



Modified LFPP to Improve the Accuracy of Matrix Pairwise Comparison Consistency Index in the Usability Evaluation

Tenia Wahyuningrum^{1*} Azhari Azhari² Suprpto Suprpto²

¹*Department of Informatics, Institut Teknologi Telkom Purwokerto, Indonesia*

²*Department of Computer Sciences, Universitas Gadjah Mada Yogyakarta, Indonesia*

* Corresponding author's Email: tenia@ittelkom-pwt.ac.id

Abstract: Usability evaluation focuses on how well users can learn and use products to achieve their goals. Evaluation of e-commerce website usability can be done using the Fuzzy Analytical Hierarchy Process (FAHP) method. Usability is considered a multi-criteria problem due to its complex structure. Logarithmic Fuzzy Preference Programming (LFPP), a non-linear programming approach, is one of the FAHP methods for weighting criteria. However, deficiencies in the LFPP method, in some cases, result in a consistency value of 0 from a consistent fuzzy comparison matrix, which should be greater than 0. This study modifies the LFPP method with the Modified LFPP (MLFPP) models. Modifications were made as a development of the LFPP method in improving the accuracy of the consistency index (λ^*) before weighting criteria. The MLFPP makes improvements to the functional constraints, while at the same time improving the determination of criteria weights. The MLFPP model uses arctan to limit functions to improve consistency index. The test results show that the MLFPP can improve the accuracy of the consistency index values in the pairwise comparison matrix given by experts to be more than 0, with an average increase of 14.9%. The amount of consistency index (λ^*) changes has improved the usability score by 6.59%.

Keywords: Usability, Arctan, E-commerce, Logarithmic fuzzy preference programming.

1. Introduction

Usability is a part of the software quality relating to the satisfaction of a particular user to achieve a set of tasks in the environment [1]. On an e-commerce website, a good interface model makes users feel comfortable and satisfied [2], and will most likely revisit the website. Conversely, if the website is difficult to use, the user will feel uncomfortable and then leave it [3]. Measurement of usability is considered necessary to utilize every function of the website optimally according to the user experience. It can be said that usability is the critical success factor in business [4]. One measure for the evaluation of usability is the sum of products between the weights of the criteria and the value of each rule [5–8].

Weights and criteria are essential factors in the usability evaluation. An objective assessment of a website product needs the standards used to be determined or calculated appropriately. The current

usability measurement method does not yet have the right uniformity and standard agreement on the software, both in terms of measurement methods, and the many criteria used [9]. Different usability criteria on e-commerce website become a separate issue in the calculation. The effect of criterion weighting is then considered a multi-criteria decision-making issue due to its intricate structure, including the size of each tangible and intangible one [10]. Therefore, an appropriate method is needed to determine the proper weighting based on expert judgment. Based on [11, 12], they considered the weight by the crisp number using AHP method. Based on the AHP method, the expert judgment represented by crisp numbers, but the decision of expert is vague and uncertain, so fuzzy numbers are better to use. Usability is then obtained by the Fuzzy Analytical Hierarchy Process (FAHP) method to support subjective human judgment and hierarchical structure [13]. According to [14], Logarithmic Fuzzy

Preference Programming (LFPP) is one of FAHP techniques to derive crisp weight from the fuzzy comparison matrix. LFPP uses logarithmic functions of natural numbers can produce a single solution in the determination of positive weights. Nevertheless, in experiments conducted by [14], they found some optimal solution (λ^*) experimental results often yield a value of 0. This incident shows a substantial inconsistency between fuzzy assessment and the optimal solution on a fuzzy pairwise comparison matrix.

The inconsistency of fuzzy pairwise comparison matrix will cause invalid weights. The development of the LFPP method is done by modifying the boundary of its objective function using the arctan to replace \ln . By optimizing the accuracy of the consistency index, we get valid weights in the usability evaluations. This improvement can help the developers and owners to see how far the ease of website interacting with users and make website enhancement. This paper is organized as follows: Chapter 2 presents the literature study and an explanation of Triangular Fuzzy Number (TFN) and LFPP. Chapter 3 discusses the proposed method, namely Modified LFPP. Chapter 4 explains the case study, and Chapter 5 presents the conclusion and the suggestion for some areas for further researches.

2. Literature study

2.1 Triangular fuzzy number

Based on Chang (1996), a fuzzy number A on R to be a triangular fuzzy number if its membership function $\mu_A(x): R \rightarrow [0,1]$ is equal to

$$\mu_A(x) = \begin{cases} \frac{(x-l)}{(m-l)}, & x \in [l, m], \\ \frac{(u-x)}{(u-m)}, & x \in [m, u], \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where $l \leq m \leq u$, l is lower, u is upper value of the support of A respectively, and m is modal value. The notation of the triangular fuzzy number (TFN) is (l, m, u) . The use of human language indicates the subjectivity of the assessment then represented by a fuzzy linguistic variable [16–18]. Table 1 shows the nine basic linguistic terms used in this study [13].

