Making a More Inclusive SUS for Mobile Usable Security and Privacy

Ann-Marie Horcher, Ph.D. *Central Michigan University*

Abstract

Accessibility for the instruments used when measuring usable security and privacy for diverse populations is an increasing concern. Populations not included in measures taken in these studies are subsequently not included in design principles or recommendations.

Increasingly security has moved to mobile platforms, necessitating measures of usable security that reflect the challenges and realities of interacting with that platform. This research addresses the twin challenges of accessibility and security interaction with mobile devices.

This research looks at the delivery of the System Usability Survey (SUS) questionnaire through an instrument that can be tailored by the user to use American Sign Language (ASL), and other languages. In the study the SUS is used across a population that included ASL native speakers, speakers of ASL as a second language, and non-ASL speakers. The study measures user satisfaction with a security input keyboard designed for a mobile platform.

Introduction

The System Usability Survey (SUS) is accepted as a reliable measure of usability in systems including privacy and security [24]. However, security usability is a greater challenge because as security researcher Bruce Schneier says, users don't want to see security at all [25]. Any interaction with security is an inconvenience.

But when looking at security interfaces for which users already have low tolerance, the design of instrument is key to quality in the reported data [21]. The SUS as an instrument has been proven as a valid instrument to measure application usability for mobile phones [13]. In addition, re-using survey question that have been previously validated is a best practice in the use of surveys for measurement. But in measuring security usability, the context in the survey instrument becomes more critical [21]. To collect more valid data, distractors like ads on a webpage, or routing the user to outside websites just to collect the survey data should be minimized.

In security interfaces it is important to require the users to exercise the least amount of cognitive processes possible [20]. Similarly in survey response, the quality of the data collected is influenced by the cognitive load on the subject to comprehend and respond. (Fig. 1)

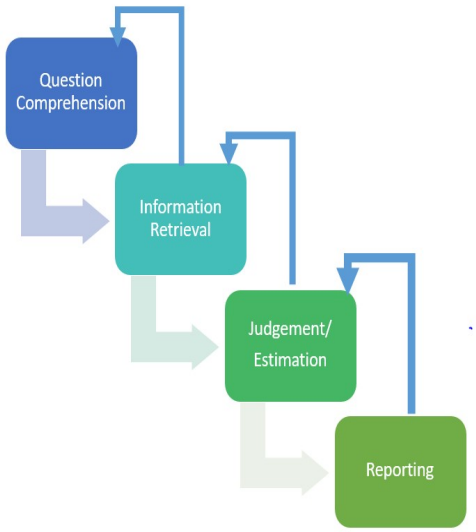

Fig 1: Cognitive Process of Survey Response [4]

USENIX Symposium on Usable Privacy and Security (SOUPS) 2021. August 8-10, 2021..

This cognitive load was a particular concern for usability researchers at Rochester Institute of Technology in their research on deaf and hard of hearing (DHH) subjects in usability tests [1]. Adding DHH representation in evaluation studies required an instrument in their native language, ASL, to reduce the cognitive load on the DHH survey respondents. This research looks to reduce cognitive load in the Question Comprehension and Reporting steps.

Background

One of the most popular, and well-validated, is the Standardized Usability Scale (SUS), a standardized questionnaire created by Brooke (1996) at Digital Equipment Corporation (DEC) as a quick and dirty assessment of usability. Over 500 additional research studies applying SUS have proven that the scale is quick, but not so dirty assessment [23]. This questionnaire is considered the best of open-source norm available [9].

In addition its importance within the United States, ASL is also used worldwide [27]. Other English-speaking countries such as Britain and Australia have their own sign language, as well as France and Japan. ASL is the lingua franca of sign language world, and learned internationally as a second signed language. This increases the value of an ASL-SUS to the research community for recruiting diverse subjects.

Reading in a non-native languages requires a greater cognitive load for any study participant.[1] Half-million people consider ASL their primary communication mode. English literacy among U.S. deaf adults is lower than their hearing peers [28].

