

The Effect of CAPTCHA on User Experience among Users with and without Learning Disabilities

Ruti Gafni

Idan Nagar

The Academic College of Tel Aviv-Yaffo The Academic College of Tel Aviv-Yaffo

rutigafn@mta.ac.il

idan.nagar@gmail.com

Abstract

CAPTCHA is a challenge-response test implemented in many Internet Websites to distinguish between malicious automatic software ("bot") and real users. However, CAPTCHA emphasizes the gaps between accessibility and security on the Internet, as it poses an obstacle and a significant challenge for the learning-impaired in the reading and comprehension of what is presented in the test. Attempts are being made to develop CAPTCHA tests of various types in order to address accessibility and security issues. The objective of this study is to investigate the differences between various CAPTCHA tests and examine how they affect user experience among populations with and without learning disabilities. A questionnaire accompanied by experiencing five different tests was performed among 212 users, 60 of them with learning disabilities. Response times for each test and rate of success were collected automatically. Findings suggest that users with learning disabilities have more difficulties in solving the tests, especially those with distorted texts, have more negative attitudes towards the CAPTCHA tests, but the response time has no statistical difference from users without learning disabilities. With this information, we can assess how to apply solutions that are more suitable for many users and especially for this sensitive population.

Keywords: CAPTCHA, Cyber Security, User Experience, Learning Disabilities.

Introduction

One of the main issues of cyber security deals with the question whether the agent trying to reach a Website is a real person or a malicious automated program ("bot"). The most common solution to decide is called CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart), invented in 2000, and still used (Singh & Pal, 2014). The test performs a challenge-response authentication process, presenting a challenge to the user, and the right to access the Website is given if solved. The user must use cognitive abilities, which are impossible to machines (Saini & Bala, 2013). CAPTCHA tests are an efficient approach for this goal, but are difficult for users with learning disabilities (LD), who have difficulties in reading, understanding, and performing the tests. CAPTCHA tests must be, on the one hand, very easy for the user in order to pass, and, on the other hand, very difficult for the bots to pass.

Cyber threats became abundant, and the attempts to reach computers by unauthorized agents are growing. Therefore, CAPTCHA tests are widely used in all kind of Websites, millions of times a day (Wikipedia, 2015).

Types of CAPTCHA tests

Text-based CAPTCHA tests – Are the most used tests (Figure 1), called "reCAPTCHA", consisting of a sequence of numbers and letters, twisted and shown in a distorted manner. The user needs to enter the exact sequence into a text box.



Figure 1. Text-based CAPTCHA

In 2013, reCAPTCHA began implementing behavioral analysis of the browser's interactions with CAPTCHA. This analysis (Figure 2), occurs before displaying the CAPTCHA and presents a more difficult test in cases there are reasons to think the user is a bot. From 2014 this mechanism, called "No CAPTCHA reCAPTCHA", started to be used in most of Google services (Wikipedia, 2015).

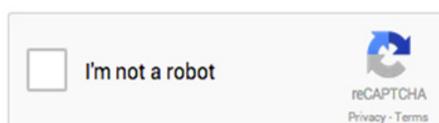


Figure 2. No CAPTCHA reCAPTCHA

Arithmetic operation based CAPTCHA tests – contain a very basic arithmetic operation, (Figure 3), which can be performed by almost every human being. The user needs to enter the result of the operation into a text box.

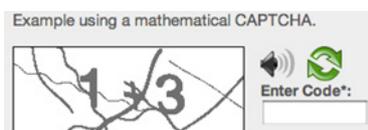


Figure 3. Arithmetic operation CAPTCHA

Picture based CAPTCHA tests –a number of pictures are shown (Figure 4), with a simple question. The user must identify the pictures, and select those which are the correct answer. There is no need to write any text. There are some variations of the picture-based CAPTCHA, for example, sliding distorted pictures to arrange them.



Figure 4. Picture based CAPTCHA

Game based CAPTCHA tests – include puzzles or interactive games (Mohamed, Sachdeva, Georgescu, Gao, Saxena, Zhang, & Chen, 2013). In Figure 5, an example of a game can be seen, where the user needs to drag the food to the baby's mouth.