2.2 Logarithmic fuzzy preference programming

Wang and Chin (2011) argued that some FAHP settlement cases use the FPP approach resulting in a negative membership degree and arising in multiple

Table 1. Linguistic value and triangular fuzzy number

Crisp Number	Fuzzy Number	Definition
1	(1, 1, 1)	The two comparable elements are equally important (equal).
2	(1, 2, 3)	Assessment is between equally important and slightly more important (equal-moderate).
3	(2, 3, 4)	The first element is slightly more important than the second element (moderate).
4	(3, 4, 5)	Assessment is between slightly more important and more important (moderate-fairly strong).
5	(4, 5, 6)	The first element is more important than the second element (fairly strong).
6	(5, 6, 7)	Assessment is between more important and obviously more important (fairly strong-very strong).
7	(6, 7, 8)	The first element is more important than the second element (very strong).
8	(7, 8, 9)	Assessment of the obvious is more important and absolutely more important (very strong-absolute).
9	(8, 9, 9)	The absolute first element is more important than the second element (absolute).

optimal solutions. A negative value makes the solution expected to be less valid [14, 19–21]. Wang and Chin (2011) refined the FPP method [22] by adding natural logarithms function to improve the weakness. The fuzzy pairwise comparison matrix from expert judgment can be expressed as

$$A = (a_{ij})_{k \times k} = \begin{pmatrix} 1 & (l_{12}, m_{12}, u_{12}) & \dots & (l_{1k}, m_{1k}, u_{1k}) \\ (l_{21}, m_{21}, u_{21}) & 1 & \dots & (l_{2k}, m_{2k}, u_{2k}) \\ \vdots & \vdots & \ddots & \vdots \\ (l_{k1}, m_{k1}, u_{k1}) & (l_{k2}, m_{k2}, u_{k2}) & \dots & 1 \end{pmatrix} \quad (2)$$

where k is the number of criteria, $l_{ij} = 1/u_{ji}$, $m_{ij} = 1/m_{ji}$, $u_{ij} = 1/l_{ji}$ and $0 < l_{ij} < m_{ij} < u_{ij}$ for all $i, j = 1, 2, \dots, k, j \neq i$. To find a crisp priority vector $W = (w_1, w_2, \dots, w_k)^T > 0$ with $\sum_{i=1}^k w_i = 1$ for the fuzzy pairwise comparison matrix [14]. The approximate equation uses natural logarithmic numbers for the improvement of fuzzy pairwise

comparison matrix. The LFPP method is formulated as Eq. (3).

Minimize

$$J = (1 - \lambda)^2 + P \sum_{i=1}^{k-1} \sum_{j=i+1}^k (\delta_{ij}^2 + \eta_{ij}^2) \quad (3)$$

Subject to

$$\begin{cases} x_i - x_j - \lambda \ln\left(\frac{m_{ij}}{l_{ij}}\right) + \delta_{ij} \geq \ln(l_{ij}), \\ i = 1, 2, \dots, k-1; j = i+1, \dots, k, \\ -x_i + x_j - \lambda \ln\left(\frac{u_{ij}}{m_{ij}}\right) + \eta_{ij} \geq -\ln(u_{ij}), \\ i = 1, 2, \dots, k-1; j = i+1, \dots, k, \\ \lambda, x_i \geq 0, i = 1, 2, \dots, k, \\ \delta_{ij}, \eta_{ij} \geq 0, i = 1, 2, \dots, k-1; j = i+1, \dots, k, \end{cases}$$

where $x_i = \ln w_i$ for $i = 1, 2, \dots, k$ and P is a specified large number such as $P = 10^3$. Eq. (4) can be used to calculate the weight of each criterion.

$$w_i^* = \frac{\exp(x_i^*)}{\sum_{j=1}^k \exp(x_j^*)}, i = 1, 2, \dots, k, \quad (4)$$

where $\exp()$ is the exponential function $\exp(x_i^*) = e^{x_i^*}$ for $i = 1, 2, \dots, k$, and k is the number of criteria.

3. Modified logarithmic fuzzy preference programming

Research by Wang and Chin (2011) mentioned that expert judgment is considered inconsistent if the optimal value $\lambda^* = 0$ unless otherwise $\delta^* = \sum_{i=1}^{k-1} \sum_{j=i+1}^k (\delta_{ij}^{*2} + \eta_{ij}^{*2}) = 0$. The more significant λ^* , the stronger the inconsistency between fuzzy judgments. Thus, the values of λ^* and δ^* can be treated as a measure of inconsistency in a fuzzy pairwise comparison matrix which can be expressed as a Proposition 1.

Proposition 1. If $(\delta^* > 0) \wedge (\lambda^* = 0)$ then "strong inconsistency". For example, if $p = (\delta^* > 0) \wedge (\lambda^* = 0)$ and $q = \text{"strong inconsistency"}$, then the proposition can be written as $p \rightarrow q$. The contrapositive of the **Proposition 1** is $\neg q \rightarrow \neg p$. It also can be explained as

$$\begin{aligned} \neg q \rightarrow \neg((\delta^* > 0) \wedge (\lambda^* = 0)); \\ \neg q \rightarrow \neg(\delta^* > 0) \vee \neg(\lambda^* = 0); \\ \neg q \rightarrow (\delta^* \leq 0) \vee (\lambda^* \neq 0) = \text{true}. \end{aligned}$$

So, if $r = (\delta^* \leq 0)$, and $s = (\lambda^* \neq 0)$, then $\neg q \rightarrow r \vee s$.

Table 2. The truth table of "OR" operations

r	s	$\neg q \rightarrow r \vee s$
0	0	0
0	1	1
1	0	1
1	1	1

Because the $\delta^* = \sum_{i=1}^{k-1} \sum_{j=i+1}^k (\delta_{ij}^{*2} + \eta_{ij}^{*2}) = 0$ is always greater than equal to 0, then the proposition r is always false. For the statement to be true, the proposition of s must be true. It can be said that the consistent of the fuzzy pairwise matrix must be not equal zero $\lambda^* \neq 0$. The condition of this proposition complies with line two from truth table that shown in Table 2.

In a numerical experiment conducted by [14] using a pairwise fuzzy comparison matrix has several calculations that yield the optimal value $\lambda^* = 0$. The fuzzy pairwise comparison matrix used in (Wang and Chin, 2011) had been calculated and expressed consistently utilizing the Extent Analysis (EA) method by [23]. The result of optimal value λ^* contradicts the **Proposition 1**, so a deeper research to investigate whether the resulting value represents the fuzzy inconsistency of the assessment is needed. Natural number logarithmic functions in the LFPP method derives a single solution in determining the value of importance weights despite its less precise amount.

