

# Web Site Usability Evaluation: An Exploratory Study on the Web Site of Directorate General of Higher Education

Imam Azhari\*, Agus Harjoko\*\*

\* Department of Information Systems, Universitas Ahmad Dahlan

\*\* Department of Computer Science and Electronics, Universitas Gadjah Mada

---

## Keywords:

Usability measures  
effectiveness  
efficiency  
satisfaction  
usability testing

---

## ABSTRACT

Website becomes an integral part how people access information. Thus, it now becomes a challenge to develop a usable and effective website to meet users-needs. Good websites combine both two aspects: aesthetics and ease of use. *Dikti.go.id* site is one of very popular and highly accessed sites among Indonesian universities. It contains the latest news and information related to the organization of higher education in Indonesia such as policy and legislation, scholarship information, grant programs and many various academic competitions for universities. This paper presents the result of evaluation of the usability of the *Dikti.go.id* site which includes aspects of its effectiveness, efficiency, and satisfaction perception. Based on the data analysis, it shows that the effectiveness of the site is high but its level of success is low. This means that there are problems related to the essential way how users navigate the site. This, finally, makes the satisfaction perception of this site be not high enough.

Copyright © 2013 Information Systems International Conference.  
All rights reserved.

---

## Corresponding Author:

Imam Azhari  
Department of Information Systems, Universitas Ahmad Dahlan  
Jalan Prof. Dr. Soepomo, S.H., Janturan, Yogyakarta, Indonesia.  
Email: azhari@uad.ac.id

---

## 1. INTRODUCTION

International Standard Organization (ISO) defines usability as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use [1]. Usability of a web site is a level in which a user finds convenience when conducting exploration, finding information, and interacting with a web site [2].

According to Pearrow [3], when designing a web site, the needs for gaining both a great view and usability are tangible and should be made balanced. A good web site combines aspects of both aesthetics and ease of use. Usability testing is used to measure a web site's effectiveness, efficiency, and satisfaction perception. It is an effective method to evaluate a system design from the user's point of view. Nielsen [4] revealed a fact that the return on investment (ROI) of a usability criterion based re-designed web can increase up to 83%. ROI itself is based on business metrics which include aspects like conversion rate, traffic numbers, user performance, and feature usage targets.

Directorate General of Higher Education's website (*Dikti.go.id*) contains information about management policies of higher education in Indonesia and the latest information on government programs for universities such as further studies scholarship offers, competitions or partnership grant programs, as well as invitation to participate in the activities initiated by *Dikti*. In short, it can be said that this web site functions as a medium of information dissemination from *Dikti* to more than 4000 universities in Indonesia. This role, in turn, causes the site to be very highly accessed by a wide range of user characteristics. However, despite of the very role it holds, there have been very few researches done about it, especially the one related to its usability. Based on that fact, the study of the site's usability is something very essential to do in order to ensure the quality of the user's experience.

This paper presents the result of an evaluation of the usability of *Dikti.go.id* site covering aspects of its effectiveness, efficiency, and satisfaction perception. A usability testing is conducted under a carefully designed scenario. The effectiveness aspect is seen through the user's achievement and accuracy in

completing the given tasks. The effectiveness indicators used is the participant's level of success in completing the tasks. Furthermore, the efficiency aspect is gained by measuring the time taken by each participant in completing the tasks, while the satisfaction perception is measured using a questionnaire provided at the end of the test.

## 2. USABILITY TESTING

Usability - one focus of the study of Human-Computer Interaction (HCI) - is an important thing in judging the ease of use (ease of use) of a product or system. International standard, ISO 9241-11 defines usability as:

... *"The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use."*[1]

Based on the above definition, it is obvious that there are three main dimensions of usability i.e effectiveness, efficiency, and satisfaction. Quesenberry [1] differently defines usability as the 5E dimensions, covering Effective, Efficient, Engaging, Error Tolerant, and Easy to Learn. Meanwhile, Nielsen [5] and Shneiderman [6] define usability dimension with slightly different terms (Table 1).

The interest rate on the usability dimension is a dependent requirement (context of use). This implies that a usability testing can be particularly done and focused on the usability dimensions based on the level of importance it has. Further, the combination among usability dimensions can be used not only as a direction in doing interface design but also helping test the usability appropriately.