Figure 5. Game based CAPTCHA

Many studies have been conducted on CAPTCHA, focusing on the security analysis (Azad, 2013), on differences between types of tests (Singh & Pal, 2014; Foley, 2012), and examining the combination between security and user friendliness (Gossweiler, Kamvar & Baluja, 2009). Only few focused on the user experience (Tangmanee & Sujarit-apirak, 2013; Fidas, Voyiatzis & Avour, 2011). This research deals with the influence of CAPTCHA tests among users having LD, examining user experience, actual performance and success.

Theoretical Background

The different types of CAPTCHA have some disadvantages (Singh & Pal, 2014). In the text-based type, the user experiences difficulties deciphering the characters due to the blurring. Furthermore, a "bot" has the ability to recognize the character sequence using an Optical Character Recognition (OCR) software. Mostly, CAPTCHA tests are based on Latin letters, causing difficulties to users who are not used to Latin letters (Tangmanee & Sujarit-Apirak, 2013; Fidas et al., 2011; Bursztein, Bethard, Fabry, Mitchell & Jurafski, 2010). In the picture-based test, visually impaired people may perceive difficulties in identifying the right picture. Hsu and Lee (2011) found that older users show greater difficulty in passing a text-based CAPTCHA in comparison to younger users, and that even not disabled users may encounter difficulties recognizing and understanding the distorted characters.

CAPTCHA tests must: be easy for the user to pass successfully; and be sufficiently difficult to prevent the "bots" from passing them. However, most studies focus primarily on how to make the tests more difficult for bots, in response to the growing number of security threats. Azad (2013) suggested to raise security of text-based CAPTCHA by adding "noise", increasing the level of distortion of the characters and aligning the characters more closely; yet, this would make it more difficult for users to identify the characters, causing more mistakes.

Picture-based CAPTCHA tests require recognition and selection of images with a similar or an exceptional meaning, out of a sequence of images and may cause confusion, as the images can sometimes be interpreted as having different meanings (Ahn, Kim & Kim, 2013). Some studies propose variations of the tests, in order to simplify them, but these propositions are not widely used yet (Gossweiler et al., 2009).

Bevan's (2009) study describes the user experience as including emotions, beliefs, preferences, conceptions, psychological and physical reactions, behaviours, and achievements occurring before, during, and after usage. According to Onwudebelu, Sanjo, Obi and Alaba (2010), the use of CAPTCHA tests is a nuisance. Some users feel threatened by this tests, irritated and frustrated, as they are unable to understand the need for it. Other, reported the text displayed is unclear, and they struggle to solve it. Most, reported the tests slow them down and interfere with their activity on Websites.

Aside from accessibility difficulties, frustration, and poor user experience (Ghazarian, 2014); user abandonment and a drop off Internet conversion rates are additional consequences of CAPTCHA tests suffering from user-unfriendliness (Mujumdar, D., Poliseti, S., 2011).

CAPTCHA tests pose an obstacle for LD users, in reading and comprehending the tests. According to Foley (2012), the most common LD is dyslexia. Dyslexia, (from Greek, dys=difficulty + lexis=words) known as reading disorder, is a LD characterized by trouble reading despite normal intelligence. Problems may include spelling or writing words, reading quickly and understanding what one reads. The cause of dyslexia is believed to involve both genetic and environmental factors (Wikipedia, 2015). It is estimated that 8-10% of the

population suffers from dyslexia. It is even estimated that one-third of Internet users suffer from certain symptoms of dyslexia (McCarthy & Swierenga, 2010).

There are many recommendations for the development of appropriate and friendly Websites to be accessible towards these users (McCarthy & Swierenga, 2010; Foley, 2012), including guidelines (W3C, 2008) and clear principles (Matej, 2013).

Research Questions and Hypotheses

The main research questions are:

RQ1: Are there any differences in the response time of users with or without LD?

H1: The performance time of users with LD will be longer in text and arithmetic based tests

RQ2: Are there any differences in the success rate of users with or without LD?

H2.1: The success rate of users without LD will be greater.

H2.2: Users with LD will succeed better in tests which do not include letters.

RQ3: Are there any differences in the attitude and user experience of users with or without LD?

H3.1: Attitude and user experience of users with LD will be more negative.

