Scale Development to Measure Usability of Text-Based CAPTCHA

Samar Swaid

Philander Smith College, 900 Daisy Bates Dr., Little Rock, AR
sswaid@philander.edu

Abstract. Completely Automated Public Turing Tests to Tell Computers and Humans Apart (CAPTCHA) is one of the important security mechanisms to avoid spamming and hacking on websites. This study focuses on identifying the usability dimensions of text-based CAPTCHA based on multi-method approach. Research methods of Think-Aloud protocol, factor analysis and equation modeling are applied. Analytical work shows that usability of CAPTCHA is formed of (i) content; (ii) format; (iii) distortion and (iv) services. The scale is used to evaluate the usability of ReCAPTCHA, JCAPTCHA and GIMBY. The main contribution of the study is constructing a scale, systematically, to quantitatively evaluate the usability of text-based CAPTCHA.

Keywords: Usability, Security, CAPTCHA, Usable Security, Users Perceptions, Design, Factor Analysis.

1 Introduction

Today, CAPTCHA is a widely used security tool to differentiate humans from automated agents by requesting a solution to a problem. Because text-based CAPTCHA is the most widely applied one [12], designers of text-based CAPTCHAs should give more attention not just to create robust ones, but also usable CAPTCHAs [7]. Recent studies underpins the necessity to focus on usability challenges in CAPTCHA [18,12] mainly for three reasons: (i) text-based CAPTCHA are the most common ones utilized by websites and the most ones preferred by web users [12] (ii) users do not necessarily share common cognitive backgrounds in which might affect their perception of the CAPTCHA and its usability [2]; and (iii) CAPTCHA generation is derived by an economic reasons to stop any abuse can be caused by automated agents [14], which make it an essential security tool. To my knowledge, no study has been dedicated to construct a scale to measure the usability of text-based CAPTCHA. This study is designed to cover this gap.

2 Literature Review

It is common knowledge that CAPTCHA should be secure and usable. Unlike the large number of studies on how to develop robust CAPTCHA, it is surprising to find

C. Stephanidis (Ed.): Posters, Part I, HCII 2013, CCIS 373, pp. 159–163, 2013.

that there has been little number of attempts to understand CAPTCHA usability. For example,W3C Working Group listed some concerns to problems posed by CAPTCHA for users with disability and suggested some alternatives to text-based CAPTCHA[15]. Another study recognized the impact of different distortion techniques on the usability of CAPTCHA designed by Microsoft [3]. Others applied usability criteria of accuracy, response time and perceived difficulty to measure usability (e.g., [8]), but no criteria were proposed. Another study examined usability aspects of CAPTCHA in a form of proposed framework of three dimensions of distortion, content and presentation[16]. Although this study forms a serious attempt to understand CAPTCHA usability, its framework cannot be used to quantify usability. Similarly, the study of [14] analyzed different properties of text-based CAPTCHAs that might explain usability.

3 Study and Findings

In this study a multi-method approach is applied. First, think-aloud protocol is used to understand usability measures and dimensions. Second, a scale is constructed to measure text-based CAPTCHAs implementing factor analysis. Finally, equation modeling is implemented to confirm reliability and validity of scale [9].

3.1 Think-Aloud Protocol

Think-aloud protocol is a well documented method in usability studies to understand cognitive processes. This method has been a proven way to highlight features and usability issues that can be improved [7].Six college students, participated in think-aloud study. Their ages ranged from 21 to 29 years old with 50% females and one male student with English as a second language. Participants were provided with six different text-based CAPTCHAs to solve. Input from participants was matched with identified usability criteria found in literature (see Table 1). It is interesting to note 66% of participants proposed issues related the “service” dimension.

