

# Subjective and Objective Assessment of Mashup Tools

Tihomir Orehovački<sup>1</sup> and Toni Granollers<sup>2</sup>

<sup>1</sup> University of Zagreb, Faculty of Organization and Informatics  
Pavliška 2, 42000 Varaždin, Croatia  
tihomir.orehovacki@foi.hr

<sup>2</sup> University of Lleida, Polytechnic School, GRIHO Research Group  
C. de Jaume II, 69, 25001 Lleida, Spain  
tonig@diei.udl.cat

**Abstract.** Mashup tools are platforms that enable end users to combine components from multiple sources and thus create, deploy, and share their own web applications. The aim of this paper is to present and discuss the findings of two empirical studies on evaluation of mashup tools by means of subjective and objective measuring instruments. In the first study, data were gathered with the employment of SUS post-use questionnaire and Tobii eye-tracking device, and in the second, by applying the retrospective thinking aloud method. The analysis of collected data uncovered strengths and weaknesses of evaluated mashup tools. In addition, the relevance that users assign to quality in use attributes in the context of mashup tools has been determined.

**Keywords:** Mashup Tools, Subjective and Objective Evaluation, Post-use Questionnaire, Eye-tracking, Retrospective Thinking Aloud, Empirical Findings.

## 1 Introduction

Mashup tools are specific breed of Web 2.0 applications [16] that integrate disparate components, such as RSS/Atom feeds, services, content provided by third parties, or widgets, which are readily available on the Web. They are meant for the development of novel situational applications, services, or functionalities in a value-adding fashion. Requiring no programming skills from their users, mashup tools are online development environments that enable everyone to create their own web applications. In that respect, they represent shift from product-oriented software development to consumer-oriented composition [14]. Malinga et al. [13] distinguish four types of mashup tools. Process-oriented mashup tools enable users to simulate business procedures [7] and integrate assorted processes [8] in an enterprise. Data-oriented mashup tools facilitate the aggregation of data from heterogeneous remotely hosted online sources. Knowledge management mashup tools improve the interaction among members of the organization. Interface-oriented mashup tools help users to develop interactive input menus and query result filters. Hoyer and Fischer [9] have categorized mashup tools into two main groups. Consumer mashup tools enable individuals to create mashups in the form of personalized web portals that combine data elements from more than one source

(e.g. news, weather forecast, e-mails, etc.) and display them on a simple graphical interface. This individualized browser start pages are frequently referred to as front-end [22] or presentation layer [21] mashups. Enterprise mashup tools are intended for the development of applications that integrate various resources from multiple sources deployed in an enterprise ecosystem.

Evaluation plays an important role in a lifecycle of every web application. Although literature offers various models and methodologies dealing with the evaluation of web sites [15][20] and Web 2.0 applications [17][18], approaches addressing the assessment of mashup tools are less common. This paper reports findings of two empirical studies in which mashup tools were evaluated by means of subjective and objective measuring instruments.

The remainder of the paper is structured as follows. Brief overview of current advances in the assessment of mashups is contained in the next section. Results of conducted studies are presented in the third and the fourth section. Concluding remarks and future research directions are provided in the last section.

## 2 Background to the Research

Recent research on the assessment of mashups was focused on exploring diverse aspects of their quality and usability. Drawing on recent standard [11] aimed for evaluating different facets of software quality, ergonomic criteria introduced by Bastien and Scapin [2], and usability guidelines for Web design, Insfran et al. [10] proposed the Mashup Usability Model in which usability has been decomposed into following six sub-characteristics: appropriateness recognisability, learnability, operability, user error protection, user interface aesthetics, and accessibility. Koschmider et al. [12] emphasize that selection of quality metrics depends on the type of mashups. The same authors argue that presentation and extraction mashups should be evaluated in terms of consistent graphical representation, data mashups have to meet metrics meant for the assessment of efficiency in modularization and connectivity, functionality mashups need to satisfy metrics related to accessibility of disparate mashed data and the composition itself, flow mashups should address metrics intended for evaluating connectivity, availability of components, and error rates, client- and server-side mashups have to comply with metrics related to efficient security policies and right management, while the quality of enterprise mashups should be measured in terms of availability, error rates, popularity, user rating, and comments. Quality of the mashup can be observed from the perspective of heterogeneous components that constitute it [5] as well as from the aspect of the final composition [19]. In that respect, Cappelletto et al. [4] developed the model which enables evaluation of following quality dimensions: data quality (accuracy, timeliness, completeness, availability, and consistency), presentation quality (usability and accessibility), and composition quality (added value, component suitability, component usage, consistency, and availability). Nevertheless, results of a systematic mapping study [6] suggest that there is still a need for empirical research addressing the assessment of security, usability, and reliability of mashups.