H3.2: Users with LD will report better experience in tests which do not include letters.

Methodology

In order to collect the data, an experiment and a questionnaire were created, which included the actual usage of five different types of CAPTCHA tests. The respondents had to perform the tests and rate their experience using them (1 – 'Strongly disagree', 5 – 'Strongly agree').

Meanwhile, a hidden script, automatically accumulated the data about the success or failure of the respondent in using each test, and the time it took to complete it. The combination of both research methods was based on the study conducted by Abrich, Berbenetz and Thrope (2011), which defined the quality of user experience on whether the user was correct or not when taking a test, as well as on the level of test difficulty the user reported.

The questionnaire was composed of 11 demographic and general information questions, and other 10 questions, answered once for each test after experiencing them.

The types of CAPTCHA that were examined were the five mentioned in the Introduction of this paper.

The questions and the experiment were constructed using a plug-in based on WordPress, a free web software. The questionnaire was delivered using social media like Facebook, and through sites and forums of LD.

The answers were collected during one week (December 2014), and then gathered and analyzed using IBM® SPSS® Statistics.

Results

There were 212 Israeli respondents, 60 reporting having LD or thinking they have but not diagnosed yet (28%) and 152 without LD (72%). The frequency of Internet usage was similar and high in the two groups (4.7 in LD and 4.88 in non-LD).

Table 1 summarizes the familiarity of the users with each CAPTCHA test. The mean value and standard deviation are shown. As seen, the text-based CAPTCHA is the most familiar to the users, and the arithmetic operation based one is also known. Most of the users are not so familiar with the other three types.

Table 1. Familiarity with the CAPTCHA tests (1 – 'Never', 5 – 'Very often')

| CAPTCHA type | LD users (n=60) | Non-LD users (n=152) | Total (n=212) |
|-------------------------------|-----------------|----------------------|----------------|
| 1. Text based | 3.88 (1.01) | 3.92 (0.94) | 3.91 (0.96) |
| 2. Arithmetic operation based | 2.48 (1.25) | 2.19 (1.26) | 2.28 (1.26) |
| 3. Picture based | 1.03 (0.18) | 1.2 (0.63) | 1.15 (0.55) |
| 4. Game based | 1.13 (0.5) | 1.1 (0.43) | 1.11 (0.45) |
| 5. "NO CAPTCHA" based | 1.33 (0.88) | 1.69 (1.13) | 1.59 (1.08) |

After performing each of the CAPTCHA tests, the participants were asked to rate their agreement with the statements in Appendix A. Table 2 specifies the items mean and standard deviation and the reliability, measured by Cronbach's alpha.

Table 2. User experience constructs definition

| Item | Mean (SD) | Cronbach's alpha | Construct | Mean (SD) |
|------|-------------|------------------|----------------------------------|-------------|
| 1.1 | 3.94 (1.28) | .802 | Frustration | 3.86 (1.23) |
| 1.2 | 3.78 (1.40) | | | |
| 1.3 | 2.94 (1.49) | Separated | Enjoyable | 2.94 (1.49) |
| 2.1 | 3.07 (1.48) | .954 | Readiness for future use | 3.05 (1.43) |
| 2.2 | 3.09 (1.48) | | | |
| 2.3 | 3.00 (1.52) | | | |
| 3.1 | 2.89 (1.34) | Deleted | | |
| 3.2 | 4.29 (1.07) | .904 | Unwillingness to use the Website | 4.28 (1.03) |
| 3.3 | 4.26 (1.09) | | | |

The results (Table 3) indicate that there is a significant difference between LD and non-LD participants in two user experience constructs: 'Frustrating' and 'Unwillingness to use the Website'.

Table 3. Statistical results for User experience

| User Experience construct | LD User | N | Mean | SD | t | df | Sig. (2-tailed) |
|----------------------------------|---------|-----|------|------|-------|--------|-----------------|
| Frustrating | No LD | 760 | 3.96 | 1.19 | 3.961 | 508.79 | .000 |
| | LD | 300 | 3.62 | 1.29 | | | |
| Enjoyable | No LD | 760 | 2.95 | 1.49 | .412 | 1058 | .680 |
| | LD | 300 | 2.91 | 1.50 | | | |
| Readiness for future use | No LD | 760 | 3.07 | 1.43 | .627 | 1058 | .531 |
| | LD | 300 | 3.01 | 1.43 | | | |
| Unwillingness to use the Website | No LD | 760 | 4.34 | 0.96 | 3.123 | 463.18 | .002 |
| | LD | 300 | 4.10 | 1.18 | | | |

Table 4 summarizes the findings of the user experience of each of the CAPTCHA tests, according to the constructs defined. The gamed-based was found the most enjoyable for LD users, while the text-based was the least enjoyable and most frustrating.

