequences of mobile phone usability: Linking simplicity and interactivity to satisfaction, trust, and brand loyalty. *Information & Management*, 52(3), 295-304.
- [19] MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly* 35(2), 293-334.

[20] Pew (2014). Mobile Technology Fact Sheet. Pew Research Center. <http://www.pewinternet.org/fact-sheets/mobile-technology-fact-sheet/>

[21] Pew (2015a). Technology Device Ownership; 2015. Pew Research Center Report. <http://www.pewinternet.org/2015/10/29/technology-device-ownership-2015/>

[22] Pew (2015b). State of the News Media; 2015. Pew Research Center Report. <http://www.journalism.org/2015/04/29/state-of-the-news-media-2015/>

[23] Sauro, J., & Lewis, J. R. (2011). When designing usability questionnaires, does it hurt to be positive? In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 2215-2224). ACM.

[24] Tullis, T. S., & Stetson, J. N. (2004). A comparison of questionnaires for assessing website usability. In Usability Professional Association Conference (pp. 1-12).

[25] Wilson, E. V. & Lankton, N. K. (2012). Some unfortunate consequences of non-randomized, grouped-item survey administration in IS research. Proceedings of the 2012 International Conference on Information Systems (ICIS), Orlando, FL.