

What Happens When Evaluating Social Media's Usability?

Virginica Rusu¹, Cristian Rusu², Daniela Quiñones²(✉),
Silvana Roncagliolo², and César A. Collazos³

¹ Universidad de Playa Ancha de Ciencias de la Educación, Valparaíso, Chile
virginica.rusu@upla.cl

² Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile
{cristian.rusu, silvana.roncagliolo}@pucv.cl,
danielacqo@gmail.com

³ Universidad del Cauca, Popayán, Colombia
ccollazo@unicauca.edu.co

Abstract. The paper presents a study on the perception of 16 evaluators over a set of usability heuristics for smartphones - SMASH. All participants were asked to perform a heuristic evaluation of the mobile version of Facebook. Later on a survey was conducted based on the standard questionnaire that we developed. Evaluators perceive SMASH as an appropriate instrument to evaluate social media's usability, and they intent to use it in future evaluations.

Keywords: Usability · Heuristic evaluation · Usability heuristics · Social media · Facebook

1 Introduction

The ISO 9241-210 defines usability as “the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [1]. Lewis identifies two approaches on evaluating usability: (1) summative, “measurement-based usability”, and (2) formative, “diagnostic usability” [2].

Heuristic evaluation is one of the most popular usability evaluation methods [3]. Generic or specific heuristics may be used. Generic heuristics are familiar to evaluators and therefore easy to apply, but they can miss specific usability issues. Specific heuristics can detect relevant domain related usability issues.

The paper presents a study on the perception of 16 evaluators over a set of usability heuristics for smartphones - SMASH [4]. All participants were asked to perform a heuristic evaluation of the mobile version of Facebook. Later on a survey was conducted based on a standard questionnaire that we developed. Section 2 briefly reviews the concepts of usability and its evaluation. Section 3 describes SMASH, the set of usability heuristics that we used. Section 4 presents the survey on evaluators' perception after evaluating the mobile version of Facebook. Section 5 highlights conclusions and future work.

2 Usability and Usability Evaluation

Known for decades, the usability concept is still evolving. A widely accepted usability definition was proposed by the ISO 9241 standard back in 1998 [5]. The ISO 9241 standard was updated in 2010 [1], but a new revision started briefly after, in 2011 [6].

There is no general agreement on either usability definition or its dimensions, but several aspects are recurrent in all definitions: effectiveness, efficiency, satisfaction, and the context of use. The ISO 9241 current approach relates usability to user and business requirements: effectiveness means success in achieving goals, efficiency means not wasting time, and satisfaction means willingness to use the system [6].

Usability evaluation does not limit to measuring effectiveness, efficiency and satisfaction. Several classifications were proposed for usability evaluation methods. Usually methods are classified as: (1) empirical usability testing, based on users' participation [7], and (2) inspection methods, based on experts' judgment [8].

Heuristic evaluation is arguably the most common usability inspection method. Usability specialists (evaluators) analyze every interactive element and dialog following a set of established usability design principles called heuristics [3]. Generic or specific heuristics may be used. Generic heuristics, as Nielsen's ones, are familiar to evaluators and therefore easy to apply; however they are not universally suitable, and can miss specific usability issues. Specific heuristics can detect relevant usability issues related to the application area.

We developed sets of specific usability heuristics for smartphones [4], touchscreen-based mobile applications [9], grid computing applications [10], virtual worlds [11], interactive digital television [12], transactional web applications [13], driving simulators [14], u-Learning applications [15], and cultural aspects [16], among others. We used a methodology that we proposed backed in 2011 [17]. The methodology is currently under review; some changes have been already proposed [18].

We systematically conduct studies on evaluators' perception over generic (Nielsen's) and specific usability heuristics. We developed a standard questionnaire, concerning 4 dimensions: D1 - *Utility*, D2 - *Clarity*, D3 - *Ease of use*, D4 - *Necessity of additional checklist*. All dimensions are evaluated using a 5 points Likert scale. The studies offer an important feedback for both teaching and research. Some results have been published [19, 20].

3 A Set of Usability Heuristics for Smartphones

We developed a set of usability heuristics for smartphones - SMASH [4]. It includes 12 heuristics and it is based on a set of usability heuristics for touchscreen-based mobile applications [9]. SMASH heuristics are briefly described below.