Table 4. Comparison of means and sd - user experience for CAPTCHA types

| CAPTCHA type | LD User | Post-test experience | | Readiness for future use | Unwillingness to use the Website |
|-------------------------------|---------|----------------------|-------------|--------------------------|----------------------------------|
| | | Frustrating | Enjoyable | | |
| 1. Text based | No LD | 3.56 (1.10) | 2.05 (1.17) | 2.42 (1.12) | 4.30 (0.87) |
| | LD | 3.15 (1.17) | 1.85 (1.15) | 2.09 (1.04) | 4.98 (1.03) |
| 2. Arithmetic operation based | No LD | 4.52 (1.03) | 3.01 (1.47) | 3.30 (1.42) | 4.45 (0.85) |
| | LD | 4.69 (1.20) | 2.72 (1.28) | 3.07 (1.27) | 4.11 (1.26) |
| 3. Picture based | No LD | 3.94 (1.15) | 3.34 (1.42) | 3.31(1.44) | 4.37 (0.95) |
| | LD | 3.68 (1.36) | 3.22 (1.52) | 3.26 (1.51) | 4.17 (1.19) |
| 4. Game based | No LD | 3.86 (1.21) | 3.24 (1.38) | 3.07 (1.37) | 4.30 (0.97) |
| | LD | 3.68 (1.23) | 3.55 (1.47) | 3.19 (1.39) | 4.29 (1.13) |
| 5. "NO CAPTCHA" based | No LD | 4.20 (1.31) | 3.11 (1.63) | 3.24 (1.56) | 4.28 (1.13) |
| | LD | 3.88 (1.42) | 3.20 (1.49) | 3.44 (1.53) | 3.98 (1.27) |

Statistical differences between the LD-users and the Non-LD users were found in the 'Frustration' construct in the text-based ($t = -2.36$, $df = 210$, $sig = 0.019$) and arithmetic-operation-based ($t = -3.294$, $df = 210$, $sig = 0.001$) tests.

As for response time, surprisingly, there were not differences between LD and non-LD participants for any of the CAPTCHA tests (Table 5).

Table 5. Comparison of the CAPTCHA types – Response Time

| CAPTCHA type | Response Time (in seconds) | | | | t-test | Sig (2-tailed) |
|-------------------------------|----------------------------|----------------------|-----------------|----------------------|--------|----------------|
| | Mean (SD) | | median | | | |
| | LD users (n=60) | Non-LD users (n=152) | LD users (n=60) | Non-LD users (n=152) | | |
| 1. Text based | 22.25 (12.73) | 19.36 (13.69) | 18 | 16 | -1.41 | 0.16 |
| 2. Arithmetic operation based | 11.4 (10.85) | 9.16 (8.3) | 9 | 8 | -1.61 | 0.11 |
| 3. Picture based | 26.78 (29.53) | 24.34 (23.55) | 18 | 18.5 | -0.63 | 0.53 |
| 4. Game based | 21.07 (13.05) | 20.34 (16.33) | 17 | 17 | -0.31 | 0.76 |
| 5. "NO CAPTCHA" based | 19.32 (22.74) | 18.97 (23.6) | 11.5 | 13 | -0.10 | 0.92 |

The success rate was calculated for each test. Table 6 presents the results and the statistical comparison, using the Pearson Chi-square test ($p < 0.05$), indicating that there is a significant dependence between the test success and the existence of LD only for text-based CAPTCHA, bringing the LD users to fail more often. The arithmetic test is more difficult for all populations.

