Computers in Human Behavior 49 (2015) 208–212

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

The comparison of mobile devices to computers for web-based assessments

Kyle C. Huff *

School of Business, Georgia Gwinnett College, 1000 University Center Lane, Lawrenceville, GA 30043, United States

ARTICLE INFO

Article history:
Available online 17 March 2015

Keywords:
Mobile assessment
Personality
Computer
Usability
Time

ABSTRACT

This research investigated the completing of a web-based personality assessment using smart phones and computers. Data were collected from 47 undergraduate students using a within subjects design. Results indicated that the usability and the time to complete the assessment of a web-based non-optimized questionnaire is significantly different when completed with a smart phone versus a computer. However, there were no significant differences in personality scores.

Published by Elsevier Ltd.

1. Introduction

The impact of internet technology on the science and practice of psychology testing and measurement has been amazing. In the last several years, it seems just about all areas have been influenced in some way. The first wave of the internet technologies to impact this area has led to the shift away from traditional paper based tests, questionnaires, and surveys to web-based versions of those assessments. As the technologies continued to improve, new changes came. One of these changes the completion of psychological assessments by people using mobile devices.

However, there has not been much published research on the use of mobile devices for completing psychological assessments. What research that does exists (Arthur, Doverspike, Muñoz, Taylor, & Carr, 2014; Illingworth, Morelli, Scott, & Boyd, 2014; Morelli, Mahan, & Illingworth, 2014; Sanchez & Branaghan, 2011; Sanchez & Goolsbee, 2010) demonstrates that assessments can be impacted by the type of device that was used to completed it and that this impact is not necessarily consistent across different types of assessments. What little we know about the differences that exist is limited by the lack of experiments investigating issues that may cause differences on assessments across devices. This paper seeks to address this gap in research by investigating the effect on assessments of no optimization of assessment for the device type. Results of this study would help indicate the worst case scenario for individuals or organizations wanting to implement a web-based assessments for their data collection needs as well as serve as a starting point for evaluating improvements in formatting of web-based assessments and optimizing assessments for mobile devices and traditional computers.

Currently, mobile devices have a sizable share of the entire US mobile market. As of January 2014, 58% of American adults have a smart phone and 42% own a tablet computer. In comparison, 78% of American adults reported owning a traditional computer (Pew Research, 2015). Research has shown that both organizations and individuals are interested in using this new technology for the completion of web-based assessments. For example, 14% of companies surveyed reported that their employee selection systems were accessible by mobile device where as 19% of companies surveyed reported that they had candidates for employment requesting the ability to complete the pre-employment assessments on mobile devices. These numbers both increased from the prior year, which were 7% and 9% respectively (Fallaw, Kantrowitz, & Dawson, 2012). Research by Arthur et al. (2014) revealed that 1.9% (almost 70,000 individuals) of a sample of more than 3.5 million completed unproctored assessments on a mobile device. Similarly, research by Illingworth et al. (2014) revealed that just over .8%, or about 8000 applicants of their sample of over 900,000 completed pre-employment assessments on a mobile device whereas Morelli et al. (2014) reported that .3% and .5% of their samples of over 209,000 and 375,000, respectively. If we assume similar trends in all other web-based assessments, such as surveys and web-based research studies, then a sizeable portion of the respondents are likely to attempt to complete the web-based assessments with mobile devices. Numbers such as these have led test developers, survey designers, researchers, and others to try to figure out how to capitalize on this type of device that allows a person to complete an

* Tel.: +1 678 407 5842.
E-mail address: khuff@ggc.edu

http://dx.doi.org/10.1016/j.chb.2015.03.008
0747-5632/Published by Elsevier Ltd.

assessment in anyplace and at any time by optimizing the assessments for the device. However, organizations and individuals who neither possess the resources to acquire those services or the knowledge to develop assessments optimized for mobile devices themselves may not be able to optimize their surveys and questionnaires for these types of devices. Since we know from prior studies that people who completed an assessment on a mobile device may perform differently than people on regular computer, an important question becomes, how much variance are we introducing into web-based assessments when we do not optimize for mobile devices?

A major difference, and probably the most obvious one, between a computer and a mobile device is the interface. While computers use large keyboards, mice or track pads, and have relatively big monitors, mobile devices tend to not have these qualities. This is an important difference since prior research has shown that the user interface can have an impact on computer and web-based test scores (Bridgeman, Lennon, & Jackenthal, 2003; Huff, Cline, & Guynes, 2012; Květon, Jelínek, Vobořil, & Klimusová, 2007; Schroeders and Wilhelm, 2010; Wästlund et al's, 2008). More specifically to this research, very recent research have found differences in mean test scores between mobile devices and traditional computers (Arthur et al., 2014; Illingworth et al., 2014; Morelli et al., 2014; Sanchez & Branaghan, 2011; Sanchez & Goolsbee, 2010). However, the effect sizes of those differences have traditionally been small and very little is known as to why the differences exist at all.

