Standard Description Guidelines and Star Ratings are Not Good Measurements for Usability Issues of Mobile Applications That Support Training in Diagnostic Decision Making for Radiologists

Awatef A. Ben Ramadan, MD, MPH1, Martina A. Clarke, MS1, Mia K. Markey, PhD2,3, Michael R. Aro, MD4, Kraig J. Lage, MD4, Kevin L. Ingalls, MD4, Vivek Sindhwani, MD4, Min S. Kim, PhD1,5

1Informatics Institute, University of Missouri, Columbia, MO, 2Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX, 3Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, 4Department of Radiology, University of Missouri School of Medicine, Columbia, MO, 5Department of Health Management and Informatics, University of Missouri School of Medicine, Columbia, MO

Abstract

The description of a mobile application used for training decision making in radiology informs users how to effectively and efficiently operate the app. We investigated statistical relationships of description guidelines created by Microsoft Development Center (MDC) and user star ratings against usability issues for the 48 studied radiology mobile applications. The study found no statistical evidence that description guidelines and star ratings could be used as indicators of usability issues of radiology mobile applications.

Introduction

Many mobile applications have been created with the purpose of improving clinical training in radiology. A user-centered design of these applications may eventually decrease medical errors, cognitive overload, and may increase the quality of healthcare services. Further usability studies of medical mobile applications used in clinical training will assure effective and efficient use of these applications and increase satisfaction among the clinician users. Microsoft Development Center (MDC) recommended ten standard description guidelines to assist in improving the quality of an application’s description section. Star ratings have been widely used to provide a summary of the overall quality of the mobile applications from users’ perspective. While these have been widely used, there was no evidence on the effectiveness of these schemes in the evaluation of health care applications. The objective of this study was to investigate the extent to which the star ratings and description guideline are effective indicators of usability issues for radiology-training mobile applications.

Method

The investigators searched two major online stores (Google Play and iTunes) for radiology-training mobile applications using key terms to identify the maximum number of the marketed radiology applications. The terms used were: radiology, ultrasound, MRI, CT, radiography, nuclear medicine, and fluoroscopy. The investigators chose only applications that strengthened the clinical decision making skills of radiologists, clinicians, and medical students. The applications underwent a multi-step review process by three usability investigators and five radiology experts to identify eligible applications. User reviews were extracted from applications that met the eligibility criteria. Two independent usability investigators (AAB, MAC) extracted usability reviews of the eligible applications and cross examined them using Nielsen’s Heuristic evaluation guidelines and consulted an experienced usability investigator (MSK) to approve the evaluation process and to reach agreement. Out of 381 applications, 107 applications were considered eligible to the study. From these 107 applications, 48 applications contained user reviews, which produced 79 usability issues according to Nielsen’s heuristic evaluation guidelines.

The investigators used nine of the ten MDC’s standard description guidelines and excluded the element, “Get ideas by reviewing descriptions of similar apps in the store”, could not be measured by the investigators and therefore deemed ineligible. Instead, an additional guideline, “the presence of screen shots,” was included to evaluate the description section of the final 48 applications. The guideline, “presence of screen shots,” was included because it
gives a user a preview of what the application looks like and illustrates its key features. The final description guidelines used in this study were:

1. Grab attention in the first few sentences: The introduction of the description should grab a shopper’s attention
2. Make it easy to learn about your application: The description section should describe additional benefits, in-app purchase opportunities, and other details about the application that customers may want to know
3. Use lists and short paragraphs: Breaking up the description content by using short paragraphs and lists
4. Avoid dry language: Use interesting and engaging language
5. Use a length that is just right: The right length is somewhere over 200 words, but well under 3000
6. Be clear about trial periods and in-app purchases: The customers should understand how long they are able to freely use the application and/or which features are limited
7. Use standard capitalization and punctuation
8. Don't forget to check the spelling and grammar
9. Don't include links, or info that belongs elsewhere
10. Presence of screen shots: images of the application that are displayed to the customers

The investigators measured if the selected applications followed these description guidelines suggested by MDC. The investigators also measured the relationship between the usability issues of the studied radiology-training applications and average star ratings.

Results

Of the selected 48 radiology applications, the three standard description guidelines that the applications most frequently followed were: Presence of screen shots (48), Don't forget to check the spelling and grammar (45), and Grab attention in the first few sentences (42). The three standard description guidelines that were the least frequently complied with were: Be clear about trial period and purchases (4), Use a length that is just right (20), and Use lists and short paragraphs (33). The number of description guidelines that were followed per application ranged from 4 to 10 in this study. Ten applications followed five or less of the standard description guidelines and twenty applications complied with nine or more description guidelines. The average number of guidelines followed per application was eight, and the median was 8 per application.

Using the number of usability issues as an objective benchmark, we assessed Pearson correlation test between: (1) the number of standard description guidelines followed and the number of usability issues, and (2) average star ratings and the number of usability issues. We found a very weak negative relationship between the standard description guidelines and the usability issues ($r = -0.07$, $p$ value = 0.64, two-tailed). In addition, the correlation between star ratings and usability issues per applications was also negative and weak ($r = -0.03$, $p$ value = 0.84, two-tailed). However, the results were not statistically significant. This means that neither the completeness of an applications’ description nor the users’ star rating may be effective indicators of how many usability issues may arise in an application.

Conclusion

Less than half (20) of the applications evaluated complied with nine or more of the total ten description guidelines, and the investigators recommended that mobile applications developers should pay closer attention to writing a good description for their applications before marketing. The study’s limitation was the investigators extracted usability issues from user reviews of the studied applications using only Google play and iTunes online stores. Future research should include searches in more online stores and user testing to reveal more usability issues in real practice environment.

References

2. Microsoft Development Center. Your app’s Description. 2014