SMASH1 - *Visibility of system status*: The device should keep the user informed about all the processes and state changes through feedback and in a reasonable time.
 SMASH2 - *Match between system and the real world*: The device should speak the users' language instead of system oriented concepts and technicalities. The device

should follow the real world conventions and display the information in a logical and natural order.

SMASH3 - *User control and freedom*: The device should allow the user to undo and redo his/her actions, and provide clearly pointed “emergency exits” to leave unwanted states. These options should be available preferably through a physical button or equivalent.

SMASH4 - *Consistency and standards*: The device should follow the established conventions, allowing the user to do things in a familiar, standard and consistent way.

SMASH5 - *Error prevention*: The device should hide or deactivate unavailable functionalities, warn users about critical actions and provide access to additional information.

SMASH6 - *Minimize the user's memory load*: The device should offer visible objects, actions and options in order to prevent users from having to memorize information from one part of the dialog to another.

SMASH7 - *Customization and shortcuts*: The device should provide basic and advanced configuration options, allow definition and customization of shortcuts to frequent actions.

SMASH8 - *Efficiency of use and performance*: The device should be able to load and display the required information in a reasonable time and minimize the required steps to perform a task. Animations and transitions should be displayed smoothly.

SMASH9 - *Esthetic and minimalist design*: The device should avoid displaying unwanted information overloading the screen.

SMASH10 - *Help users recognize, diagnose, and recover from errors*: The device should display error messages in a language familiar to the user, indicating the issue in a precise way and suggesting a constructive solution.

SMASH11 - *Help and documentation*: The device should provide easy-to-find documentation and help, centered on the user's current task and indicating concrete steps to follow.

SMASH12 - *Physical interaction and ergonomics*: The device should provide physical buttons or the equivalent for main functionalities, located in positions recognizable by the user, which should fit the natural posture (and reach) of the user's dominant hand.

4 Evaluating Facebook's Usability: Evaluators' Perception

We made an experiment with 16 undergraduate Computer Science students at Pontificia Universidad Católica de Valparaíso, Chile. They performed a heuristic evaluation of the mobile version of Facebook, based on SMASH. They all had (low) previous experience in heuristic evaluations, based on Nielsen's heuristics. They were all frequent users of Facebook, mainly in its mobile version.

After performing the heuristic evaluation, all participants were asked to rate SMASH heuristics, based on a standard questionnaire, using a 5 points Likert scale. The average scores are presented in Table 1.

Table 1. Average scores of evaluators' perception on SMASH

	D1 - utility	D2 - clarity	D3 - ease of use	D4 - necessity of additional checklist
SMASH1	4.44	4.00	3.31	3.81
SMASH2	3.81	3.44	3.25	4.13
SMASH3	3.81	3.56	3.25	4.38
SMASH4	4.06	3.75	3.25	4.06
SMASH5	4.31	3.75	3.44	4.38
SMASH6	3.81	3.56	3.06	4.06
SMASH7	3.63	3.44	3.50	4.25
SMASH8	4.00	3.75	3.38	4.06
SMASH9	4.31	3.94	3.81	4.00
SMASH10	4.19	3.75	3.31	4.06
SMASH11	3.75	4.06	3.81	4.00
SMASH12	3.88	3.87	3.43	3.94
Average score by dimension	4.00	3.74	3.40	4.09

When evaluating Facebook, SMASH heuristics are perceived as useful (average score 4.00) and clear (average score 3.74). However, they are perceived as not quite easy to use (average score 3.40), and therefore there is a necessity of additional checklist (average score 4.09).

SMASH1 - *Visibility of system status* is perceived as the most useful heuristic (4.44); it is also perceived as clear (4.00), and the necessity of additional checklist is the lowest one (3.81). SMASH9 - *Esthetic and minimalist design* is also positively perceived: useful (4.31), clear (3.94), and easy to use (3.81). On the opposite side, SMASH7 - *Customization and shortcuts* is perceived as the less useful (3.63) and less clear (3.44) heuristic; the necessity of additional checklist is high (4.25).

Some heuristics are perceived as useful, relatively clear, but not quite easy to use: SMASH4 - *Consistency and standards*, SMASH5 - *Error prevention*, SMASH8 - *Efficiency of use and performance*, and SMASH10 - *Help users recognize, diagnose, and recover from errors*. Their associated necessities of additional checklist are quite high.