Thus, the boundary in the LFPP objective function should be sought with other alternatives, such as trigonometric functions. The function of arcsin and arccos is considered to be less relevant to use, due to its domain limitation in value 1, whereas the use of triangular fuzzy numbers consists of a combination of divisions of numbers 1 to 9. On the other hand, the logarithmic of natural number (\ln) and \arctan has some similarity of curve shape especially in quadrant 1 (if x and y positive) showed in Fig. 1. Example: Using the plot function on MATLAB software, the graphic function $f(x) = \ln(x)$, $f(x) = \arctan(x)$, where $1 \leq x \leq 10$ can be shown in Fig. 1.

The arctan function as a substitute for \ln in Eq. (3) can be explained by using the positive number (1 to 9) which is represented in the fuzzy number in Table 3. Based on Table 3, for $l = 1$ and $m = 1$, $\ln(m/l) = \ln(1)$ obtained 0, and $\arctan(1)$ obtained 0.785. The significant differences allegedly caused the λ^* become 0, although the fuzzy pairwise matrix is consistent. While the value of arctan function shows an amount more than 0, even close to 1. Therefore, arctan trigonometric functions are considered more

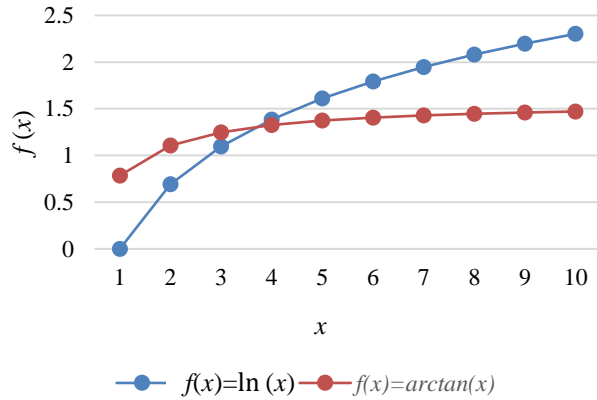


Figure. 1 Graph of ln and arctan function

appropriate to replace ln function in LFPP constraints. The use of fuzzy numbers in assessments by experts from practitioners or academics who know usability issues can be shaped into a matrix of pairwise comparisons (see Eq. (2)).

Table 3. Function of ln and arctan on fuzzy number

<i>l</i>	<i>m</i>	<i>u</i>	ln (<i>m/l</i>)	ln (<i>u/m</i>)	arctan (<i>m/l</i>)	arctan (<i>u/m</i>)
1	1	1	0.000	0.000	0.785	0.785
1	2	3	0.693	0.405	1.107	0.983
2	3	4	0.405	0.288	0.983	0.927
3	4	5	0.288	0.223	0.927	0.896
4	5	6	0.223	0.182	0.896	0.876
5	6	7	0.182	0.154	0.876	0.862
6	7	8	0.154	0.134	0.862	0.852
7	8	9	0.134	0.118	0.852	0.844
8	9	9	0.118	0.000	0.844	0.785

To maximize the accuracy value of λ^* , we proposed a triangular fuzzy judgment a_{ij} , by substituting the function ln to be arctan within its boundaries (this method called Modified LFPP or MLFPP).

Minimize

$$J = (1 - \lambda)^2 + P \sum_{i=1}^{k-1} \sum_{j=i+1}^k (\delta_{ij}^2 + \eta_{ij}^2) \quad (5)$$

Subject to

$$\begin{cases} x_i - x_j + \lambda \arctan\left(\frac{m_{ij}}{l_{ij}}\right) + \delta_{ij} \geq \arctan(l_{ij}), \\ i = 1, 2, \dots, k - 1; j = i + 1, \dots, k, \\ -x_i + x_j - \lambda \arctan\left(\frac{u_{ij}}{m_{ij}}\right) + \eta_{ij} \geq -\arctan(u_{ij}), \\ i = 1, 2, \dots, k - 1; j = i + 1, \dots, k, \\ \lambda, x_i \geq 0, i = 1, 2, \dots, k, \\ \delta_{ij}, \eta_{ij} \geq 0, i = 1, 2, \dots, k - 1; j = i + 1, \dots, k, \end{cases}$$

If $x_i^*, i = 1, 2, \dots, k$ is the optimal solution for Eq. (5), then the normalized weight for fuzzy pairwise comparison matrix $A = (a_{ij})_{k \times k}$ is

$$w_i^* = \frac{\tan(x_i^*)}{\sum_{j=1}^k \tan(x_j^*)}, i = 1, 2, \dots, k, \quad (6)$$

where w_i^* is the weight of each criterion, and k is the number of criteria.

Using MLFPP formula (see Eq. (5)), we calculated the consistency index and compared the accuracy of both methods. We use 45 the fuzzy comparison matrix based on [14, 24–26]. In this study, the examination of the consistency ratio (CR) value also was done by the Analytical Hierarchy Process (AHP) method. CR used to ensure that matrices are consistent.

Based on Fig. 2, we were choosing the matrices to be analysed when CR is less than 10%. Fig. 2 shows that 12 fuzzy paired comparison matrices have a CR value of more than 0.1 (inconsistent) so that the form is excluded from the analysis. Analysis of data normality in this study using $(45 - 12 = 33)$ consistent pairwise comparison matrices with a value of $P = 1$. index value greater than 0.