Table 1. Definition of usability classification

ISO 9241-11	Quesenberry [1]	Nielsen [5]	Shneiderman [6]
Effectiveness	Effective Error Tolerant	Memorability Errors/Safety	Retention over Time Rate of Errors by Users
Efficiency	Efficiency Easy to Learn	Efficiency Learnability	Speed of Performance Time to Learn
Satisfaction	Engaging	Satisfaction	Subjective Satisfaction

There are a number of methods and techniques used to do an interface evaluation i.e. expert walkthrough (heuristic evaluation), guidelines checklist, cognitive walkthrough, user behavior observation, and questionnaires.

Table 2 compares several methods and techniques of usability testing. Generally, usability methods are divided into two groups i.e. (1) methods of inspection (without end-user) and (2) test methods (with the end-user). The inspection methods include heuristic evaluation, cognitive walkthrough and action analysis (keystroke-level analysis); while method tests comprise thinking aloud, observation, and questionnaires.

Table 2. Comparison of usability testing techniques

	Inspection Methods	Test Methods				
	Heuristic Evaluation	Cognitive Walkthrough	Action Analysis	Thinking Aloud	Field Observation	Questionnaires
Applicably in Phase	All	All	Design	Design	Final testing	All
Required Time	Low	Medium	High	High	Medium	Low
Needed Users	None	None	None	3+	20+	30+
Required Evaluators	3+	3+	1-2	1	1+	1
Required Equipment	Low	Low	Low	High	Medium	Low
Required Expertise	Medium	High	High	Medium	High	Low
Intrusive	No	No	No	Yes	Yes	No

In short, for a web site existence, usability is more important in its nature since it is a very important aspect of how a web site's quality is determined. It functions more than just to develop a beautiful interface aesthetically. Though the fact that aesthetics can affect usability [7] is undeniable, other factors such as aesthetics, functionality, user satisfaction and ease of use have a very important role in developing a usable system as well.

## 3. RESEARCH METHOD

### 3.1 Participants

To be considered 'enough' to perform statistical testing, the number of participants involved in a usability test is often a source of controversy. Principally, to conduct a usability testing, one single participant can be considered sufficient. However, based on the Nielsen's findings [8], it indicates that when an evaluation is based on only one participant, the result is quite small (about 25%). Nielsen, then, recommends around five participants but not less than three [9]. Another expert, Downey [10], though finds that when the

testing is done in the form of a group of users, it will simultaneously allow a test to obtain more reliable results.

In conducting the test, the study involves two groups of testers, each of which consists of 5 people. Based on the participant's experience in utilizing internet, these two groups are divided into two i.e. Low Experience group (LE) and High Experience group (HE). The Low Experience group (LE) consists of participants who use internet for less than three years, less than 15 hours online per week, require help from others when they have computer problems, and take advantage of the browser by default. In contrast, participants who use internet for more than four years, more than 15 hours online per week, are able to optimize the use of browser, able fix the computer problems by themselves and able to help others solve computer problems are categorized into High Experience (HE) group.

### 3.2 Testing Procedures and Data Analysis

The tests are carried out in three phases for each participant: pre-test, test and post-test in done in less than 45 minutes. During the test phase, participants are required to do 8 previously designed activities (Table 3). Each activity is identified under four criteria, namely: *Easy*, *Medium*, *Hard*, *Assist*, and *Fail*. The first three criteria (*Easy*, *Medium*, and *Hard*) are categorized *Success*, and the remaining two (*Assist* and *Fail*) are considered *Fail*. The time duration needed to complete each activity is carefully recorded.

The information about the web site's effectiveness is measured by two sizes. The first is the completion of every activity success rate obtained from the assessment category of success rate of completion of each activity. The second one is task completion rates or *p*. The effectiveness analysis of this study uses the measurements that have been adjusted as the result of measurements made by Lewis and Sauro [11] which says that Laplace method  $((x + 1) / (n + 2))$  is the best estimator.