English captions are often offered as an alternative communication to ASL [12]. But English and ASL are two distinct languages with separate idioms, grammar, and sentence structure[3]. Native speakers of ASL may have difficulty understanding English, similar to others whose primary language is not English [19]. For example, when communicating emergency information during disasters like hurricanes and national pandemic, ASL interpreters appear on screen to provide the deaf community whose native language is ASL with complete information.

One in seven individuals have bilateral hearing loss. Hearing loss is most prevalent among older adults, where the prevalence rises to three quarters of the population [8]. Since the use of headphones for long period has become prevalent hearing loss at younger ages has accelerated [14]. As the world gets noisier as a whole, hearing loss due to acoustic trauma as well as steady long-term exposure is becoming a public health concern [15].

Finally, mobile devices are becoming the technology platform of choice for most people to interact with throughout their day [22]. More than just a phone, a mobile device can be an emotional and medical lifeline [5, 17] for marginalized communities such as the deaf and hard of hearing (DHH). When assessing the usability of security and privacy, these communities can be excluded simply by the method by which the measurement is made. If, as Peter Drucker is reported to have said, “What you can’t measure, you can’t improve” then what you don’t measure you don’t improve [6].

Research Questions and Hypotheses

The research questions are:
RQ1: Does making the inclusive SUS, make it more usable for non-English native speakers?

RQ2: Is the inclusive SUS more usable for all participants?

This leads to the following hypotheses.

H0: Non-English native speakers do not comprehend better in their native language in SUS.

H1: Non-English native speakers comprehend better in their native language SUS.

H2: The inclusive version of SUS is more usable for all users

Methodology

This research used a design science research (DSR) approach. Design research (DR) is research into or about design. DSR is research using design as a research method or technique [10]. DSR methodology has a series of steps that result in specific outputs (Fig 2). It can be an iterative process, as information from an evaluation influences the design of another element [29].

Design science research solves problems in a more effective and efficient manner by creating an artifact to represent the proposed solution [10]. Because of the nature of many design-research problems, an optimal solution may not always be possible [26]. A designer searches available alternatives until an acceptable alternative is found.

Fig 3: Mobile Security Input Keyboard.

In this research, two artifacts were created. The first was the mobile security input keyboard that the participants in the study were evaluating (Fig. 3). Details about the design choices of that artifact and tasks performed to experience the security interface is not the focus of this research, but can be found in [11].

The second artifact created was an instrument to administer the SUS survey about user satisfaction when interacting with the first artifact. Surveys are a widely accepted method for gathering this measure within both the security and usability communities [2].

The SUS uses the following response format shown in Fig. 4. It uses a 5 point scale to assess user attitudes [16]. The results of raw SUS scores when converted to percentiles yield a letter grade for the application which can be compared to other studies.

Strongly Disagree 1	2	3	4	Strongly Agree 5
O	O	O	O	O

Fig. 4 Standard SUS Reported Likert

Exposing subjects to another software interface to authenticate to collect survey data could influence the user perception of the target interface. This is the change of context discussed earlier, that can be a distractor and impact self-reported data quality. To avoid this the presentation of the survey was designed according to the same usability principles as used for mobile security input keyboard, and matched to the look and feel.

To remove the dependence on correct translation of the Likert scale by any subject, the words and numbers seen in Fig 4 were replaced with five colored buttons. The use of color with green to indicate positive and red to indicate negative, with white as neutral, con-

serves user effort by indicating meaning without requiring the user to read the screen (Fig 5). This follows the Finstad study that uses images to solicit responses to eliminate the need to read the scale [7]. The on-screen targets for responses are the recommended size of 9.2 mm to allow easy acquisition from a touchscreen [18]