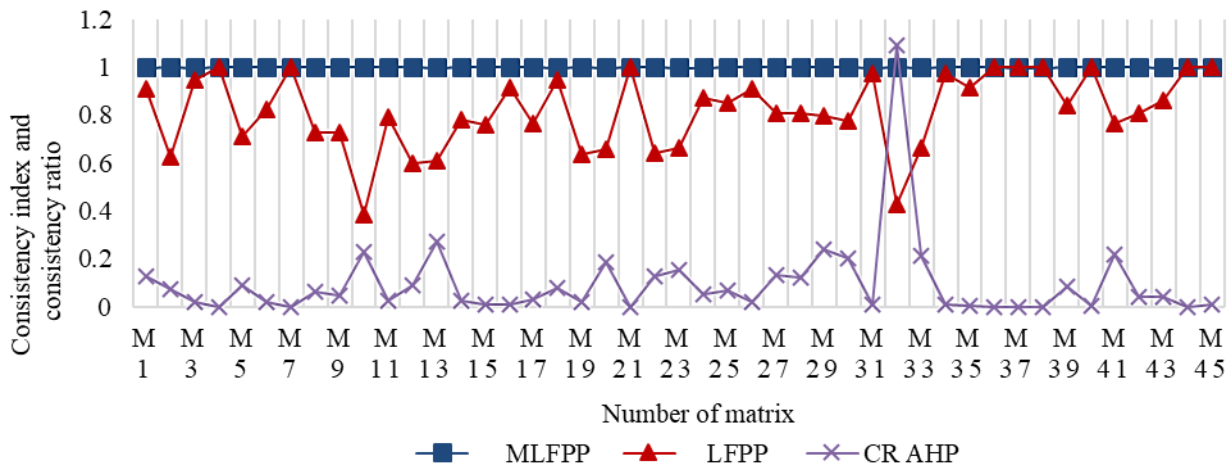


Figure. 2 Comparison of the consistency index, $P=1$

Fig. 2 shows that the two models compared, the MLFPP model more often reaches the maximum λ -value (close to 1), this is because the arctan function is limited, the value is between $(-\pi, \pi)$. Unlike the ln function, the solution can reach an infinitive value. It can be said that the arctan function produces computational amount that tend to be stable and do not cause an overflow.

Before conducting a comparison test of the two samples using the LFPP and MLFPP methods, we perform data normality testing using Kolmogorov-Smirnov test and Shapiro-Wilk test. Data normality testing is a common practice before the statistical method. Normality test is one part of the test requirements for data analysis or commonly called classical assumptions. The purpose of the normality test is to find out whether the data distribution follows or approaches the normal distribution. For a data normality test, the hypothesis are as follows:

- H_0 : Data follow a normal distribution.
- H_a : Data do not follow a normal distribution.

Using SPSS software, rejection and acceptance of hypothesis are seen from Asymp. Sig. (2-tailed) value. Because the number of samples is less than 50, the test results used are the Shapiro-Wilk test. If the Asymp. Sig. (2-tailed) value of Shapiro-Wilk test is greater than 0.05, the data are normal. If it is below 0.05, the data significantly deviate from a normal distribution. Based on Table 4, the result shows that the Asymp. Sig. (2-tailed) value of LFPP and MLFPP method are 0.004 and 0.000, thus the null hypothesis is rejected. The conclusion from the Shapiro-Wilk test shows that the data do not follow a normal distribution.

Wilcoxon signed rank test is a non-parametric test to measure the significance of differences between two groups of pairs of interval data but not normally distributed. The proposed hypothesis are as follows:

- H_0 : There is no significant difference between the consistency index of LFPP and MLFPP.
- H_a : There is a significant difference between the consistency index of LFPP and MLFPP.

Table 4. Normality test of consistency index

Method	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
MLFPP	0.468	33	0.000	0,537	33	0.000
LFPP	0.146	33	0.074	0,896	33	0.004

a. Lilliefors Significance Correction

Table 5. Wilcoxon signed rank test

MLFPP LFPP	Results	N	Mean Rank	Sum of Ranks
	Negative Ranks	3a	2.00	6.00
	Positive Ranks	24b	15.50	372.00
	Ties	6c		
	Total	33		

a MLFPP < LFPP, b MLFPP > LFPP, c MLFPP = LFPP

Table 6. Test statistics^b

Method	Z	Asymp. Sig. (2-tailed)
MLFPP - LFPP	-4.398a	0.000

a Based on negative ranks, b Wilcoxon Signed Ranks Test

Table 5 shows the Wilcoxon signed rank test result. Output interpretation of Wilcoxon test is explained as follows:

The negative ranks between the value of index consistency produced by the LFPP and MLFPP methods are 3; 2.00; 6.00. Value 3 shows that there are three pieces of data which have decreased from LFPP values to MLFPP values. The decrease mean rank is 2.00 and the sum rank is 6.00.

The positive rank between the value of the consistency of the index produced by the LFPP and MLFPP methods is 24; 15.50; 372.00. There are 24 pieces of data that have increased from the LFPP value to the MLFPP value. The increase mean rank is 15.50 and the sum rank is 372.00.

Ties are the same number of values produced by the LFPP and MLFPP methods. There are six data that have the same value in both methods.

The rules of rejecting and accepting the hypothesis can be seen from the Asymp. Sig. Value (2-tailed). If the Asymp. Sig. Value (2-tailed) is greater than 0.05 then H_0 is accepted and if the Asymp. Sig. value (2-tailed) less than 0.05 then H_a is accepted. Based on the Table 6, the Asymp. Sig. Value is 0.000; it means that H_0 is rejected. We can conclude that there is a significant difference mean of optimal value between the LFPP and MLFPP method. The MLFPP approach on a consistent fuzzy pairwise comparison matrix will always have an optimum value above 0. This result is in line with the "OR" operating legal statement.