Table 3. Scenario/Task of usability testing activities

No	Classification	Scenarios/Tasks	Completion Criteria
1	Features	You are a prospective student who will be applying to college. There is information that the Higher Education published a ranking: 50 Indonesian Promising University. You want to know whether the University of Muhammadiyah Malang is one of the promising university.	Participants found the University of Muhammadiyah Malang Muhammadiyah listed one of the promising university.
2	Features	You are a prospective student who will be applying to college. You will find information about college entrance exams (SNMPTN). What is the website address (URL) of SNMPTN?	Participants found the URL of SNMPTN: <a href="http://www.dikti.go.id/snmptn">www.dikti.go.id/snmptn</a>
3	News	You are a lecturer who was looking for information on certification of lecturers for 2008. What is the number of quota to be certified in 2008?	Participants found the number of 12,000 lecturers to be certified in 2008.
4	News	You are a team leader Institutions Competitive Grant Program (PHKI) 2008 of STMIK Putera Batam. Does your institution full proposal pass the evaluation?	Participants found that the proposal of STMIK Putera Batam funded by the Dikti.
5	Rules & Regulations	You are a Dean who are in need of referral rules regarding plagiarism prevention.	Participants found the decree No. 3298/D/T/1999 about plagiarism prevention.
6	Features	Who was the Minister of National Education of Indonesia?	Participants found the History of Higher Education page containing the names of the Minister of National Education.
7	Rules & Regulations	You are a Head of Department who are looking for supervision guidance of the Masters program. Find out about the guidelines!	Participants found the decree No 08/Dikti/Kep/2002.
8	Others	Perform the registration process to this site to get the username and password!	Participants successfully login using their username.

Efficiency aspect is measured by the average completion time of activity for each activity; that is the duration of time it takes participants to complete each activity. At the end of the test, participants are asked to fill out questionnaires of satisfaction perception. The questionnaire is adapted from the System Usability Scale (SUS) and consists of 10 questions with five answers in Likert scale [12,13].

The questionnaire provides an easy-to-understand score from 0 to 100. An SUS score above 68 would be considered above average and anything below 68 is below average [14]. The SUS done after the participants completed the testing session is a participant's direct response after performing the test.

## 4. RESULTS AND DISCUSSION

### 4.1 Effectiveness

As previously explained, the effectiveness is gained from the success rate. The success rate is measured by the level of ease. If participants can successfully complete the screenplay in one trial, it's stated *Easy*. When participants are able to complete successfully the screenplay in the second or third trial, it's

categorized *Medium*. *Hard* category is the condition when participants are only capable of completing the scenario in the fourth trial. At the time of entering the trial, participants are given five clues. If they can successfully complete them all, they're declared in *Assist* scenario. Yet, if they can't successfully complete the tasks, it's stated *Fail*.

Figure 1 shows the distribution of differences in the measurement of success criteria for each scenario. The figure shows that for *Features* scenario (*Promising* and *PHKI*), there is a consistency in both groups. This means that the task is easily carried out by most participants.

In *News* scenario (*SNMPTN* and *Certif*), however, there is a striking difference in two groups. The LE group has difficulty in finding the news in question. In contrast, the HE group can accomplish more effective scenarios, except for *Certif* scenario which shows that 60% of HE participants are assisted to search the news.

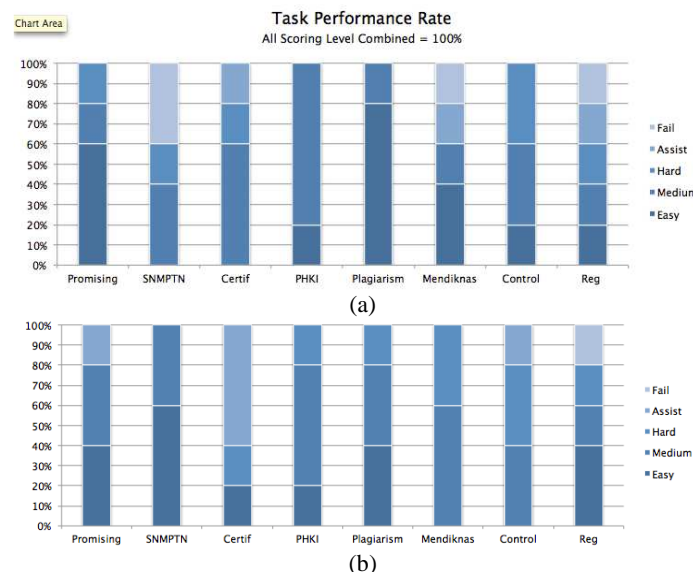


Figure 1. Task Performance Rate: (a) Low Experience, (b) High Experience

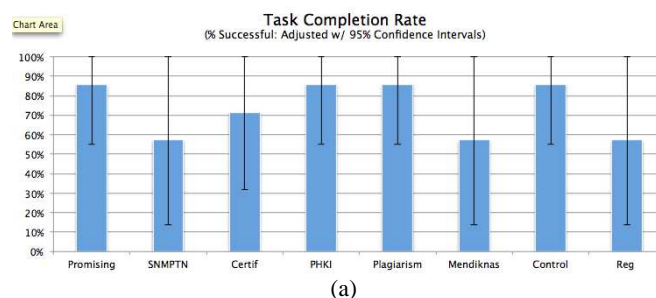
Of *Rules & Regulations* Scenario, the two group's effectiveness in completing the scenario is relatively the same. The difference lies in the scenario of *Plagiarism*. In *Plagiarism* scenario, it seems that this task is more easily solved because there are only 20% of HE group participants are identified *Hard*. In contrast, only 20% of LE group is identified *Easy* in completing the *Control* scenario.