3.2 Scale Construction

Several psychometric researchers have proposed different procedural models to develop better instruments. In general, scale construction follow the following pattern: (i) conceptualization; ii) operationalization; and (iii) purification; and (iv)testing psychometric properties of the scale. Dimensions conceptualization was made based on literature [14,16], and think-aloud study. The dimensions then, were opertionalized [6], and pilot tested generating 25 measures. The variable ease of use was adopted from literature [4]. CAPTCHA usability dimensions were treated as formative indicators of the second-order latent construct [5,6,9].

Table 1. Proposed Dimensions of Text-based CAPTCHA

Dimension	Definition	Source
Content	The extent the content embedded in CAPTCHA is understandable to users(e.g., character set, meaningfulness of words, combination of lowercase and uppercase, friendliness of words to non-English speakers, words are inoffensive, number of words)	[14],[16], think-aloud study
Format	The extent CAPTCHA content is formatted and presented to users (e.g., string length, font shape, font color, background color, background pattern, integration with website, uniformity of presentation, text box size)	[14],[16], think-aloud study
Distortion	The extent CAPTCHA content is distorted (e.g., stray lines, collapsing, and level of distortion)	[14], [16], think-aloud
Services	The solvers are provided with a number of services that make it easy to solve CAPTCHA (e.g., number of trials to solve, options to use other forms of CAPTCHA other than text, using appropriate symbols to indicate services	Think-aloud study

All items were measured using seven-point Likert-type scale ranging from (1) strongly disagree to (7) strongly agree. Data was collected from 135 participants with 76% of them females, their ages range from 19-43 years old. Factor analysis was conducted following Hair et al.'s recommendations [9], to identify the number of factors that adequately represent the underlying dimensions of CAPTCHA usability. Factor analysis resulted in identifying four-factor solution of: (i) content; (ii) format; (iii) distortion and (iv) services (see Table 2). Following the recommendation of [5,6,9], equation modeling is performed to understand the links among the dimensions of CAPTCHA usability and the "ease of use" variable. Conducting equation modeling reveals that identified factors relate significantly to overall ease of use. The developed scale demonstrated acceptable psychometric properties based on variety of reliability and validity indices drawn from exploratory factor analysis, and equation modeling, based on recommendations of [5],[6], and [13].

4 CAPTCHA Evaluation

Three of the commonly used CAPTCHAs were evaluated using the developed scale. Total of 26 participants, more than 70% of them are females with ages range from 21-30 years old, evaluated the text-based CAPTCHAs of ReCAPTCHA, JCAPTCHA and GIMBY. ReCAPTCHA scored the highest in terms of usability, then comes JCAPTCHA, to place GIMBY in the third place. Data Show that users find CAPTCHA with appropriate type of distortion that does not confuse them, supported with text of reasonable length and services that enable them to try again or switch to audio-based CAPTCHA make CAPTCHA easy to use by humans.

Table 2. Factor Solution for Usability of Text-based CAPTCHA

	Construct/ Measure	Factor Loading	VIF
	Content:		
1.	Text is clear and easy to solve	0.851	1.20
2.	Text is understandable	0.834	1.23
3.	Text is from dictionary	0.822	1.18
4.	Text is not offensive	0.820	1.33
5.	Text is appropriate for native language speakers	0.811	1.54
6.	Text is of appropriate number of words	0.811	1.70
7.	Text is of appropriate character set	0.809	1.61
	Format:		
1.	Text is presented in an appropriate way	0.901	1.21
2.	Text is of appropriate length	0.883	1.65
3.	Text is of predictable length	0.667	1.44
4.	Text is of appropriate font shapes	0.854	1.271.
5.	Text is of appropriate font sizes	0.848	71
6.	Text is of appropriate color	0.830	1.34
7.	Background is of appropriate color	0.829	1.51
8.	Background is of appropriate level of complexity	0.828	1.89
9.	Text is always presented in standard way	0.822	1.67
	Distortion:		
1.	Distortion applied does not make it hard to solve CAPTCHA	0.877	1.38
2.	Distortion is of appropriate level and do not stop me from solving CAPTCHA	0.856	1.41
3.	Noise added to text is of appropriate level and do not stop me from solving CAPTCHA	0.833	1.34
4.	Distortion type applied is appropriate and do not stop me from solving CAPTCHA	0.803	1.91
	Services:		
1.	CAPTCHA provides me with extra services and features that help me solve CAPTCHA	0.846	1.71
2.	CAPTCHA services improves my ability to solve	0.821	1.93
3.	CAPTCHA services enables me to choose other forms of CAPTCHA such as audio CAPTCHA	0.799	1.89
4.	CAPTCHA is provided with appropriate symbols such as symbol  as an icon to try again	0.767	1.75