On the other hand, even if it is perceived as clear and easy to use, heuristic SMASH11 - *Help and documentation* is not perceived as really useful. That is probably because evaluators are so familiar with Facebook; they do not feel the need for help and documentation when using this particular social network.

As observations' scale is ordinal, and no assumption of normality could be made, the survey results were analyzed using nonparametric statistics tests. In all tests $p \leq 0.05$ was used as decision rule. Spearman ρ tests were performed to check the hypothesis:

- $H_0: \rho = 0$, the dimensions D_m and D_n are independent,
- $H_1: \rho \neq 0$, the dimensions D_m and D_n are dependent.

Table 2 shows the correlations between dimensions when all 12 SMASH heuristics are considered. There is a very strong significant correlation between dimensions D2 – Clarity and D3 - Ease of use. As expected, when heuristics are perceived as clear, they are also perceived as easy to use. There are no other significant correlations.

Table 2. Spearman ρ test when all SMASH heuristics are considered

	D1 - utility	D2 - clarity	D3 - ease of use	D4 - necessity of additional checklist
D1	1	Independent	Independent	Independent
D2		1	0.802	Independent
D3			1	Independent
D4				1

We also performed Spearman ρ tests for each heuristic (Tables 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 and 14). There are only two very strong significant correlations: (1) between dimensions D1 – Utility and D3 - Ease of use, in the case of SMASH1 - Visibility of system status, and (2) between dimensions D1 – Utility and D2 – Clarity, in the case of SMASH2 - Match between system and the real world. There are few strong or moderate correlations. The most recurrent correlation occurs between dimensions D2 – Clarity and D3 - Ease of use (for 7 out of 12 heuristics).

Table 3. Spearman ρ test for SMASH1

	D1 - utility	D2 - clarity	D3 - ease of use	D4 - necessity of additional checklist
D1	1	0.510	0.864	Independent
D2		1	0.663	Independent
D3			1	Independent
D4				1

Table 4. Spearman ρ test for SMASH2

	D1 - utility	D2 - clarity	D3 - ease of use	D4 - necessity of additional checklist
D1	1	0.816	Independent	Independent
D2		1	0.776	Independent
D3			1	Independent
D4				1

Table 5. Spearman ρ test for SMASH3

	D1 - utility	D2 - clarity	D3 - ease of use	D4 - necessity of additional checklist
D1	1	Independent	Independent	Independent
D2		1	Independent	Independent
D3			1	Independent
D4				1

Table 6. Spearman ρ test for SMASH4

	D1 - utility	D2 - clarity	D3 - ease of use	D4 - necessity of additional checklist
D1	1	Independent	Independent	Independent
D2		1	0.752	Independent
D3			1	Independent
D4				1

Table 7. Spearman ρ test for SMASH5

	D1 - utility	D2 - clarity	D3 - ease of use	D4 - necessity of additional checklist
D1	1	Independent	Independent	Independent
D2		1	Independent	Independent
D3			1	Independent
D4				1

Table 8. Spearman ρ test for SMASH6

	D1 - utility	D2 - clarity	D3 - ease of use	D4 - necessity of additional checklist
D1	1	Independent	Independent	Independent
D2		1	0.683	Independent
D3			1	Independent
D4				1

Table 9. Spearman ρ test for SMASH7

	D1 - utility	D2 - clarity	D3 - ease of use	D4 - necessity of additional checklist
D1	1	Independent	Independent	Independent
D2		1	0.556	Independent
D3			1	Independent
D4				1

Table 10. Spearman ρ test for SMASH8

	D1 - utility	D2 - clarity	D3 - ease of use	D4 - necessity of additional checklist
D1	1	Independent	Independent	Independent
D2		1	Independent	Independent
D3			1	Independent
D4				1

Table 11. Spearman ρ test for SMASH9

	D1 - utility	D2 - clarity	D3 - ease of use	D4 - necessity of additional checklist
D1	1	Independent	0.576	Independent
D2		1	0.549	Independent
D3			1	Independent
D4				1

Table 12. Spearman ρ test for SMASH10

	D1 - utility	D2 - clarity	D3 - ease of use	D4 - necessity of additional checklist
D1	1	0.599	Independent	Independent
D2		1	0.659	Independent
D3			1	Independent
D4				1

Table 13. Spearman ρ test for SMASH11

	D1 - utility	D2 - clarity	D3 - ease of use	D4 - necessity of additional checklist
D1	1	0.554	Independent	Independent
D2		1	Independent	Independent
D3			1	Independent
D4				1

Table 14. Spearman ρ test for SMASH12

	D1 - utility	D2 - clarity	D3 - ease of use	D4 - necessity of additional checklist
D1	1	Independent	0.662	Independent
D2		1	Independent	Independent
D3			1	Independent
D4				1

We asked participants three additional questions; responses were evaluated using a 5 points Likert scale. Average scores are presented in Table 15.