Contrary to the aforementioned, studies related to the evaluation of tools aimed for the development of mashups are rather scarce. For that reason we initiated a research which was comprised of two experimental studies. The objective of the first study was to determine to what extent measuring instruments meant for evaluating the usability of websites are suitable for the assessment of consumer mashup tools. The purpose of the second study was to identify a set of quality in use attributes which users find relevant in the context of interaction with front-end mashup tools. Findings of both studies are described in more detail in the following two sections.

### 3 First Study

**Participants.** Ten subjects ranging in age from 19 to 34 years ( $M = 26.00$ ,  $SD = 5.696$ ) were recruited for the first study. The sample was composed of 70% male and 30% female students from the Polytechnic School at the University of Lleida in Spain. At the time when the study was carried out, the majority of participants (50%) were in the second year of a graduate programme, 40% of them were second-year undergraduate students, while 10% of them were enrolled in the third year of an undergraduate programme. Most of respondents (60%) had at least very good knowledge of using Web 2.0 applications. Study participants had been loyal users of Facebook, YouTube, Twitter, and Wikipedia (80%, 40%, 30%, and 30% of them, respectively, used those popular Web 2.0 applications at least between once and twice a day).

**Measures.** Users' performance in executing scenario tasks was examined with five disparate metrics. Task efficiency is the estimated amount of minutes needed to complete the task. Task completion refers to the ratio of number of tasks attempted and total number of tasks. Task effectiveness denotes the product of task completion and the extent to which goals of the task have been reached. Task productivity represents the proportion of task effectiveness and task efficiency. Number of fixations is total quantity of relatively stable positions of users' eyes per task. Perceived usability of both mashup tools (Netvibes and Protopage) was explored with System Usability Scale (SUS) [3] post-use questionnaire.

**Apparatus.** With an aim to create a comfortable environment, the study was conducted in a GRIHO usability lab. Data on participants' eye movements as well as their efficiency and effectiveness in completing scenario tasks were recorded with Tobii T60 eye tracker which is a 15-inch LCD display (96 dpi) with a screen resolution of 1280x1024 pixels and a sampling rate of 60 Hz. The collected data were analyzed with Tobii Studio 3.2 and Morae 3 software. The SUS was administrated online using the KwikSurveys questionnaire builder. Responses to the items were scored on a five point Likert scale (1- strongly disagree, 5 – strongly agree).

**Procedure.** The study was composed of two main parts: the execution of predefined tasks and the completion of the online questionnaire. Upon arriving at the lab, the participants were welcomed and briefly introduced about the usability evaluation

study which was followed by the explanation of the equipment to be used. The participants were then asked to complete an informed consent form. At the beginning of the tasks performance session, the form containing a list of fifteen representative tasks (see Appendix) was given to the participant. The eye-tracker was then calibrated to the participant with a short 9-point procedure during which he or she watched a circle moving on the screen. When the eye calibration succeeded, Tobii Studio launched the mashup tool in Internet Explorer web browser. The experimenter left the room while the participant started working on tasks. Data recording commenced when the participant began reading the task, and ended when the participant either completed or cancelled the task. Each participant went through the calibration procedure and tasks solving session twice – first with Netvibes and then using the Protopage. After completing all the tasks with both mashup tools, the participants were asked to fill out the post-use questionnaire. At the end of the study, respondents were debriefed, and thanked for their participation. The duration of the experiment was between 30 minutes and one hour.

**Findings.** The first study adopted a within-subjects design contrasting two consumer mashup tools. Taking into account the results of Shapiro-Wilk Tests which uncovered that at least one of the variables in a pairwise comparison violates the assumption of normality in data ( $p < .05$ ), differences between evaluated mashup tools were explored by means of Wilcoxon Signed-Rank Tests.

