

# Reliability, validity, and sensitivity of a single-item measure of online store usability

Timo Christophersen, Udo Konrad<sup>\*</sup>

*Christian-Albrechts-University at Kiel, Institute of Psychology, Olshausenstr. 40, 24 098 Kiel, Germany*

Received 23 February 2010; received in revised form 1 October 2010; accepted 25 October 2010

Communicated by G.A. Sundstrom

Available online 29 October 2010

## Abstract

In an experimental study, we examined the validity and reliability of a single-item measure for customers' assessment of online store usability. Each of the 378 participants visited two out of 35 online stores and performed three shopping related tasks. Usability was rated using a single-item and an eight-item measure. In addition to trust in the online store and aesthetics, we also measured the participant's intention to buy. Results from factor analysis and the correction for attenuation formula revealed an adequate reliability of the single-item measure. Positive correlations with both trust and aesthetics supported the convergent validity of the single-item measure for usability. The positive correlation between the single-item and the intention to buy demonstrated the high predictive validity of this measure. Finally, results support the sensitivity of the single-item measure to differentiate between the usability for each online store.

© 2010 Elsevier Ltd. All rights reserved.

*Keywords:* Usability; e-Commerce; Single-item measure; Online store; Scale development

## 1. Introduction

Besides usability, several factors, including trust and aesthetics, determine the success of an e-commerce interface (Schlosser et al., 2006; Hall and Hanna, 2004). As a result, the number of constructs and variables (e.g., user characteristics), which has to be considered in the evaluation of online stores, tends to be high. However, the number of items that can be administered in surveys is usually limited due to time and also space constraints (Braithwaite and Scott, 1991). The possibilities for integrating a large number of items are strongly limited, particularly, in online studies (Evans and Mathur, 2005). Consequently, it is necessary to include scales consisting of only a few items per construct, while still ensuring reliable and valid measurement.

According to the recommendations of numerous scholars (e.g., Churchill, 1979; Nunnally, 1978), researchers should preferably use multiple-item measures in which a latent

construct is represented by a number of manifest items. This approach is in line with the tradition of psychometrics that considers the measurement of psychological constructs as traits, states, or abilities (Kline, 1979). Empirical studies that do not include multiple-item measures are less likely to be accepted for publication in scientific journals (Bergkvist and Rossiter, 2007). In contrast, practitioners often prefer single items to multiple-item measures, as there is less effort involved and costs are reduced. A single-item measure consists of one single item for the assessment of a latent construct. In fact, researchers have successfully applied single-item measures in a broad range of different research areas, such as clinical psychology (McKenzie and Marks, 1999), quality of life research (Cunney and Perri, 1991; Hyland and Sodergran, 1996), consumer research (Bergkvist and Rossiter, 2009; Sarstedt and Wilczynski, 2009), social and personality psychology (Aron et al., 1992), and occupational psychology (Dolbier et al., 2005; Scarpello and Campbell, 1983; Stanton et al., 2002a; Wanous and Hudy, 2001). The aim of this study is to explore the reliability, content validity, predictive validity, and the sensitivity of a single-item measure for usability of online stores.

<sup>\*</sup>Corresponding author. Tel.: +49 431 880 3676;  
fax: +49 431 880 4878.

*E-mail address:* [konradt@psychologie.uni-kiel.de](mailto:konradt@psychologie.uni-kiel.de) (U. Konradt).

## 2. Theoretical and empirical background

### 2.1. The assessment of online store usability

According to the norm DIN EN ISO 9241-11 (1998), usability is defined 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”. Usability has to be considered as one of the main decisive factors in e-commerce interface success (Konradt et al., 2003; Weathers and Makienko, 2006). Consequently, usability experts emphasize that investment in improving this dimension of interface quality will usually “pay off” (Bias and Mayhew, 2005).

Scientists have developed several different methods of examining usability in order to ensure high levels of usability (Nielsen and Mack, 1994). One can generally differentiate between two kinds of evaluation purposes (Bloom et al., 1971). On the one hand, a *summative evaluation* approach focuses on the overall quality of usability. Summative methods allow comparisons between different interfaces without referring to any detailed facet of usability in the form of concrete flaws in design or usage problems (Hartson et al., 2003). On the other hand, contrary to the summative evaluation approach, *formative evaluation* methods are used to identify specific flaws in usability, e.g., bad or incorrect wording, inadequate menu structures, or low colour contrast of text against the background. Thus, formative evaluation methods allow one to draw conclusions pertaining to changes in interface design, interaction processes, and user guidance (Carroll, 1997).