Test results of *Reg* scenario show various success rates in both groups. This indicates that users face a challenge in order to be able to register to this site.

Task completion rate scenario is measured by the proportion of *Success* versus *Fail* (Figure 2). Completion rate for four scenario of 8 usability testing scenarios reached 86% by each group. The average completion rate of LE group was 73% (SD=14), slightly lower than the average rate of completion of HE group (mean 75%, SD=15).

*Features* and *News* scenarios are resolved with completion rates above 50% by both groups. Similarly, *Rules and Regulations* scenarios are solved by completion rates above 70% by both groups. However, it's important to note that the margin of error in these scenarios is quite large.

There is a difference in completion rates between LE and HE groups in *News* scenario. The LE group's completion rate scenario of *SNMPTN* is 57% while the HE group shows completion rate of 86%. The opposite occurs in the scenario of *Certif* i.e. LE group has 71% completion rate and HE group has only 43% completion rate. This suggests that the experience of using internet doesn't influence the completion rate scenario anymore.



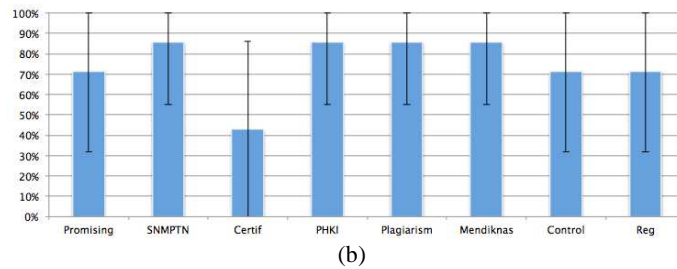


Figure 2. Task Completion Rate: (a) Low Experience, (b) High Experience

In sum, based on completion rate scenario, the effectiveness of *Dikti.go.id* website is fairly high, which is more than 70%. However, the site's success rate is quite low. This is shown through the fact that the success rate scenario of LE group is only 2.7 (SD=1.26) and the HE group is 2.75 (SD=1.08) on a scale of 4.

#### 4.2 Efficiency

Figure 3 shows the average completion time spent for completing each task, including the maximum and minimum values performed by each participant indicated by the error bars on the graph.

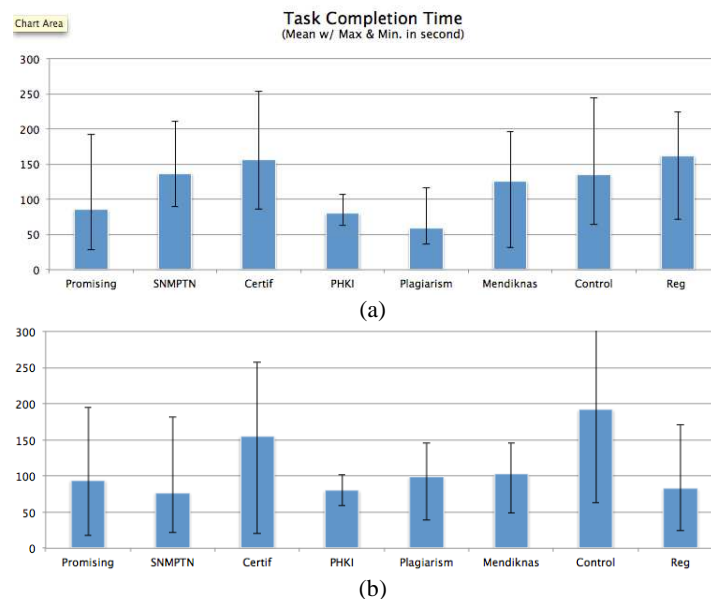


Figure 3. Task Completion Time: (a) Low Experience, (b) High Experience

Based on the gained data, it's obvious that the average completion time scenario of HE group is smaller than that of the LE group (117.64 seconds vs. 110 seconds). Nevertheless, the deviation figure is very large i.e. the HE group's Mean is 110 (SD=73), while the LE group gets 117.64 for the Mean (SD=63).