Table 15. Overall perception on SMASH when evaluating social media’s usability

Question	Average score
Q1 – <i>Easiness</i> : How easy was to perform the heuristic evaluation of Facebook?	3.00
Q2 – <i>Intention</i> : Would you use SMASH when evaluating social media in the future?	3.88
Q3 – <i>Completeness</i> : Do you think SMASH covers all usability aspects for social media?	3.94

Even if evaluators do not think the heuristic evaluation of Facebook was an easy task, they do perceive SMASH as an appropriate instrument to evaluate social media's usability, and they intent to use it in future evaluations.

We also performed a Spearman ρ test to check the correlation between Q1, Q2, Q3, and D1, D2, D3, D4. Results are presented in Table 16.

Table 16. Spearman ρ tests for Q1, Q2, Q3, and D1, D2, D3, D4

	Q1 – easiness	Q2 – intention	Q3 – completeness	D1 - utility	D2 - clarity	D3 - ease of use	D4 - necessity of additional checklist
Q1	1	Independent	Independent	Independent	Independent	0.570	Independent
Q2		1	Independent	Independent	Independent	Independent	Independent
Q3			1	0.517	Independent	Independent	0.542
D1				1	Independent	Independent	Independent
D2					1	0.802	Independent
D3						1	Independent
D4							1

Correlations between dimensions D were analyzed above. As we already mentioned, there is only one very strong significant correlation, between dimensions D2 – *Clarity* and D3 - *Ease of use*.

Analyzing other correlations, we noticed that:

- Q1 – *Easiness* is moderately correlated with D3 – *Ease of use*. When SMASH are perceived as easy to use, the heuristic evaluation is also perceived as easy to perform.
- Q3 – *Completeness* is moderately correlated with dimensions D1 – *Utility*, and D4 - *Necessity of additional checklist*. When evaluators perceived SMASH as useful, they also feel is an appropriate/complete tool. But they also think that SMASH could be complemented with an additional checklist.

5 Conclusions

Evaluating specific applications' usability is still challenging. Social media is not an exception. Even a well know inspection method, as heuristic evaluation, is hard to perform by novice evaluators. As generic heuristics can miss specific usability issues, we usually prefer to use specific heuristics, which can detect relevant domain related usability issues.

16 undergraduate Computer Science students performed a heuristic evaluation of the mobile version of Facebook, based on a set of 12 usability heuristics that we developed (SMASH). SMASH targets smartphone applications in general, but it seems to work well when evaluating social media. Evaluators do not perceive the heuristic evaluation of Facebook as an easy task, probably because they were using SMASH for

the very first time. However, they think SMASH is an appropriate instrument to evaluate social media's usability, and they intend to use it in future evaluations.

We surveyed evaluators' perception on SMASH based on 4 dimensions: D1 - *Utility*, D2 - *Clarity*, D3 - *Ease of use*, D4 - *Necessity of additional checklist*. There is only one strong significant correlation between dimensions D2 and D3, when all 12 heuristics are considered; when heuristics are perceived as clear, they are also perceived as easy to use. Correlation between dimensions D2 and D3 was also the most recurrent one identified in a previous study [20]. When performing Spearman ρ tests for each heuristic, no patterns could be identified.

As future work, we intend to complement the study with a qualitative approach, based on data collected through surveys and interviews.