After knowing there is a significant difference in the consistency index produced by LFPP and MLFPP, then Table 7 explains the magnitude of the average difference between the two models. Table 7 shows that the consistency index (λ^*) value in the pairwise comparison matrix given by the expert after the

Table 7. Descriptive statistics of consistency index

Method	N	Mean	Std. Deviation	Min	Max
MLFPP	33	0.999	0	0.999	1.000
LFPP	33	0.869	0.124	0.599	1.000

modification of the LFPP method becomes more than 0, with an average increase of λ is 0.13 points or 14.9%.

4. Case study

We apply the MLFPP method in evaluating e-commerce usability using eight evaluation stages as illustrated in Fig. 3.

Fig. 3 shows a flowchart of the MLFPP method to measure the usability that consisting of eight stages, i.e., determining usability criteria, developing the hierarchical model, developing pairwise comparison matrix, calculating consistency index, weighting criteria, normalizing data, calculating usability, determining severity, and recommendation.

Establishing usability criteria is the first step of evaluation. Developers and usability experts define essential rules in the assessment. The literature study activities are often used to collect several papers relating to usability e-commerce, to look for the right criteria in measurement. After choosing the proper criteria, the next is to build a model hierarchy based on the taxonomy specified. We developed 14 criteria which represented of usability [27].

Data collection on usability evaluation was done by several ways, i.e., direct measurement, and questionnaire. Some criteria such as *load time* (C_1), *response time* (C_2), *page rank* (C_3), *frequency of update* (C_4), *traffic* (C_5), *sitemap* (C_6), *design optimization* (C_7), *size* (C_8), *backlink* (C_9), *number of*

items (C_{10}), *accessibility error* (C_{11}), *markup validation* (C_{12}), and *broken link* (C_{13}) use a direct measurement taken by using diagnostic tools such as Pingdom.com, Bitcathca.com, Power Mapper, Google page insight, W3C Markup validation service and Alexa. While the criteria associated with ease of use can be collected using System Usability Scale (SUS) questionnaires. The SUS questionnaire has the advantage of being an established evaluation tool for software quality measurement [28–30]. The SUS questionnaire uses the Likert scale 1-5, the value one indicates "strongly disagree", then increases as the number increases, and finally the value five means "strongly agree".

We develop a fuzzy pairwise comparison matrix based on an expert judgment about the importance of each criterion. The comparison matrix is then used for consistency and weight calculation.

Minimize

$$J = (1 - \lambda)^2 + P \sum_{i=1}^{13} \sum_{j=i+1}^{14} (\delta_{ij}^2 + \eta_{ij}^2)$$

Subject to

$$\begin{cases} x_1 - x_2 + \lambda \arctan\left(\frac{1.26}{1}\right) + \delta_{12} \geq \arctan(1), \\ -x_1 + x_2 + \lambda \arctan\left(\frac{1.44}{1.26}\right) + \eta_{12} \geq -\arctan(1.44), \\ x_1 - x_3 + \lambda \arctan\left(\frac{2.41}{1.82}\right) + \delta_{13} \geq \arctan(1.82), \\ -x_1 - x_3 + \lambda \arctan\left(\frac{2.88}{2.41}\right) + \eta_{13} \geq -\arctan(2.88), \\ \vdots \\ x_{13} - x_{14} + \lambda \arctan\left(\frac{1}{0.69}\right) + \delta_{1314} \geq \arctan(0.69), \\ -x_{13} - x_{14} + \lambda \arctan\left(\frac{1.44}{1}\right) + \eta_{1314} \geq -\arctan(1), \\ \lambda, x_1, x_2, \dots, x_{14}, \delta_{12}, \eta_{12}, \delta_{13}, \eta_{13}, \dots, \delta_{1314}, \eta_{1314} \geq 0, \end{cases}$$

Determined $P = 10^{-1}$ then the consistency index λ^* is 0.87 ($\lambda^* > 0$ are considered consistent). The value of x can be explained as $x_1 = 2.06; x_2 = 2.05; x_3 = 1.58; x_4 = 0.99; x_5 = 0.83; x_6 = 0.97; x_7 = 0.91; x_8 = 0.91; x_9 = 0.62; x_{10} = 0.74; x_{11} = 0.74; x_{12} = 0.45; x_{13} = 0.65; x_{14} = 0.50$.

Because there is an x value greater than 1, the x value is transformed into an interval $[0,1]$ so that the weight obtained is positive using Eq. (7).

$$x_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)}, i = 1, 2, \dots, k, \quad (7)$$