Table 6. Comparison of the CAPTCHA types – Success Rates

| CAPTCHA type | Success Rate | LD users (n=60) | Non-LD users (n=152) | χ^2 value | df | Sig. (2-tailed) |
|-------------------------------|--------------|-----------------|----------------------|----------------|----|-----------------|
| 1. Text based | % | 58.33% | 74.34% | 5.23 | 1 | 0.022 |
| | Pass | 35 | 113 | | | |
| | Fail | 25 | 39 | | | |
| 2. Arithmetic operation based | % | 48.33% | 50% | 0.048 | 1 | 0.83 |
| | Pass | 29 | 76 | | | |
| | Fail | 31 | 76 | | | |
| 3. Picture based | % | 68.33% | 73.03% | 0.47 | 1 | 0.49 |
| | Pass | 41 | 111 | | | |
| | Fail | 19 | 41 | | | |
| 4. Game based | % | 83.33% | 82.89% | 0.01 | 1 | 0.96 |
| | Pass | 50 | 126 | | | |
| | Fail | 10 | 26 | | | |
| 5. "NO CAPTCHA" based | % | 86.67% | 89.47% | 0.34 | 1 | 0.56 |
| | Pass | 52 | 136 | | | |
| | Fail | 8 | 16 | | | |

Discussion

Findings suggest that users with LD have more difficulties in solving the tests, especially those with distorted texts and have more negative attitudes towards the CAPTCHA tests. Surprisingly, there was no significant difference found in response times between LD and non-LD users in any of the five tests types, thus, rejecting H1. The reason H1 was not supported, may be the fact that CAPTCHA tests became so common, that the users are getting used to the twisted letters. However, a significant difference in the test success between the two kind of users was found for the text-based CAPTCHA. Thus, H2.1 was rejected, supporting the position of Hsu & Lee (2011), who claim that even the non-impaired population might encounter difficulties with CAPTCHA; and H2.2 was accepted, meaning that users with LD have difficulties in reading the distorted texts, supporting Foley's study (2012), but have no problems performing other tasks. Perhaps, if taking more time to resolve the test, the success rate may be better for the LD users.

A significant difference in the user-experience attitude was found for 'Frustrating' and 'Unwillingness to use the Website' between LD users and non LD users, where LD users had a more negative attitude towards the tests, accepting H3.1. The most negative attitude of the LD users was found for the text-based and arithmetic-based CAPTCHA, supporting H3.2.

The main conclusion of this research is that all five types of tests influenced user experience in a certain level, from frustration to enjoyment. Users with LD found it more difficult to succeed in the text-based CAPTCHA, and there was a significant difference found between LD and non-LD users in most user experiences constructs examined.

The main limitations of this study concern about LD: (1) the small number of LD participants (60 with LD, and 152 without). A larger number of LD users might have provided a better representative sample. (2) the definition of users with LD – some were diagnosed, some were self reported but not diagnosed and others may be not aware of LD.