References

1. ISO 9241-210: Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems. International Organization for Standardization, Geneva (2010)
2. Lewis, J.: Usability: lessons learned... and yet to be learned. *Int. J. Hum.-Comput. Inter.* **30** (9), 663–684 (2014)
3. Nielsen, J., Mack, R.L.: *Usability Inspection Methods*. John Wiley & Sons, New York (1994)
4. Inostroza, R., Rusu, C., Roncagliolo, S., Rusu, V., Collazos, C.: Developing SMASH: a set of SMARtphone's uSability heuristics. *Comput. Stand. Interfaces* **43**, 40–52 (2016)
5. ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability. International Organization for Standardization, Geneva, (1998)
6. Bevan, N., Carter, J., Harker, S.: ISO 9241-11 revised: what have we learnt about usability since 1998? In: Kurosu, M. (ed.) *HCI 2015. LNCS*, vol. 9169, pp. 143–151. Springer, Cham (2015). doi:[10.1007/978-3-319-20901-2_13](https://doi.org/10.1007/978-3-319-20901-2_13)
7. Dumas, J., Fox, J.: Usability testing: current practice and future directions. In: Sears, A., Jacko, J. (eds.) *The Human – Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, pp. 1129–1149. Taylor & Francis, New York (2008)
8. Cockton, G., Woolrych, A., Lavery, D.: Inspection – based evaluations. In: Sears, A., Jacko, J. (eds.) *The Human – Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, pp. 1171–1189. Taylor & Francis, New York (2008)
9. Inostroza, R., Rusu, C., Roncagliolo S., Rusu V.: Usability heuristics for touchscreen-based mobile devices: update. In: *First Chilean Conference on Human - Computer Interaction (ChileCHI2013)*, pp. 24–29. ACM International Conference Proceeding Series (2013)
10. Roncagliolo, S., Rusu, V., Rusu, C., Tapia, G., Hayvar, D., Gorgan D.: Grid computing usability heuristics in practice. In: *8th International Conference on Information Technology: New Generations (ITNG2011)*, pp. 145–150. IEEE Computer Society Press (2011)
11. Rusu, C., Muñoz, R., Roncagliolo, S., Rudloff, S., Rusu, V., Figueroa A.: Usability heuristics for virtual worlds. In: *The Third International Conference on Advances in Future Internet (AFIN2011)*, pp. 16–19. IARIA (2011)
12. Solano, A., Rusu, C., Collazos, C., Arciniegas, J.: Evaluating interactive digital television applications through usability heuristics. *Ingeniare* **21**(1), 16–29 (2013)

13. Quiñones, D., Rusu C., Roncagliolo S.: Redefining usability heuristics for transactional web applications. In: 11th International Conference on Information Technology: New Generations (ITNG2014), pp. 260–265. IEEE Computer Society Press (2014)
14. Campos, A., Rusu, C., Roncagliolo, S., Sanz, F., Galvez, R., Quiñones, D.: Usability heuristics and design recommendations for driving simulators. *Adv. Intell. Syst. Comput.* **448**, 1287–1290 (2016)
15. Sanz, F., Galvez, R., Rusu, C., Roncagliolo, S., Rusu, V., Collazos, C., Cofré, J.P., Campos, A., Quiñones, D.: A set of usability heuristics and design recommendations for u-Learning applications. *Adv. Intell. Syst. Comput.* **448**, 983–993 (2016)
16. Diaz, J., Rusu, C., Collazos, C.: Experimental validation of a set of cultural-oriented usability heuristics: e-Commerce websites evaluation. *Comput. Stand. Interfaces* **50**, 160–178 (2017)
17. Rusu, C., Roncagliolo, S., Rusu, V., Collazos C.: A methodology to establish usability heuristics. In: The Fourth International Conference on Advances in Computer-Human In-teractions (ACHI2011), pp. 59–62. IARIA (2011)
18. Quiñones, D., Rusu, C., Roncagliolo, S., Rusu, V., Collazos, C.: Developing usability heuristics: a formal or informal process? *IEEE Lat. Am. Trans.* **14**(7), 3400–3409 (2016)
19. Rusu, C., Rusu, V., Roncagliolo, S., Apablaza, J., Rusu, V.Z.: User experience evaluations: challenges for newcomers. In: Marcus, A. (ed.) DUXU 2015. LNCS, vol. 9186, pp. 237–246. Springer, Cham (2015). doi:[10.1007/978-3-319-20886-2_23](https://doi.org/10.1007/978-3-319-20886-2_23)
20. Rusu, C., Rusu, V., Roncagliolo, S., Quiñones, D., Rusu, V.Z., Fardoun, H.M., Alghazzawi, D.M., Collazos, C.A.: Usability heuristics: reinventing the wheel? In: Meiselwitz, G. (ed.) SCSM 2016. LNCS, vol. 9742, pp. 59–70. Springer, Cham (2016). doi:[10.1007/978-3-319-39910-2_6](https://doi.org/10.1007/978-3-319-39910-2_6)