where $\min(x_i)$ is the minimum value of the whole

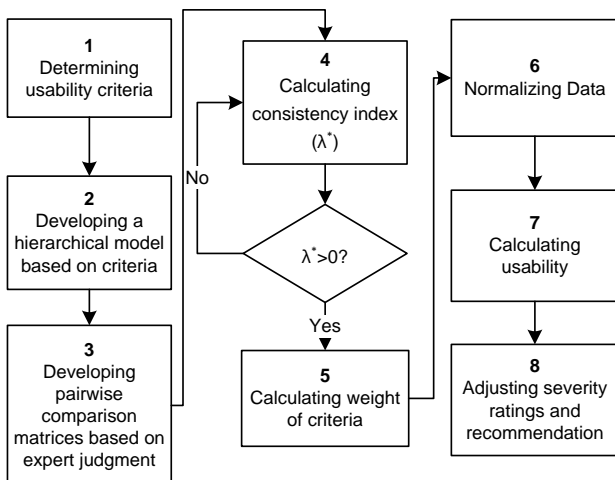


Figure. 3 Usability evaluation method

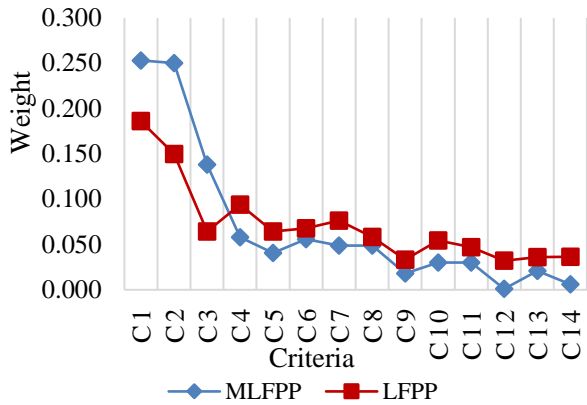


Figure. 4 Comparison of criteria weights

data x_i subtract 0.01, and $\max(x_i)$ is the maximum value of the entire data. The results of the transformation of values as follows, as $x_1 = 1.00$; $x_2 = 0.99$; $x_3 = 0.70$; $x_4 = 0.34$; $x_5 = 0.25$; $x_6 = 0.33$; $x_7 = 0.29$; $x_8 = 0.29$; $x_9 = 0.11$; $x_{10} = 0.19$; $x_{11} = 0.19$; $x_{12} = 0.01$; $x_{13} = 0.13$; $x_{14} = 0.04$.

The next step is calculating weight. Based on the results of the weight calculation using Eq. (6), it can be said that the improvement of the method by modifying the LFPP method to the MLFPP models results in differences in the criteria weights. The difference in the weights of the criteria produced by each model is more clearly shown in Fig. 4.

MLFPP and LFPP models show almost the same pattern, although they have different values. Both models have the lowest weight in the 12th criterion, the markup validation criteria (w_{12}). This opposite pattern raises different priority priorities. The normalized of the data result can be expressed using the Linear Weightage Model (LWM) [13] such as the Eq.(8).

$$r_{ij} = \begin{cases} \frac{\max-x_{ij}}{\max-\min}, & \text{for maximum threshold} \\ \frac{x_{ij}-\min}{\max-\min}, & \text{for minimum threshold} \end{cases} \quad (8)$$

where r_{ij} is the normalized value and x_{ij} is criteria value. Max is the maximum value of the particular criteria among all websites, min is minimum value of the same criteria among the whole set of websites. The threshold for the “smaller is better” criteria must be maximum, and the threshold for the “larger is better” criteria must be minimum.

The final step of this method is to determine severity ratings based on usability value obtained. Eq. (9) used to calculating usability, where l is the number of alternatives, k is the number of criteria, r_{ij} is normalized value, and w_j is weight of criteria.

$$usability = \sum_{i=1}^l \sum_{j=1}^k w_j \times r_{ij}; \quad (9)$$

$$i = 1, 2, \dots, l, j = 1, 2, \dots, k$$

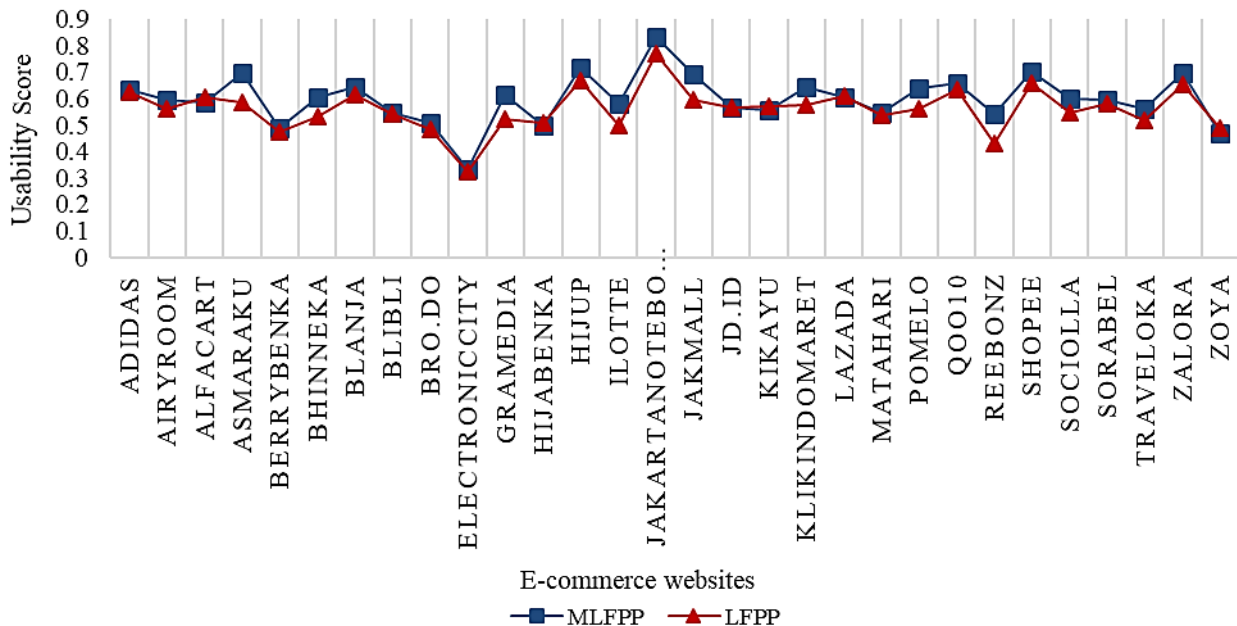


Figure. 5 Comparison of the usability score

Level severity helps developers to get recommendations on which criteria that need to be improved in term of quality. Based on Tullis and Albert (2013), the severity ratings consist of four level, i.e., unusable (90-100%), severe (51-89%), moderate (11-50%) and irritant (0-10%) [31]. If the value of the severity ratings of usability level is high, it is necessary to recommend improvement on the website design.