References

- Abrich, R., Berbenetz, V., & Thorpe, M. (2011). Distinguishing between Humans and Robots on the Web. Retrieved December 1, 2014, from richardabrigh.com: http://richardabrigh.com/abrigh-berbenetz-thorpe_thecaptchaexperiment.pdf
- Ahn, Y., Kim, N., & Kim, Y. S. (2013). A User-friendly Image-Text Fusion CAPTCHA for Secure Web Services. In *Proceedings of International Conference on Information Integration and Web-based Applications & Services*, 550-554.
- Azad, K. J. (2013). Captcha: Attacks and weaknesses against OCR technology. *Global Journal of Computer Science and Technology*, 13(3).
- Bevan, N. (2009). What is the difference between the purpose of usability and user experience evaluation methods. *UXEM'09 Workshop*, Uppsala, Sweden.
- Bursztein, E., Bethard, S., Fabry, C., Mitchell, J. C., & Jurafsky, D. (2010). How good are humans at solving CAPTCHAs? A large scale evaluation. In *Security and Privacy (SP), Symposium on Security and Privacy IEEE*, 399-413.
- Fidas, C., Voyiatzis, A., & Avouris, N. . (2011). On the Necessity of User-friendly CAPTCHA. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2623-2626.
- Foley, A. (2012, Jul). Biometric alternatives to CAPTCHA: Exploring accessible interface options. Masters Dissertation, Dublin Institute of Technology, Ireland.
- Ghazarian, A. (2014). CAPTCHAs' Effect on UX and How to Fix It. Retrieved December 15, 2014, from DesignModo: <http://designmodo.com/ux-captcha-effect>
- Gossweiler, R., Kamvar, M., & Baluja, S. (2009). What's up CAPTCHA?: A CAPTCHA based on image orientation. In *Proceedings of the 18th international conference on WWW*, 841-850.
- Hsu, C.-H., & Lee, Y.-L. (2011). Effects of age groups and distortion types on text-based CAPTCHA tasks. In Jacko, J. J., *Human-computer interaction. Users and Applications* (pp. 435-455). New York: Springer-Verlag Berlin Heidelberg publications.
- Matej, S. (2013). *Strict Standards: Introduction to CAPTCHA Accessibility*. Retrieved December 1, 2014, from CAPTCHA.com: <http://captcha.com/captcha-accessibility.html>
- McCarthy, J. E., & Swierenga, S. J. (2010). What we know about dyslexia and web accessibility: a research review. *Universal Access in the Information Society*, 9, 147-152.
- Mohamed, M., Sachdeva, N., Georgescu, M., Gao, S., Saxena, N., Zhang, C., & Chen, W. B. (2013). *Three-Way Dissection of a Game-CAPTCHA: Automated Attacks, Relay Attacks, and Usability*. Cornell University. New York: arXiv:1310.1540.
- Mujumdar, D., & Polisetti, S. (2011). *A platform to monetize usable & secure CAPTCHAs for desktop and mobile devices (PICATCHA)*. Retrieved Dec 20, 2014, from Kevinwarnock: http://www.kevinwarnock.com/wp-content/uploads/2012/05/picatcha_mims_final_report_summary_0.pdf
- Onwudebelu, U., Sanjo, F., Obi, N. C., & Alaba, O. B. (2010). CAPTCHA Malaise: Users suffer Consequences of the Anti-spam Technology while the Spammers Adapt. In *Proceeding of the 1st International Conference and Workshop on Software Engineering and Intelligent Systems*, 113-125.
- Saini, B. S. and Bala, A. (2013). A Review of Bot Protection using CAPTCHA for Web Security, *IOSR Journal of Computer Engineering*, 6, 36-42.
- Singh, P. V., & Pal, P. (2014). Survey of Different Types of CAPTCHA. *International Journal of Computer Science and Information Technologies*, 5(2), 2242-2245.
- Tangmanee, C., & Sujarit-apirak, P. (2013). Attitudes towards CAPTCHA: A Survey of Thai Internet Users. *The Journal of Global Business Management*, 9, 29-41.

W3C. (2008). *Web Content Accessibility Guidelines (WCAG) 2.0*. Retrieved December 10, 2014, from World Wide Web Consortium: <http://www.w3.org/TR/WCAG20>

Wikipedia. (2015). *Dyslexia*. Retrieved October 3, 2015, from Wikipedia: <http://en.wikipedia.org/wiki/Dyslexia>

Wikipedia. (2015). *ReCAPTCHA*. Retrieved January 2, 2015, from Wikipedia: <http://en.wikipedia.org/wiki/ReCAPTCHA>

Appendix A

1. The experience after the CAPTCHA test –

"How would you describe your feeling after taking this CAPTCHA test?"

Item 1.1: *"Frustrating – the test was **difficult and unclear**"*

Item 1.2: *"A waste of time – It took **too much time** to do it"*

Item 1.3: *"Comfortable/Enjoyable – It was a **nice** test"*

2. The position towards this CAPTCHA test – "If you have to take this CAPTCHA test again, what will your position be towards this test?"

Item 2.1: *"I would rather do it again – since it's **easy and clear**"*

Item 2.2: *"I would rather do it again – since it takes only a **short time** to do it"*

Item 2.3: *"I would rather do it again – since it's **nice and comfortable**"*

3. The position towards the Website - "If you have to take this CAPTCHA test again, what will your position be towards the Website that displays this test?"

Item 3.1: *"My position will be **positive**"*

Item 3.2: *"I will not cooperate with this site – **I will leave this site immediately**"*

Item 3.3: *"I will not cooperate with this site – **I will not perform actions** such as registration, buying, etc."*