Mean values of the time-on-task metric are presented in Table 1. Overall, the participants have spent significantly more time ( $Z = -2.805$ ,  $p = .005$ ,  $r = -.63$ ) on completing all scenario tasks by means of Netvibes (Mdn = 22.865) than when they employed Protopage (Mdn = 11.855) for the equivalent purpose. More specifically, the participants were significantly more efficient ( $Z = -2.803$ ,  $p = .005$ ,  $r = -.63$ ) in adding widgets to the tab entitled Business when they were using Protopage (Mdn = 1.075) than when they performed the same task with Netvibes (Mdn = 2.445). As a follow up, the analysis of recordings revealed that participants had difficulties in finding functionality for adding widgets to the Netvibes. On the other hand, the participants had to invest significantly less time ( $Z = -2.191$ ,  $p < .05$ ,  $r = -.49$ ) to personalize the Google Maps widget on Netvibes (Mdn = .400) because during the execution of the identical task with Protopage (Mdn = 1.365) they could not figure out how to change the default location displayed in the widget. Data analysis also uncovered that participants were significantly less efficient ( $Z = -2.805$ ,  $p = .005$ ,  $r = -.63$ ) in changing the background image on Netvibes (Mdn = 3.090) than completing the equivalent task on Protopage (Mdn = .450). The rationale behind this finding is that participants were unable to find the specified background theme in a timely manner. It was also found that Netvibes (Mdn = 2.425) does not support changing the tab color which is the reason why the participants needed significantly less time ( $Z = -2.803$ ,  $p = .005$ ,  $r = -.63$ ) to complete the eleventh task with Protopage (Mdn = .190). The participant were significantly more efficient ( $Z = -2.805$ ,  $p = .005$ ,  $r = -.63$ ) in sorting widgets when they were using Protopage (Mdn = .240) than when they were performing the identical task with Netvibes (Mdn = .775). In addition, the participants invested significantly less time ( $Z = -2.599$ ,  $p < .01$ ,  $r = -.58$ ) in changing the

background image to the default value on Protopage (Mdn = .225) than when they were doing the same on Netvibes (Mdn = 1.040). Finally, the participants spent significantly more time ( $Z = -2.312$ ,  $p < .05$ ,  $r = -.52$ ) on logging out from Netvibes (Mdn = .195) than completing the equivalent task with Protopage (Mdn = .140). All the aforementioned effects are large in size.

There was no significant difference ( $Z = -.912$ ,  $p = .362$ ) between two mashup tools in terms of the extent to which the participants succeeded to complete all scenario tasks. Since Protopage (Mdn = 0) does not provide the functionality of changing the interface language, the participants were significantly more successful ( $Z = -2.828$ ,  $p = .005$ ,  $r = -.63$ ) in completing the identical task with Netvibes (Mdn = 1). On the other hand, the participants were significantly more effective ( $Z = -2.000$ ,  $p < .05$ ,  $r = -.45$ ) in personalizing the weather widget on Netvibes (Mdn = 1) than when they were in doing the same on Protopage (Mdn = 1). Given that participants experienced difficulties in adding news feed for Esport3.cat website on Protopage (Mdn = 0.75), they were significantly more successful ( $Z = -2.070$ ,  $p < .05$ ,  $r = -.46$ ) in completing this task on Netvibes (Mdn = 1). It was also found that participants were significantly more effective ( $Z = -2.000$ ,  $p < .05$ ,  $r = -.45$ ) in customizing the Google Maps widget on Netvibes (Mdn = 1) than they were in executing the same task on Protopage (Mdn = 0). Taking into account that participants were unable to change tab color on Netvibes (Mdn = 0), they were significantly more effective ( $Z = -3.162$ ,  $p < .005$ ,  $r = -.71$ ) in completing this task on Protopage (Mdn = 1). Finally, the participants were significantly more successful ( $Z = -2.449$ ,  $p < .05$ ,  $r = -.55$ ) in changing the background image to the default value when they were using Protopage than when they were employing Netvibes for the equivalent purpose. The effects related to the fifth, sixth, and eighth task are medium in size while remaining reported effects are large in size.