Usability questionnaires can be classified as a set of subjective evaluation methods that primarily provide quantitative data (Dumas, 2003). Scientists have developed different usability scales, which been proven to be valid instruments in both practical contexts of user interface evaluation and scientific studies (cf. Gediga et al., 1999; Hornbæk, 2006; Lewis, 2002). A closer look at usability questionnaires reveals that they vary in many aspects, including the answer format (e.g., Likert-type,<sup>1</sup> open answer field, and ‘no opinion’-options) or the format of the items (e.g., statement, question, and semantic differential). Usability questionnaires also differ regarding the object of evaluation (Kirakowski, 1997). There are instruments that generally focus on the usability of software (e.g., PUTQ by Lin et al., 1997), interactive products (e.g., AttrakDiff by Hassenzahl, 2004), and Websites (e.g., WAMMI by Kirakowski et al., 1998). Furthermore, usability measures for specific domains have also been developed, for example the Usability Questionnaire for Online Stores (UFOS; Konradt et al., 2003).

Usability questionnaires vary concerning the assumed dimensionality of the construct. While some instruments include subscales and thus cover more than one dimension

of usability (e.g., QUIS by Chin et al. (1988); IsoMetrics by Gediga et al. (1999)), others represent a single dimension (PSSUQ by Lewis (2002); SUS by Brooke (1996)). As a consequence, the number of items included in questionnaires strongly varies. For example, the After-Scenario-Questionnaire (ASQ) consists of only three items assessing the user’s global impression of usability (Lewis, 1991). In contrast, the IsoMetrics Questionnaire comprises 90 items, which also cover different facets of usability (Gediga et al., 1999). Few scholars have applied single-items for the assessment of usability (Tractinsky et al., 2000), and unfortunately neglected the subsequent analysis of the single-item measures’ psychometric properties. Consequently, scholars have criticized the single-item approach and questioned its methodology for measuring usability (Hassenzahl, 2004).

### 2.2. Usability as a success factor in e-commerce

Research has shown that the integration of usable design elements in the interfaces has a positive impact on success measures (e.g., Lee and Lee, 2004; Shankar et al., 2003; Weathers and Makienko, 2006). Moreover, flaws in design that are related to usability have a negative impact on success, because the online transaction becomes more inconvenient and more time-consuming (Ceaparu et al., 2004). Empirical studies have shown that usability and related constructs of perceived ease of use and perceived usefulness positively influence the intention to buy (Gefen and Straub, 2000; Konradt et al., 2003; Muthitacharoen et al., 2006).

### 2.3. Application of single-item measures in empirical science

The most important benefit of single-item measures concerns the economy of survey design. Compared to multiple-item measures, a single-item measure shortens surveys and thus reduces the time needed for completing a questionnaire (Nagy, 2002). This results in respondents being more willing to participate in the study (Wanous et al., 1997). Moreover, multiple-item measures often show strong redundancy and include a large number of items that appear to be similar, which leads to participants’ fatigue, frustration, and boredom (Robins et al., 2001), resulting in a large amount of missing information and thus incomplete data (Dolbier et al., 2005). Furthermore, high redundancy leads to lower cognitive participation of respondents and therefore to invalid answers (Stanton et al., 2002b). This is a particular concern in online survey studies because self-administered surveys suffer due to the lack of human interaction (Evans and Mathur, 2005). Also, respondents to multi-item surveys tend to interpret redundancy within a survey as an indication of low face validity.

Due to the shortened time required for both the administration and completion of surveys, single-items may be considered to be more cost-effective than multiple-item measures. For example, the costs for telephone interview

<sup>1</sup>A Likert-type scale is a method that uses standardized response categories to record reactions to items.

Download English Version:

<https://daneshyari.com/en/article/400736>

Download Persian Version:

<https://daneshyari.com/article/400736>

[Daneshyari.com](https://daneshyari.com)