Summary Brief
Revising the Mobile Phone Usability Questionnaire (MPUQ): An Item Response Theory Approach

Young-Sam Ryu, Texas State University—San Marcos
Taewon Suh, Texas State University—San Marcos
Hyun Seok Kim, Georgia Institute of Technology

The Mobile Phone Usability Questionnaire (MPUQ) (Ryu and Smith-Jackson, 2005; 2006) is one of the latest developments in responding to the need for more tailored usability questionnaires. This research attempts to demonstrate that item response theory offers an alternative perspective to the MPUQ. The revised instrument presented will permit the more precise and systematic use of the MPUQ by allowing marketers and researchers to operationalize and utilize it with confidence. As a result, a quite simple four-factor structure of 28 items was obtained for the revised MPUQ.

Introduction

Usability should be a huge issue for any marketer who wants to develop new products to better serve customers, build customer relationships through product usages and customer services, sell products or services online, or use the Web to project their brand or corporate image. It is even more important for electronic consumer products since usability-based models can be more directly tied to system design characteristics, and, more basically, usability has been an important criterion of decision making for consumers, end-users, product designers and software developers for their respective purposes. Particularly, the use of handheld devices such as mobile phones is rapidly growing, and has become an important phenomenon in marketing (i.e., through mobile commerce and communications) (Chu et al., 2005).

Therefore, a distinctive approach and questionnaire regarding usability would be helpful for the evaluation of electronic consumer products, particularly handheld devices such as mobile phones. Although usability has been defined by many researchers in many ways and there have been many efforts to develop usability questionnaires, marketers and product developers are still complaining about deficiencies in existing questionnaires and unfulfilled need for more tailored questionnaires. One of the recent developments against the backdrops in the mobile phone category is the Mobile Phone Usability Questionnaire (MPUQ). The MPUQ consisting of 72 items was proven to be valid and reliable with satisfactory reliability and validity according to CTT in the previous study (Ryu & Smith-Jackson, 2006). However, since the authors only analyzed the scale at the test level as a whole, not at each item level using IRT, it is imperative to further examine whether each of 72 items in the MPUQ is also acceptable according to the IRT approach. Therefore, this paper attempts to demonstrate that item response theory offers an alternative perspective to the MPUQ. The revised instrument presented will permit the more precise and systematic use of the MPUQ by allowing marketers and researchers to operationalize and utilize it with confidence.

Revising The MPUQ

The MPUQ was proven to be a valid and reliable questionnaire with satisfactory reliability and validity according to CTT in the previous study (Ryu & Smith-Jackson, 2006). However, since the previous study analyzed the scale at the test level as a whole, not at each item level using IRT, it is important to further examine whether each of 72 items in MPUQ is also acceptable according to the IRT approach.

First, the polytomous IRT model, GPCM, was utilized to analyze each item of the MPUQ. To distinguish good and bad items, the *item discrimination parameter* and *item information* were estimated using the PARSCALE program. As a result, it was found that 12 items (item # 24, 28, 38, 39, 41, 43, 45, 46, 47, 50, 51, and 58) were unacceptable in terms of the item discriminating power and *item information*. They showed very low levels of *item discrimination parameters* ($\alpha_i < 0.3$); usual range for α_i is between 0 and 2. Because *item information* depends on the item discriminating power, they showed also poor *item information* values. On the other hand, the rest of items (60) showed marginal or high values of *item discrimination parameter* and *item information*. To provide a graphical comparison between bad (#24) and good (#62) items, the item characteristic curves (ICC) and item information curves (IIC) of the two items were displayed in Figure 1. As shown in Figure 1, item 24 has a slow slope of ICC ($\alpha_i = 0.26$) and an almost flat IIC, whereas item 62 shows ideal shapes of ICC ($\alpha_i = 1.16$) and the IIC. Also, in the IIC of item 62, the highest information value (2.4) at the scale score of about -1 indicates that this item measures the trait (i.e., usability) most accurately at the examinee's trait level (θ) of about -1. The 12 unacceptable items with low levels of discriminating power and *information* were eliminated from the MPUQ.