**Table 1.** Average values of task efficiency and task completion metrics for selected mashup tools

Tasks	Task Efficiency *		Task Completion **	
	Netvibes	Protopage	Netvibes	Protopage
<b>Task 1</b>	0.63	0.70	1.00	1.00
<b>Task 2</b>	0.85	0.68	1.00	1.00
<b>Task 3</b>	1.92	1.74	0.80	0.00
<b>Task 4</b>	3.09	1.20	0.90	1.00
<b>Task 5</b>	0.61	0.86	1.00	0.60
<b>Task 6</b>	1.68	1.74	1.00	0.65
<b>Task 7</b>	1.68	2.18	0.85	0.80
<b>Task 8</b>	0.58	1.31	0.80	0.40
<b>Task 9</b>	3.12	0.49	0.70	0.90
<b>Task 10</b>	0.56	0.27	1.00	0.90
<b>Task 11</b>	2.50	0.23	0.00	1.00
<b>Task 12</b>	0.95	0.25	0.90	1.00
<b>Task 13</b>	1.53	1.09	0.90	0.75
<b>Task 14</b>	1.78	0.23	0.30	0.90
<b>Task 15</b>	0.31	0.14	1.00	1.00

\* A lower value means a better result, \*\* A higher value means a better result

Considering the values shown in Table 2, the participants were significantly more effective when they were performing all scenario tasks with Netvibes (Mdn = .775) than when they were doing the same by means of Protopage (Mdn = .680). More specifically, the participants could complete the third task significantly more accurately ( $Z = -2.828$ ,  $p = .005$ ,  $r = -.63$ ) when they were using Netvibes (Mdn = 1) than they did when they were interacting with Protopage (Mdn = 0). On contrary, the participants were significantly more effective ( $Z = -2.724$ ,  $p < .01$ ,  $r = -.61$ ) in completing the fourth task with Protopage (Mdn = 1), than they were with Netvibes (Mdn = .660). Moreover, Wilcoxon Signed-Rank Test revealed that participants were significantly more effective ( $Z = -3.162$ ,  $p < .005$ ,  $r = -.71$ ) in executing the fifth task by means of Netvibes (Mdn = 1) than when they were working on the identical task with Protopage (Mdn = 0). According to results of data analysis, the participants could complete the sixth task significantly more accurately ( $Z = -2.070$ ,  $p < .05$ ,  $r = -.46$ ) when they employed Netvibes (Mdn = 1) than when they applied Protopage (Mdn = .625) for the equivalent purpose. It was also found that participants were significantly less effective ( $Z = -2.646$ ,  $p < .01$ ,  $r = -.59$ ) in executing the eighth task using the Protopage (Mdn = 0) than when they were solving the identical task with Netvibes (Mdn = 1). On the other hand, the participants were able to complete the eleventh task significantly more accurately ( $Z = -3.000$ ,  $p < .005$ ,  $r = -.67$ ) by means of Protopage (Mdn = 1) than they were with Netvibes (Mdn = 0). Finally, the participants were significantly less effective ( $Z = -2.070$ ,  $p < .05$ ,  $r = -.46$ ) in performing the thirteenth task with Protopage (Mdn = .625) than when they used Netvibes (Mdn = 1) for the equivalent purpose. Effects related to the third, fourth, fifth, eighth, and eleventh task are large in size while effects referring to sixth and thirteenth task are medium in size.

According to the mean values of task productivity presented in Table 2, the participants were in general 50.97% more productive in completing all scenario tasks when they were using Protopage than when they were applying Netvibes for the same purpose.