Method modification was applied to 30 e-commerce websites to evaluate the resulting usability score. Thus, we tested improvement in the usability score (US) before and after the method modification. Based on Fig. 5, it can be seen that the website with the highest usability score is Jakartanotebook, while the lowest is Electroniccity. Electroniccity has a low usability value, so the severity rating is high and requires improvement recommendations.

Recommendations given for Electroniccity is to increase page load times by making fewer HTTP requests, using cookie-free domains, and adding expired headers. Reducing the number of components on a page reduces the HTTP requests needed to create the page, resulting in faster page loading. Some ways to reduce the elements include combining files, combining several scripts into one script, combining several CSS files into one style sheet, and using CSS sprites and image maps.

The next hypothesis is to measure a significant difference between the usability score before and after the method modification. Comparative testing uses a paired sample t-test, which is a test used to compare two averages in one sample with two different events, assuming the data is standard. The proposed hypothesis are as follows:

H_0 : There is no significant difference between the usability score of LFPP and MLFPP.

H_a : There is a significant difference between the usability score of LFPP and MLFPP.

Testing the normal distribution using the Kolmogorov-Smirnov test produces Sig. 0.200* whereas in the Shapiro-Wilk test the Sig. 0.414 (MLFPP); 0.483 (LFPP) according to Table 8. Thus, the overall value of Sig. > 0.05, this shows that all three models produce usability scores that are normally distributed so that they meet the requirements of the t-test.

Table 9 shows the statistics of usability scores for both models. Obtained usability score before method modification (LFPP) is 0.567, while after method modification (MLFPP) 0.598. The standard deviations obtained were 0.081988 (LFPP), and 0.092531 (MLFPP). Based on the average collected,

Table 8. Normality test of usability score

Method	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
MLFPP	0.103	30	0.200*	0.965	30	0.414
LFPP	0.78	30	0.200*	0.968	30	0.483

a. Lilliefors Significance Correction

Table 9. Paired sample statistic

Method	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 MLFPP	0.59850	30	0.092531	0.016894
LFPP	0.56133	30	0.081988	0.014969

Table 10. Paired samples correlations

Method	N	Correlation	Sig.
Pair 1 MLFPP & LFPP	30	0.907	0.000

Table 11. Paired samples t-test

Paired Differences			t	df	Sig.(2-tailed)
Mean	Std. Dev	Std. Error Mean			
0.03716	0.039098	0.007138	5.207	29	0.000

method modification increased the usability score of around 6.59%.

Table 10 shows the output of paired samples correlations. The correlation value between the two variables is 0.907, and the significance value gained 0.000. Because of the amount of Sig. <0.05, it can be concluded that the relationship between the two variables is very real/close.

Table 11 shows paired samples t-test on the MLFPP and LFPP models. The difference in the average usability score in the two models is 0.037167, with a standard deviation difference of 0.039089. Rejection and acceptance of the hypothesis can be seen from the t value or the Sig. (2-tailed). H_0 is accepted if it meets the requirements $-t \text{ table} < t \text{ count} < t \text{ table}$ or $\text{Sig. (2-tailed)} > 0.05$, and the opposite applies.

Table 11 shows the calculated t value of 5.207, while the t table is obtained from the t distribution table by looking at the two-party test, with a significance of 5% and degrees of freedom/df = 30-1 = 29. So, we get the value of t table = 2.045. Because t value is more than t table, H_0 is rejected. Based on Sig. (2-tailed) obtained a value of 0.000 <0.05 then H_0 is rejected. Thus, it can be concluded that the modification of the LFPP method using the MLFPP model has a significant impact on usability score. The MLFPP model not only significantly increased the

consistency index, but also increased the usability score by 0.037167 points or 6.59%.

5. Conclusion

The weakness of the LFPP method in determining the consistency matrix (called the consistency index) has been improved by modifying the LFPP to increase the accuracy of the value of the consistency index. The modification of the usability evaluation model on the e-commerce website has modifying the functional limitations of the LFPP model into a new model called MLFPP.

The accuracy value of the consistency index has increased with an average increase 14.9%. The usability score comparison shows that based on the average obtained, the method modification increased a usability score 6.59%. Future work is possible in calculating the complexity of the algorithm in determining the P value on the MLFPP objective function.

Acknowledgments

This research was supported by the Doctoral grant of the Indonesia endowment fund for education (LPDP) ministry of Finance Republic of Indonesia (No. PRJ-4811/LPDP.3/2016).