Second, exploratory factor analysis was conducted with the remaining 60 items of the MPUQ which were acceptable according to IRT. Because underlying latent common factors of the MPUQ are themselves intercorrleated with nonzero correlations, oblique rotation (promax) method was chosen.

Although the normalized varimax method of orthogonal rotation has been exceedingly popular for exploratory factor analysis, the orthogonality assumption of common factors is an artifact of the method because the assumption is hardly met in reality. Four factors were extracted from the 60 items considering the *proportion of total variance explained* (95%) and interpretability. Then, based on the *simple structure criteria* that each variable loads on one or as few factors as possible with at least one zero loading, 32 items with cross-loadings were excluded. As a result, a quite simple four-factor structure of 28 item was obtained for the revised MPUQ as shown in Table 1. The inter-factor correlations were also reported in Table 2. They were all positive and moderate.

Finally, the revised MPUQ has satisfactory psychometric properties in both CTT and IRT. The internal consistency-reliabilities (coefficient alphas) for the total scale, factor 1, factor 2, factor 3, and factor 4 are 0.93, 0.90, 0.86, 0.84, and 0.78, respectively. The item-test correlations within each factor ranged from 0.49 to 0.74. The coefficient alpha of 0.70 or higher is the widely accepted level of adequate internal consistency, and the item-test correlation near 1 indicates that the item measures the same thing that is being measured by the test. The *item discrimination parameters* and *item information* for 28 items each item was also provided in Table 3.

Discussion and Conclusions

Most measurement approaches in marketing rest on classical test theory (CTT). Although new alternative measurement approaches have been developed in the area of psychometrics and educational psychology, the marketing literature has tended to focus solely on CTT-based approaches (Singh, 2004). This paper, for that reason, aims to draw attention to one such alternative approach—the item response theory (IRT) approach—that appears particularly promising for addressing contemporary measurement problems. The specific purpose of this paper is to provide a comprehensible review of IRT by developing a measure for a marketing-related construct, mission fulfillment. This paper, thus, has attempted to demonstrate that IRT offers an alternative perspective to the traditional CTT in developing usability questionnaire. IRT deserves the serious attention of marketing researchers. This paper also hopes to initiate marketers' interest in usability questionnaires.

However, like any other model, IRT imposes its own constraints and limitations. It is obvious that IRT are more complex, lack the inherent simplicity of CTT methods. Although several authors have compared and contrasted IRT and CTT approaches, it is important to emphasize that (1) the two methods have different assumptions (e.g., linearity/nonlinearity), (2) the two methods provide different information, and (3) neither method is without flaw.

There can be several implications for marketers in developing and revising a usability questionnaire. First, this instrument can have a substantial and positive effect on evaluating the usability of mobile products for the purpose of making marketing decisions among competing product variations in the end-users market, alternative prototypes during the development process, and evolving versions of a same product during an iterative design process. Second, marketers will be able to take advantage of this instrument or the subscales to expedite their decision making in the comparative evaluation of their mobile products and prototypes. Third, the instrument can serve as a tool for finding diagnostic information to improve specific usability dimensions and related interface elements. Lastly, the more parsimonious instrument can help conduct usability-based research more conveniently.

References

Chu, Y.P., Lee, S.S., Helander M.G., Chong, J., and Lawan, T. (2005), "Injecting usability in m-commerce," *International Journal of Product Development*, 2 (4): 386-397.

Ryu, Y.S., and Smith-Jackson, T.L. (2006), "Reliability and validity of the mobile phone usability questionnaire (MPUQ).," *Journal of Usability Studies*, 2: 39-53.

Ryu, Y.S., and Smith-Jackson, T.L. (2005), "Usability questionnaire for mobile products and content validity," In *Proceedings of HCI International 2005*, Las Vegas.

Singh, Jagdip (2004), "Tackling measurement problems with Item Response Theory: Principles, characteristics, and assessment, with an illustrative example," *Journal of Business Research*, 57, 184-208.