**Table 2.** Average values of task effectiveness and task productivity metrics for selected mashup tools (note that a higher value means a better result)

Tasks	Task Effectiveness		Task Productivity	
	Netvibes	Protopage	Netvibes	Protopage
<b>Task 1</b>	1.00	1.00	1.58	1.44
<b>Task 2</b>	1.00	1.00	1.17	1.47
<b>Task 3</b>	0.80	0.00	0.42	0.00
<b>Task 4</b>	0.56	1.00	0.18	0.83
<b>Task 5</b>	1.00	0.00	1.63	0.00
<b>Task 6</b>	1.00	0.55	0.60	0.32
<b>Task 7</b>	0.70	0.68	0.42	0.31
<b>Task 8</b>	0.80	0.10	1.39	0.08
<b>Task 9</b>	0.60	0.90	0.19	1.83
<b>Task 10</b>	0.80	0.90	1.44	3.40
<b>Task 11</b>	0.00	0.90	0.00	4.00
<b>Task 12</b>	0.90	1.00	0.95	3.95
<b>Task 13</b>	0.90	0.63	0.59	0.57
<b>Task 14</b>	0.30	0.40	0.17	1.75
<b>Task 15</b>	0.90	1.00	2.95	7.06

An analysis of collected data revealed that the evaluated mashup tools differ significantly ( $Z = -2.293$ ,  $p < .05$ ,  $r = -.51$ ) in terms of the overall number of fixations that were recorded during the implementation of all fifteen tasks. Results of Wilcoxon Signed-Rank Tests presented in Table 3 indicate that participants were significantly more effective in carrying out the fourth, ninth, eleventh, twelfth, and fifteenth task

**Table 3.** Average number of fixations per task for evaluated mashup tools (note that a lower value means a better result)

Tasks	Overall number of fixations		Z	p	r
	Netvibes	Protopage			
Task 1	47.12	52.35	-.357	.721	N/A
Task 2	65.71	52.10	-1.274	.203	N/A
Task 3	145.39	135.74	-.051	.959	N/A
Task 4	244.08	91.67	-2.293	.022	-.51
Task 5	44.49	62.49	-.561	.575	N/A
Task 6	143.79	140.77	-.663	.508	N/A
Task 7	126.95	168.08	-.561	.575	N/A
Task 8	47.04	102.97	-2.191	.028	-.49
Task 9	228.58	38.64	-2.803	.005	-.63
Task 10	37.30	20.72	-1.682	.093	N/A
Task 11	184.40	17.55	-2.803	.005	-.63
Task 12	68.52	19.28	-2.803	.005	-.63
Task 13	103.37	81.88	-1.274	.203	N/A
Task 14	139.28	17.51	-2.599	.009	-.58
Task 15	22.37	11.16	-2.395	.017	-.54

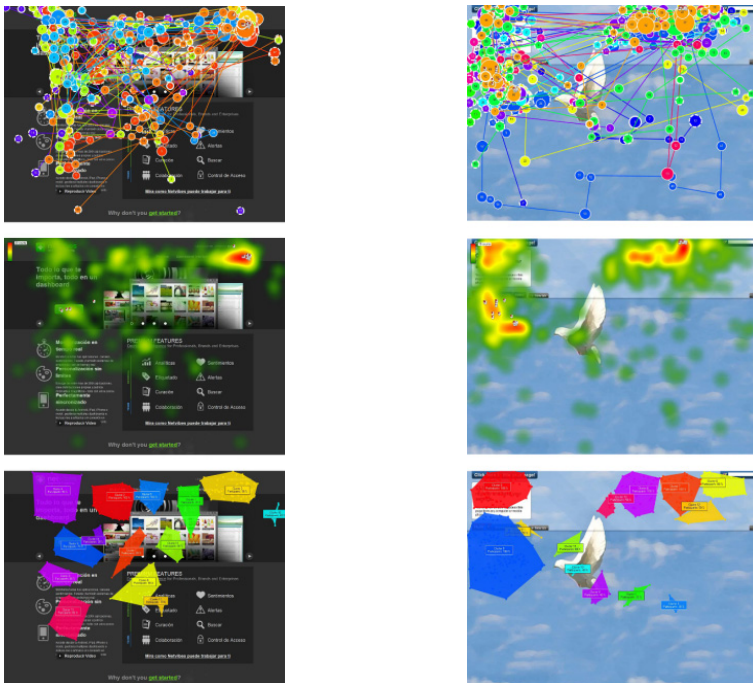
**Table 4.** Perceived usability of selected mashup tools