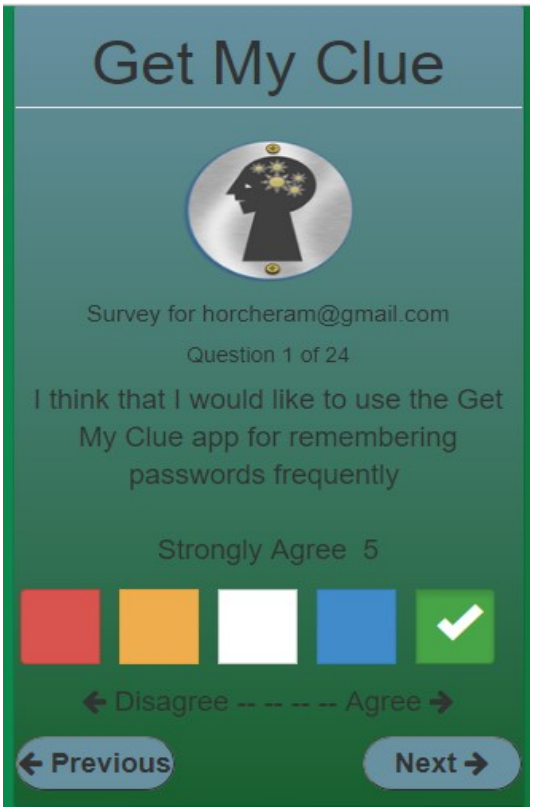

Fig 5. Mobile-optimized Response Format with Color Coding

As the user enters the SUS instrument they are prompted to choose a language, with ASL as an option. In place of the words of the SUS question, a video option appears. The user can click to view the question in ASL. The ASL versions of the SUS survey were created by researchers at the Rochester Institute of Technology [1] by native ASL signers. These recordings have been made available for use by other researchers in hopes of creating more inclusive HCI related studies.

Evaluation

The design choices decreased the cognitive load process (Fig. 1) in the Question Comprehension stage, and the Response Stage. Providing the native language, ASL, of potential DHH subjects reduces the cognitive load of comprehension. Providing color-coded response buttons, which also the correct size to be defined as usable on a mobile phone, reduces the cognitive load of response. The artifact is designed to be stand-alone and reusable in multiple studies.

The single question Net Promoter Score (NPS) and Adjective scales, which are also available in ASL, will be used to rate the usability of the accessible SUS instrument [1] This is used to separate feedback on the two artifacts (Fig.6) .

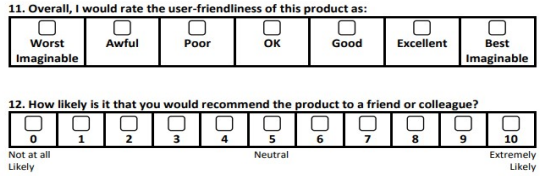

Fig. 6 – Single Question NPS and Adjective Scales

Current Status and Future

The pilot group of participants will be advanced undergraduate students in ASL. The students range between 18-30 years of age and are non-native speakers of ASL. After the pilot, recruitment of native ASL speakers will begin from the deaf community near the university. Recruitment and pilot phases have been delayed due to restrictions imposed by the Covid-19 pandemic.

After the ASL version is evaluated, the option for other foreign languages will be added. Translations of the SUS are also available for Polish, German, French, and Spanish [1].

References

[1] Berke, L., Huenerfauth, M. and Patel, K. Design and Psychometric Evaluation of American Sign Language