5 Conclusion

It has become a standard practice to employ CAPTCHA to differentiate between human users and automated agents. The purpose of this study is to present results of

an in-progress study, with the aim to construct a scale to measure text-based CAPTCHA. The next phase of the study aims into collecting another dataset and confirm the dimensionality of the scale and its properties. This study contributes to efforts related to usable security on the World Wide.

References

1. Baecher, P., Fischlin, F., Gordon, L., Langenberg, R., Lutzow, M., Schroder, D.: CAPTCHAs: The Good, the Bad, and the Ugly, http://www.dominique-schroeder.de/resources/captchas.pdf
2. Belk, M., Fidas, C., Germankos, P., Samaras, G.: Do Cognitive Styles of Users Affect Preference and Performance Related to CAPTCHA Challenges? In: CHI 2012, Austin, Texas, pp. 1487–1492 (2012)
3. Chellapilla, K., Larson, K., Simard, P.Y., Czerwinski, M.: Building Segmentation Based Human-Friendly Human Interaction Proofs (HIPs). In: Baird, H.S., Lopresti, D.P. (eds.) HIP 2005. LNCS, vol. 3517, pp. 1–26. Springer, Heidelberg (2005)
4. Davis, F.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 13(3), 319–340 (1989)
5. Diamantopoulos, A., Siguaw, J.: Formative Versus Reflective Indicators in Organizational Measure Development. JMR 17, 263 (2006)
6. Diamantopoulos, A., Winklhofer, H.: Index Construction with Formative Indicators: An Alternative to Scale Development. JMR 38, 369 (2001)
7. Erricsson, K., Simon, H.: Verbal Reports as Data. Psychol. Rev. 87(3), 215–251 (1980)
8. Fidas, C., Voyiatzis, A., Avouris, N.: On the Necessity of User-Friendly CAPTCHA. In: CHI 2012, pp. 2623–2626 (2012)
9. Hair, J., Anderson, R., Tatham, R., Black, W.: Multivariate Data Analysis, 5th edn. Prentice – Hall, Englewood Cliffs (1998)
10. Hwang, K., Haung, C., You, G.: A Spelling CAPTCHA Systems By Using Click. In: International Symposium on Biometrics and Security Technologies, pp. 1–8 (2012)
11. Nunnally, J.C.: Psychometric theory, 2nd edn. McGraw-Hill, New York (1978)
12. Sauer, G.H.: Towards a Universally Usable CAPTCHA. In: 4th Symposium on Usable Privacy and Security, SOUPS 2008, Pittsburgh, pp. 8–12 (2008)
13. Straub, D., Boudreau, M.-C., Gefen, D.: Validation Guidelines for is Positivist Research. Communications of the Association of IS 13, 380–427 (2004)
14. Sutherland, C.: Usability and Security of Text-Based CAPTCHAs. In: UMM CSci Senior Seminar Conference, Morris, MN (2012)
15. W3C Working Group.: Inaccessibility of CAPTCHA-Alternatives to Visual Turing Test on the Web, http://www.w3.org/TR/turingtest
16. Yan, J., El-Ahmad, A.: Usability of CAPTCHA or Usability Issues in CAPTCHA Design. In: Proceedings of the 4th Symposium on Usable Privacy and Security, pp. 44–52. ACM, New York (2008)