SUS Items	Netvibes	Protopage
I think that I would like to use this mashup tool frequently.	3.10	3.00
I found the mashup tool unnecessarily complex	3.10	2.90
I thought the mashup tool was easy to use.	3.40	3.60
I think that I would need the support of a technical person to be able to use this mashup tool.	2.50	2.20
I found the various functions in this mashup tool were well integrated.	3.30	3.10
I thought there was too much inconsistency in this mashup tool.	2.70	2.60
I would imagine that most people would learn to use this mashup tool very quickly.	3.50	3.70
I found the mashup tool very cumbersome to use.	2.60	2.40
I felt very confident using the mashup tool.	3.00	3.40
I needed to learn a lot of things before I could get going with this mashup tool.	2.50	2.10
<b>SUS Score</b>	<b>57.25</b>	<b>61.50</b>

when they were using Protopage than when they were interacting with Netvibes. On contrary, the participants have made significantly less fixations when they were working on the eighth task in Netvibes than when they were performing the identical task by means of Protopage. All reported effects are large in size.

According to the values presented in Table 4, Protopage was perceived as more usable mashup tool than Netvibes. Given that SUS scores of both Netvibes and Protopage are below the acceptable threshold [1], their usability should be improved. Nevertheless, the selected mashup tools do not differ significantly ( $p > .05$ ) in terms of responses to particular SUS items nor in the context of the overall SUS scores.

Gaze plots, heat maps, and clusters which represent the distribution of fixations made by all participants during the execution of the first task are shown in Figure 1. The aforementioned illustrations imply that in the case of Netvibes the participants had difficulties in finding a login form while in the context of Protopage their attention was interfered with advertisements and comics which were displayed on the home page.



**Fig. 1.** Gaze plots, heat maps, and clusters (top to bottom) illustrating the distribution of fixations in the first task (left: Netvibes, right: Protopage)

## 4 Second Study

**Participants.** A total of 110 respondents (84.55% male, 15.45% female), aged 20.25 years ( $SD = .960$ ) on average, were involved in the second study. At the time when the study took place, majority (90%) of research subjects were enrolled in the second



year of an undergraduate programme at Faculty of Organization and Informatics, University of Zagreb in Croatia. All of them had been frequent users of social networking site Facebook and video podcasting service YouTube (90.91% and 82.73%, respectively, had been using those Web 2.0 applications at least between once and twice a day). Most of students (95.45%) had at least good knowledge of interacting with Web 2.0 applications.

**Procedure.** During one semester students were asked to employ mashup tools as Personal Learning Environments, that is, for receiving updates on heterogeneous educational artifacts created and integrated with various Web 2.0 applications. At the end of the semester, participants had to deliver a critical review of the used mashup tools in the form of a written report. Considering the specificities of the Retrospective Thinking Aloud method, a two-phase analysis of qualitative data was carried out. Firstly, positive and negative statements related to particular mashup tool were extracted from the reports. Thereafter, each statement was assigned to the quality in use attribute [17] whose definition it most closely fits. Based on the sum of positive and negative statements, a set of attributes that should be considered when evaluating the quality in use of mashup tools was determined. In addition, drawing on the overall number of negative statements allocated, the most common issues users are facing when interacting with mashup tools were identified.

**Table 5.** Results of the analysis of data collected applying the Retrospective Thinking Aloud (RTA) method

Attributes	Netvibes		Protopage	
	Strengths	Weaknesses	Strengths	Weaknesses
Aesthetics	36	10	20	25
Availability		4	4	4
Controllability	3	4	3	1
Customizability	25	4	13	5
Ease of Use	54	13	56	13
Effectiveness	20	3	30	4
Efficiency	9	2	10	
Learnability	8	7	5	7
Loyalty	1	1	3	1
Navigability	16	4	18	7
Reliability	10	4	7	4
Responsiveness		2		1
Satisfaction	26	6	29	4
Uniqueness		23	1	23
Usefulness	7		12	1

**Findings.** With the employment of the RTA method the participants generated the set of 613 statements related to the strengths (69.49%) and weaknesses (30.51%) of evaluated mashup tools. Majority of statements were related to the following quality in use attributes: ease of use (22.19%), aesthetics (14.85%), satisfaction (10.60%),

effectiveness (9.30%), customizability (7.67%), uniqueness (7.67%), and navigability (7.34%). Statements on weaknesses of selected mashup tools were mostly attached to the uniqueness (24.60%), aesthetics (18.72%), ease of use (13.90%), learnability (7.49%), navigability (5.88%), satisfaction (5.35%), and customizability (4.81%). On the other hand, strengths that were reported in majority of cases were assigned to the ease of use (25.82%), aesthetics (13.15%), satisfaction (12.91%), effectiveness (11.74%), customizability (8.92%), and navigability (7.98%). The results of data analysis presented in Table 5 indicate that Protopage (47.39%) has better ratio of strengths and weaknesses than Netvibes (40.47%).