Translations of Usability Questionnaires. *ACM Trans. Access. Comput.*, 12, 2 (2019), Article 6.
[2] Bowen, J., Reeves, S. and Schweer, A. A tale of two studies. *Proceedings of the Fourteenth Australasian User Interface Conference - Volume 139* (2013), 81-89.
[3] Bragg, D., Rector, K. and Ladner, R. E. A User-Powered American Sign Language Dictionary. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing* (2015), 1837–1848.
[4] Cannell, C. F., Marquis, K. H. and Laurent, A. *A summary of studies of interviewing methodology*. Department of Health, Education, and Welfare, Public Health Service, Health …, 1976.
[5] Chen, Y.-F. and Katz, J. E. Extending family to school life: College students' use of the mobile phone. *International Journal of Human-Computer Studies*, 67, 2 (2009), 179-191.
[6] Drucker, P. F. *Classic Drucker: essential wisdom of Peter Drucker from the pages of Harvard Business Review*. Harvard Business Press, 2006.
[7] Finstad, K. Response interpolation and scale sensitivity: evidence against 5-point scales. *J. Usability Studies*, 5, 3 (2010), 104-110.
[8] Goman, A. M. and Lin, F. R. Prevalence of hearing loss by severity in the United States. *American journal of public health*, 106, 10 (2016), 1820-1822.
[9] Heeringa, S., West, B. and Berglund, P. *Applied survey data analysis*. Chapman & Hall/CRC, 2010.
[10] Hevner, A. R., March, S. T., Park, J. and Ram, S. Design Science in Information Systems Research. *Management Information Systems Quarterly*, 28, 1 (2004).
[11] Horcher, A.-M. *One size does not fit mobile: Designing usable security input on mobile devices*. City, 2018.
[12] Kawas, S., Karalis, G., Wen, T. and Ladner, R. E. Improving Real-Time Captioning Experiences for Deaf and Hard of Hearing Students. *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility* (2016), 15–23.
[13] Kortum, P. and Sorber, M. Measuring the Usability of Mobile Applications for Phones and Tablets. *International Journal of Human-Computer Interaction*, 31, 8 (2015), 518-529.
[14] Kujawa, S. G. and Liberman, M. C. Acceleration of age-related hearing loss by early noise exposure: evidence of a misspent youth. *Journal of Neuroscience*, 26, 7 (2006), 2115-2123.
[15] Le, T. N., Straatman, L. V., Lea, J. and Westerberg, B. Current insights in noise-induced hearing loss: a literature review of the underlying mechanism, pathophysiology, asymmetry, and management options. *Journal of Otolaryngology-Head & Neck Surgery*, 46, 1 (2017), 1-15.
[16] Likert, R. A technique for the measurement of attitudes. *Archives of Psychology*, 22 140 (1932), 55.
[17] Osmani, V., Balasubramaniam, S. and Botvich, D. Human activity recognition in pervasive health-care: Supporting efficient remote collaboration. *Journal of Network and Computer Applications*, 31, 4 (2008), 628-655.
[18] Parhi, P., Karlson, A. K. and Bederson, B. B. Target size study for one-handed thumb use on small touchscreen devices. *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services* (2006), 203-210.
[19] Patrickis, C. *Closed captions vs. ASL interpreters: Why the latter is a COVID-19 lifeline for the deaf*. 2020.[Online] Available: https://wjla.com/news/local/deaf-community-faces-extreme-lack-of-access-during-covid-19-pandemic.
[20] Rajivan, P., Moriano, P., Kelley, T. and Camp, L. J. Factors in an end user security expertise instrument. *Information & Computer Security*, 25, 2 (2017), 190-205.
[21] Redmiles, E. M., Acar, Y., Fahl, S. and Mazurek, M. L. *A summary of survey methodology best practices for security and privacy researchers*. 2017.
[22] Saha, D. and Mukherjee, A. Pervasive computing: a paradigm for the 21st century. *Computer*, 36, 3 (2003), 25-31.
[23] Sauro, J. *Measuring usability with the system usability scale (SUS)*. City, 2011.
[24] Sauro, J. and Lewis, J. R. When designing usability questionnaires, does it hurt to be positive? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), 2215-2224.
[25] Schneier, B. Security in the real world: How to evaluate security technology. *Computer security journal*, 15 (1999), 1-14.
[26] Simon, H. A. *The Sciences of the Artificial*. MIT Press, 1996.
[27] Thompson, R. L., Vinson, D. P. and Vigliocco, G. The link between form and meaning in American Sign Language: lexical processing effects. *J Exp Psychol Learn Mem Cogn*, 35, 2 (2009), 550-557.
[28] Traxler, C. B. The Stanford Achievement Test: National norming and performance standards for deaf and hard-of-hearing students. *Journal of deaf studies and deaf education*, 5, 4 (2000), 337-348.
[29] Vaishnavi, V. and Kuechler, B. *Design Science Research in Information Systems*. DESRIST, 2004.[Online] Available: http://desrist.org/desrist