References

- [1] H. Yang, "Measuring software product quality with ISO standards base on fuzzy logic technique", *Advances in Intelligent and Soft Computing*, Vol. 137, pp. 59–67, 2012.
- [2] Restyandito and E. Kurniawan, "The Effect of Screen Size and Interaction Style on Mobile Device Usability", *Engineering Letters*, Vol. 25, No. 4, pp. 6, 2017.
- [3] N. B. N. Rozali and M. Y. B. Said, "Usability testing on government agencies web portal: A study on Ministry of Education Malaysia (MOE) web portal", In: *Proc. of Malaysian Software Engineering Conference*, pp. 37–42, 2016.
- [4] H. Roder, "Specifying usability features with patterns and templates", In: *Proc. of International Workshop on Usability and Accessibility Focused Requirements Engineering*, pp. 6–11, 2012.
- [5] T. P. Bowen, G. B. Wigle, and J. T. Tsai, *Specification of Software Quality Attributes*, New York, 1985.
- [6] J. S. Challa, A. Paul, Y. Dada, V. Nerella, and P. R. Srivastava, "Quantification of Software Quality Parameters Using Fuzzy Multi Criteria Approach", In: *Proc. of International Conference on Process Automation, Control and Computing*, pp. 1–6, 2011.
- [7] S. K. Dubey, A. Gulati, and P. A. Rana, "Usability Evaluation of Software Systems using Fuzzy Multi-Criteria Approach", *International Journal Computer Science Issues*, Vol. 9, No. 3, pp. 404–409, 2012.
- [8] S. K. Dubey and S. Pandey, "Measurement of Usability of Office Application Using a Fuzzy Multi-Criteria Technique", *International Journal Information Technology and Computer Science*, Vol. 4, pp. 64–72, 2015.
- [9] H. R. Hartson, T. S. T. Andre, and R. R. C. Williges, "Criteria for evaluating usability evaluation methods", *Int. J. Hum. Comput. Interact.*, Vol. 13, No. 4, pp. 1–35, 2001. doi: 10.1207/S15327590IJHC1304_03
- [10] Y. C. Tang and T. W. Lin, "Application of the fuzzy analytic hierarchy process to the lead-free equipment selection decision", *Int. J. Bus. Syst. Res.*, Vol. 5, No. 1, p. 35, 2011. doi: 10.1504/IJBSR.2011.037289
- [11] A. L. M. Villota, *Usability of Websites*, University of Birmingham, 2009.
- [12] B. Zhao and Y. Cheng, "Research on B2C E-commerce Website Service Quality Evaluation Based on Analytic Hierarchy Process", In: *Proc. of International Conference on Information Science and Technology*, pp. 364–367, 2014.
- [13] P. D. D. Dominic and H. Jati, "A comparison of Asian airlines websites quality: using a non-parametric test", *International Journal Business Innovation and Research*, Vol. 5, No. 5, 2011.
- [14] Y. M. Wang and K. S. Chin, "Fuzzy analytic hierarchy process: A logarithmic fuzzy preference programming methodology", *International Journal of Approximate Reasoning*, Vol. 52, No. 4, pp. 541–553, 2011.
- [15] D.-Y. Chang, "Applications of the extent analysis method on fuzzy AHP", *European Journal of Operational Research*, Vol. 95, No. 3, pp. 649–655, 1996.
- [16] A. R. Afshari, R. Yusuff, and A. R. Derayatifar, "Project manager selection by using fuzzy simple additive weighting method", In: *Proc. of International Conference on Innovation, Management and Technology Research*, pp. 412–416, 2012.
- [17] C. C. Sun, "A performance evaluation model by integrating fuzzy AHP and fuzzy TOPSIS methods", *Expert Systems with Applications*, Vol. 37, No. 12, pp. 7745–7754, 2010.
- [18] T. C. Wang and Y. H. Chen, "Applying fuzzy linguistic preference relations to the improvement of consistency of fuzzy AHP", *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 3, 2020 DOI: 10.22266/ijies2020.0630.36

- Information Sciences*, Vol. 178, No. 19, pp. 3755–3765, 2008.
- [19] S. Dožić, T. Lutovac, and M. Kalić, “Fuzzy AHP approach to passenger aircraft type selection”, *Journal of Air Transport Management*, Vol. 68, 2017.
- [20] M. Balouchi and E. Khanmohammadi, “Using logarithmic fuzzy preference programming to prioritization social media utilization based on tourists’ perspective”, *Foundations of Management*, Vol. 7, No. 1, pp. 7–18, 2015.
- [21] M. Momeni, A. Sasani, M. R. Fathi, and E. Khanmohammadi, “Applying Logarithmic Fuzzy Preference Programming for Ranking of Effective Organizational Factors on Creativity : A Case Study Mansour Momeni Faculty of Management University of Tehran”, *International Journal of Business and Social Science*, Vol. 3, No. 14, pp. 83–95, 2012.
- [22] L. Mikhailov, “Deriving priorities from fuzzy pairwise comparison judgements”, *Fuzzy Sets and Systems*, Vol. 134, No. 3, pp. 365–385, 2003.
- [23] M. Celik, I. D. Er, and A. F. Ozok, “Application of fuzzy extended AHP methodology on shipping registry selection : The case of Turkish maritime industry”, *Expert Systems with Applications*, Vol. 36, No. 1, pp. 190–198, 2009.
- [24] C. Kahraman, U. Cebeci, and D. Ruan, “Multi-attribute comparison of catering service companies using fuzzy AHP: The case of Turkey”, *International Journal of Production Economics*, Vol. 87, No. 2, pp. 171–184, 2004.
- [25] L. Mikhailov, “Group prioritization in the AHP by fuzzy preference programming method”, *Computers and Operations Research*, Vol. 31, No. 2, pp. 293–301, 2004.
- [26] R. K. Goyal and S. Kaushal, “Effect of utility based functions on fuzzy-AHP based network selection in heterogenous wireless networks”, In: *Proc. of International Conference on Recent Advances in Engineering & Computational Sciences*, pp. 1–5, 2015.
- [27] T. Wahyuningrum, Azhari, and Suprpto, “A Comparison of Extent Analysis and Fuzzy Preference Programming for Evaluating B2C Website Usability”, In: *Proc. of International Conference on Control System, Computing and Engineering*, pp. 150–155, 2018.
- [28] J. Brooke, “SUS : A Retrospective”, *Journal of Usability Studies*, Vol. 8, No. 2, pp. 29–40, 2013.
- [29] J. R. Lewis and J. Sauro, “The Factor Structure of the System Usability Scale”, In: *Proc. of International Conference on Human Centered Design*, pp. 94–103, 2009.
- [30] J. J. R. Lewis, “Revisiting the Factor Structure of the System Usability Scale”, *Journal of Usability Studies*, Vol. 12, No. 4, pp. 183–192, 2017.
- [31] T. Tullis and B. Albert, *Measuring The User Experience*. USA: Morgan Kaufmann, 2013.