## 5 Conclusion

The objective of the work presented in this paper was to examine consumer mashup tools from subjective and objective perspective. For that purpose, two empirical studies were carried out. Results of the first study imply that objective metrics have a higher level of validity than subjective ones. In that respect, metrics of task efficiency, completion, effectiveness, and productivity can be used to determine the significant difference among mashup tools. On the other hand, mashup tools cannot be compared and ranked by mean values of responses to the SUS post-use questionnaire items.

Findings of the second study suggest that each front-end mashup tool should: be simple to operate (ease of use); have an attractive interface (aesthetics); meet users' expectations (satisfaction); enable users to execute tasks accurately and completely (effectiveness); be distinctive among other mashup tools (uniqueness); be customizable to users' specificities (customizability); have intuitive user interface (navigability); enable users to quickly become proficient in using its interface functionalities (learnability); be dependable, stable, and bug-free (reliability); enable users to quickly complete tasks (efficiency); be beneficial in a specified context of use (usefulness); allow use without registration and enable signing in with existing user account (availability); give users full freedom in carrying out tasks (controllability); encourage visitors to become regular users (loyalty); and react promptly to users' actions (responsiveness).

The aforementioned set of attributes will be used as a basis for our future work which will be focused on the development of a measuring instrument that would enable the assessment of all relevant pragmatic and hedonic facets of the quality in use in the context of mashup tools. With an aim to explore the robustness of findings presented in this paper and to draw generalizable sound conclusions, further studies will be carried out during which heterogeneous groups of users will evaluate various types of mashup tools.

## References

1. Bangor, A., Kortum, P., Miller, J.: Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies* 4(3), 114–123 (2009)
2. Bastien, J.M., Scapin, D.L.: Ergonomic Criteria for the Evaluation of Human-Computer Interfaces. Technical Report no. 156. INRIA, Rocquencourt (1993)