The fastest time in LE group is in *Plagiarism* scenario (59 seconds, the time spans between 36 up to 116 seconds). Meanwhile, the scenario of *PHKI* becomes the fastest time for the HE group (77 seconds, the time extends between 22 up to 181 seconds). Seen from the length of time spent, LE group spends the longest time in *Reg* scenario (162 seconds, the time reaches between 72 to 224 seconds), while HE group spends its longest time in *Control* scenario (192 seconds, the time spreads between 63 to 332 seconds).

#### 4.3 Satisfaction Perception

The results of the SUS score for the LE group is 46.5 while HE group gets the score of 45.5. This fact indicates that the two groups have similar perceptions about *Dikti.go.id* site i.e. the site is not satisfactory. Based on the question item analysis of SUS questionnaire, it appears that most of the participants assessing *Dikti.go.id* site find it too complex and contains quite a lot of inconsistencies (items no 2 and 6). In terms of navigation, the site is categorized not easy to explore (items no 3, 7, and 8).

#### 5. CONCLUSION



To summarize therefore, seen from the completion rate scenario, the Effectiveness of *Dikti.go.id* website is categorized high but its success rate is quite low. Meanwhile, from the point of view of Efficiency, it indicates that the website's level of efficiency has very large deviation. This means that there are essential problems related to the way how users navigate the site. This, at the end, makes the satisfaction perception

for the site is not high enough. However, it's important to note that the result of this study is descriptive in nature which means that it can't be generalized for other different research settings.

## REFERENCES

- [1] W. Quesenberry, "Dimensions of usability: Defining the conversation, driving the process," *Proceeding of the UPA 2003 Conference* (<http://www.upa.org>), 2003.
- [2] K. Goto and E. Cotler, "Web redesign 2.0: Workflow that works," Berkeley, CA: Peachpit Press, 2005.
- [3] M. Pearrow, "Web site usability handbook," USA: Charles River Media, Inc., 2000.
- [4] J. Nielsen, "Jacob's Nielsen's alert box: Usability ROI declining, but still strong," [cited 2008 January 22]. Available from <http://www.useit.com/alertbox/roi.html>.
- [5] J. Nielsen, "Usability engineering," Cambridge, MA: Academic Press, 1993.
- [6] B. Shneiderman, "Designing the user interface: Strategies for effective human-computer interaction," Reading, MA: Addison-Wesley Longman, 1998.
- [7] M. Kampf, "What is usable is usable. Master thesis presentation," [cited 2008 April 5]. Available from <http://www.marijn.org/downloads/Thesis%20presentation%20M.%20Kampf%2026%20august%202004.ppt>.
- [8] J. Nielsen, "How to conduct a heuristic evaluation," [cited 2006 December 20]. Available from [http://www.useit.com/papers/heuristic/heuristic\\_evaluation.html](http://www.useit.com/papers/heuristic/heuristic_evaluation.html), 2005.
- [9] J. Nielsen, "Why you only need to test with 5 users," [cited 2006 December 20]. Available from <http://www.useit.com/alertbox/20000319.html>, 2000.
- [10] Downey, "Group usability testing: Evolution in usability techniques," *Journal of Usability Studies*, vol. 2 no. 3, pp. 133-144, 2007.
- [11] J.R. Lewis dan J. Sauro, "When 100% really isn't 100%: Improving the accuracy of small-sample estimates of completion rates," *Journal of Usability Studies*, vol. 1 no. 3, pp. 136-150, 2003.
- [12] J. Brook, "SUS – A quick and dirty usability scale", [cited 2008 July 3]. Available from <http://www.usabilitynet.org/trump/documents/Suschapt.doc>.
- [13] J. Sauro, "Performance satisfaction and perception satisfaction," [cited 2013 September 10]. Available from <http://www.measuringusability.com/blog/test-task-sat.php>.
- [14] A. Bangor, P. Kortum, and J. Miller, "Determining what individual SUS scores mean: adding an adjective rating scale," *Journal of Usability Studies*, vol. 4 no. 3, pp. 114-123, 2009.

## BIBLIOGRAPHY OF THE AUTHORS

	<p>Imam Azhari is a faculty member in the Dept. of Information Systems, Universitas Ahmad Dahlan. He teaches Human-Computer Interaction and User Interface Design. His research is focused on the usability studies both on the desktop/web or mobile devices.</p>
	<p>Agus Harjoko completed his doctoral program at the University of New Brunswick, Canada 1996 in the field of Image Proc. and Computer Vision. This time he is listed as a lecturer in the Department of Computer Science and Electronics, Universitas Gadjah Mada.</p>