Research by Leeson (2006) summarized research on a number of factors that have been shown to contribute to differences in a number of different types of psychological assessments between different mediums of assessment. In her review of the literature, she found that various demographic characteristics, user-interface legibility, and user-interface interactivity influence the performance of computer based assessments versus paper and pencil assessments. Indeed, these issues could help to explain why a number of studies have found differences in computer and web based assessments (Bridgeman et al., 2003; Huff et al., 2012; Květon et al., 2007; Mead and Drasgow, 1993; Schroeders and Wilhelm, 2010; Wästlund et al., 2008), at either the construct level or the mean level, while others have not (Mead and Drasgow, 1993; Templer & Lange, 2008; Weigold, Weigold, & Russell, 2013) when assessments have been administered in different conditions. This, however, represents a difficulty in that most of the published research on differences between the testing mediums has not taken the information presented by Leeson (2006) into account. In other words, most of this research considers the implantation of the assessments, e.g. number of items per page, response option presentation, etc., that they used in their study equivalent to every other possible implementation of the assessment. In addition, devices such as smart phones, laptops, and PCs, used to complete the assessments have been considered equivalent across groups (Illingworth et al., 2014; Květon et al., 2007). This is of course a major problem in that all we are left with is that differences in the testing medium may cause differences, but we know very little as to how they do it or why.

One of the earliest and most often explanations offered as to why differences between testing mediums exist has to do with scrolling (Bridgeman et al., 2003; Leeson, 2006; Morelli et al., 2014; Sanchez & Branaghan, 2011; Sanchez & Goolsbee, 2010; Wästlund et al., 2008). Basically, test takers who had to scroll in order to complete the test generally performed worse when compared to test takers who did not have to scroll to complete the assessment. This seems particularly true with reading comprehension tests but also can apply to other types of assessments (Sanchez & Branaghan, 2011). Based on these results, we know that optimizing the display of assessments has an impact, but optimization, or the lack there of, would be different for different device types, i.e. optimization for a desktop computer would look different for optimization for a smart phone. Since these differences do exist, and optimization is a degree ranging from completely not optimized to completely optimized, it would be helpful to look at differences in assessments when the level of optimization is held constant. The simplest way to hold optimization constant is to have assessments that are not optimized for either device. This type of assessment would seem to be the type used by individuals or organizations that do not have many resources to invest in web-based assessment. In this way, it can be determined what the differences are, if any, between mobile devices and traditional computers. In addition, this would help answer Leeson's (2006) call for research investigating the presentation format that leads to greatest equivalence between test modes.

One way to approach this problem is to focus on the differences in usability between traditional computers and mobile devices. Usability is defined by the ISO as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" (UsabilityNet, 2006). When assessing usability, Lewis (1995) stated that proper usability evaluations of a computer system would include not just what a person sees on the screen, but also the keyboard, mouse, and other hardware that is being used. The first part of the definition, effectiveness, is traditionally thought of as whether or not the user can use the system to accomplish the task at hand. In regards to the present research, this would become whether a respondent could effectively complete a web-based questionnaire on both a mobile device and a traditional computer. If a user is able to effectively complete the assessment, then there should be an absence of errors on the questionnaire. Prior research has found some mean score differences between mobile and traditional computers for a variety of assessments (Arthur et al., 2014; Illingworth et al., 2014; Morelli et al., 2014) for questionnaires with unknown levels of optimization. However most of these studies reported small to negligible effect sizes. With no optimization, the results could be greater. This brings us to the first hypothesis:

H1. Device type will influence the number of errors on the questionnaire.

The second part of the usability definition is about the efficiency of the system. This is often measured in terms of time to complete the task on the system (Lundby and Mack, 2003; Nielson, 2003; UsabilityNet, 2006). Therefore in this study, time to complete the task would be the time it took a respondent to complete the questionnaire. If the two devices are equal in usability, then there would be no differences in the time it would take to complete the questionnaires. However, since computers and mobile devices are different in the amount and/or quality of information that can be displayed on the screen (Sanchez & Branaghan, 2011; Sanchez & Goolsbee, 2010) and differences in screen size have been associated with differences in time to complete the assessments on optimized web-based assessments (Arthur et al., 2014; Sanchez & Branaghan, 2011), then differences in time to complete the assessment would be expected in this situation as well. This leads to the second hypothesis:

H2. Device type will influence time to complete the assessment.

Finally, usability is also related to user satisfaction (Lundby and Mack, 2003; Nielson, 2003; UsabilityNet, 2006). User satisfaction is generally assessed with a questionnaire. Although user satisfaction is generally lacking in the measurement equivalency research, a study by Huff et al. (2012) did report differences in usability ratings