3. Brooke, J.: SUS: a “quick and dirty” usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, I.L. (eds.) *Usability Evaluation in Industry*, pp. 189–194. Taylor and Francis, London (1996)
4. Cappiello, C., Daniel, F., Koschmider, A., Matera, M., Picozzi, M.: A Quality Model for Mashups. In: Auer, S., Díaz, O., Papadopoulos, G.A. (eds.) *ICWE 2011. LNCS*, vol. 6757, pp. 137–151. Springer, Heidelberg (2011)
5. Cappiello, C., Daniel, F., Matera, M.: A Quality Model for Mashup Components. In: Gaedke, M., Grossniklaus, M., Díaz, O. (eds.) *ICWE 2009. LNCS*, vol. 5648, pp. 236–250. Springer, Heidelberg (2009)
6. Cedillo, P., Fernandez, A., Insfran, E., Abrahão, S.: Quality of Web Mashups: A Systematic Mapping Study. In: Sheng, Q.Z., Kjeldskov, J. (eds.) *ICWE Workshops 2013. LNCS*, vol. 8295, pp. 66–78. Springer, Heidelberg (2013)
7. De Vrieze, P., Xu, L., Bouguettaya, A., Yang, J., Chen, J.: Process-Oriented Enterprise Mashups. In: *Proceedings of the Workshops at the 4th International Conference on Grid and Pervasive Computing*, pp. 64–71. IEEE, Geneva (2009)
8. Fichter, D., Wisniewski, J.: They Grow Up So Fast: Mashups in the Enterprise. *Online* 33(3), 54–57 (2009)
9. Hoyer, V., Fischer, M.: Market Overview of Enterprise Mashup Tools. In: Bouguettaya, A., Krueger, I., Margaria, T. (eds.) *ICSOC 2008. LNCS*, vol. 5364, pp. 708–721. Springer, Heidelberg (2008)
10. Insfran, E., Cedillo, P., Fernández, A., Abrahão, S., Matera, M.: Evaluating the Usability of Mashups Applications. In: *Proceedings of the 8th International Conference on the Quality of Information and Communications Technology*, pp. 323–326. IEEE, Lisbon (2012)
11. ISO/IEC 25010:2011. Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuARE) – System and software quality models (2011)
12. Koschmider, A., Hoyer, V., Giessmann, A.: Quality Metrics for Mashups. In: *Proceedings of the Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists*, pp. 376–380. ACM, Bela-Bela (2010)
13. Malinga, M., Gruner, S., Koschmider, A.: Quality and usability of mashup tools: criteria and evaluation. In: *Proceedings of the Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists*, pp. 154–159. ACM, East London (2013)
14. Nestler, T.: Towards a mashup-driven end-user programming of SOA-based applications. In: *Proceedings of the 10th International Conference on Information Integration and Web-based Applications & Services*, pp. 551–554. ACM, Linz (2008)
15. Olsina, L., Rossi, G.: Measuring Web Application Quality with WebQEM. *IEEE Multimedia* 9(4), 20–29 (2002)
16. Orehovački, T., Bubaš, G., Kovačić, A.: Taxonomy of Web 2.0 Applications with Educational Potential. In: Cheal, C., Coughlin, J., Moore, S. (eds.) *Transformation in Teaching: Social Media Strategies in Higher Education*, pp. 43–72. Informing Science Press, Santa Rosa (2012)
17. Orehovački, T., Granić, A., Kermek, D.: Evaluating the Perceived and Estimated Quality in Use of Web 2.0 Applications. *The Journal of Systems and Software* 86(12), 3039–3059 (2013)
18. Pang, M., Suh, W., Hong, J., Kim, J., Lee, H.: A New Web Site Quality Assessment Model for the Web 2.0 Era. In: Murugesan, S. (ed.) *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications*, pp. 387–410. IGI Global, Hershey (2010)

19. Picozzi, M., Rodolfi, M., Cappiello, C., Matera, M.: Quality-Based Recommendations for Mashup Composition. In: Daniel, F., Facca, F.M. (eds.) ICWE 2010 Workshops. LNCS, vol. 6385, pp. 360–371. Springer, Heidelberg (2010)
20. Seffah, A., Donyae, M., Kline, R.B., Padda, H.K.: Usability measurement and metrics: A consolidated model. *Software Quality Journal* 14(2), 159–178 (2006)
21. Young, G.O.: The Mashup Opportunity – How To Make Mone in The Evolving Mashup Ecosystem. Forrester Research (2008), [http://mdc.jackbe.com/enterprise-mashup/sites/default/files/Forrester\\_The\\_Mashup\\_Opportunity\\_May2008\\_0.pdf](http://mdc.jackbe.com/enterprise-mashup/sites/default/files/Forrester_The_Mashup_Opportunity_May2008_0.pdf)
22. Zhao, Z., Bhattarai, S., Liu, J., Crespi, N.: Mashup Services to Daily Activities – End-user Perspective in Designing a Consumer Mashups. In: Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services, pp. 222–229. ACM, Ho Chi Minh City (2011)

## Appendix

Please complete the following tasks with Netvibes first and then by means of Protopage:

1. Log in with the predefined username and password.
2. Create three new tabs: Business, News, and Fun.
3. Change the interface language to Spanish. If it is already in Spanish, change it to English.
4. Add following widgets to the Business tab: To Do List, Weather, and Calendar.
5. Personalize the Weather widget in a manner that it displays the weather in Barcelona.
6. Add following widgets to the News tab: CNN.com and Esport3.cat.
7. Add following widgets to the Fun tab: Facebook and Google Maps.
8. Personalize the Google Maps widget in a manner that it displays the map of London.
9. Change the background image to Coffee Beans.
10. Change the title of the page to Lleida.
11. Change to color of the News tab to purple.
12. Sort widgets on Business tab in following order: Calendar, To Do List, Weather.
13. Delete all three tabs and the title of the page.
14. Change background image to the default value.
15. Log out and press ESC.