

Biljana Rondović¹
Julija Cerović
Tamara Đuričković
Boban Melović

THE IMPORTANCE OF OBSERVING THE DIFFERENCE IN WEBSITE EVALUATIONS OBTAINED FROM DIFFERENT PERSPECTIVES

Article info:

Received 07.03.2017

Accepted 10.04.2017

UDC – [519.8+005.33](004.738)

DOI – 10.18421/IJQR11.02-11

Abstract: *The purpose of this paper is to show how website evaluations differ if performed by different evaluators despite the same methodological framework being used for the assessment.*

A quantitative analysis of websites was performed according to the modified evaluation framework based on the 2QCV3Q-model. The AHP method was used for the verification of the weight criteria, while a single-factor ANOVA was used to check the difference in the average scores of the evaluations performed by three groups of evaluators

The results obtained by the single-factor ANOVA showed a statistically significant difference in the average scores of internal evaluation when compared to the evaluation done by clients and independent evaluators.

The results obtained can be used in the banking industry as guidelines and recommendations for the future design or redesign of websites, as well as a literary base for comparative analyses of this kind in developing countries.

Keywords: *Website Evaluation, Website Service, Website Quality*

1. Introduction

Due to an increase in the number of users of bank websites and the clients preference towards a so-called *permission* approach on the one hand, and the growing needs of banks to use web platforms to promote their services and to generate income (Xue et al., 2011; Trujillo-Ponce, 2013), on the other, the last decade has seen various efforts on the part of the professional and scientific public to draw attention to both the potential (Hernando and Nieto, 2007; Callaway, 2011)

and also the limitations of websites (Kesharwani and Singh, 2012).

Given that, on the one hand, in the digital economy, from the viewpoint of banks, websites have been recognized as the most important internet marketing service, and that, on the other, it is the most susceptible platform in terms of customer perception, it is clear that website quality metrics affects the profitability of banks to a great degree.

Although it is undisputed that the mere existence of a website implies the existence of metrics and the application of measuring instruments, marketing sectors generally seem to ignore the recommendations of experts to use methodologically based

¹ Corresponding author: Biljana Rondović
email: biljaro@ac.me

approaches for these purposes. Only by applying an effective methodology can quality website evaluation be performed, the issue of subjective human assessments reduced, and as a consequence, the right actions taken towards the improvement of overall website quality.

An assessment of website quality requires the selection of criteria and their attributes which describe what is expected from a website. The set of criteria, attributes and their relationships constitutes a model for the evaluation of the site.

There is no universally applicable method of evaluation, that is, the evaluation framework and criteria selection should be tailored to the target market and the industry whose sites are evaluated. Selection criteria for the evaluation of sites must be different and dependant on what the goal metric is.

Most of these methods are based on hierarchical lists of criteria and their attributes. The criteria can be assigned weights, that is, the criteria can be weighted.

There is extensive research literature on the application of specific website evaluation models in various industries such as education (Manzari and Trinidad 2013), e-government (Burmaoglu and Kazancoglu, 2012), e-commerce (Davidavičienė and Tolvaišas, 2011), banking (Ariff et al., 2013; Klaus and Nguyen, 2013) and so on. A full understanding of industrial differences would be a prerequisite for the effective design of commercial web sites Huizingh, E. (2002). The majority of these papers recognize the importance of website evaluation from the perspective of the user. Only a small number of authors have recognized the importance of evaluation both from the users and from the employees' perspectives in their research. (Molla and Licker, 2001; Zhu, 2004; Quaddus and Achjari, 2005), but no one has done consolidated research on the topic as of yet. The lack of consolidated research prompted us to conduct research on the banking sector and eliminate the literal gap in terms of a

multidimensional approach to evaluation.

The paper features the results of an internal and external evaluation based on the modified 2QCV3Q model (Mich et al., 2003), complemented by the results of an evaluation performed by the authors, acting as independent evaluators. A single-factor analysis of variance was performed to determine the variation in average scores. The 2QCV3Q model attributes have been adapted to the characteristics of the banking sector and unlike the original model, weight coefficients have been allocated to them. The AHP method was used to confirm the correctness of the assigned weight coefficients.

The rest of the paper is organized as follows: In section 2, reference is made to the findings of previous research in this area and a review is offered of the papers which prompted us to conduct our own research and write this paper. An explanation of the research methodology, research tools and the characteristics of the sample follows. In the section entitled *Results and discussion*, the research findings are presented. At the end, the authors offer concluding remarks, state some limitations of the research and provide recommendations for future research in this field.

2. Related works

The research published up to 2000 placed greater emphasis on software evaluation as compared to website or web application evaluations. Web engineering spurred significant progress in this respect and very often the models for website assessment were developed based on the basic models for software evaluation.

Two dilemmas have always existed when it comes to a research of this kind:

1. What should be assessed, which criteria and attributes should be chosen?
2. Which methodological approach should be employed to reach the final assessment of website quality?

2.1 Criteria selection for website evaluation

When we analyze the selection of criteria and attributes we find that the first research on website evaluation generally focused on the selection and evaluation of those criteria that are important from the user's standpoint. However, no research resulted in a generally accepted list of criteria to measure user satisfaction with a website.

In the field of e-business, the role of the owner and employees is also of great importance (Ghandour et al., 2011), and therefore, in addition to website evaluation from the user's standpoint, it seems logical to conduct periodic internal evaluations.

There are a limited number of papers on internal evaluation and this research gap was what prompted us to devote part of our own research to internal evaluation.

We deem this to be important since employees are also users and might often spot problems better than clients. In addition, internal evaluation can serve as an “early warning system” (Lai, 2006), but also as an instrument for measuring the level of return on an investment in ICT (Tallon and Kraemer, 2002). In the study A strategic framework for website evaluation based on a review of the literature from 1995–2006 the authors, with a view to strategy, recommend conducting an internal evaluation first and then proceeding with the external evaluation (Chiou et al., 2010).

Some authors are of the opinion that Web design affects the client's immediate satisfaction (Yang et al., 2005). On the other hand, research has also shown that design influences the user's future behaviour (Liang et al., 2008).

In part of the published research (Sénécal et al., 2005) it has been shown that different categories of users have different perceptions in terms of design, and that when creating an online presentation and performing website analysis a user quality evaluation segmentation should be done.

In some research, the authors emphasize the importance of the navigability for a website quality. Kalbach (2007) believes that navigability is “the systematic organization of links to provide access to information and to make meaningful associations in a way that enhances understanding, reflects brand, and lends to overall credibility of a site”.

Some recent research suggests methods and instruments for measuring the quality of website navigation. (Vaucher and Sahraoui, 2010; Fang et al., 2012).

The simplicity of a website is measured by the level of the ease of use. Ahn et al., (2007) reminds us that for users “ease of use means a clear system which is easy to be understood, skilled and controllable”.

A synthetic overview of the studies which employed the criteria content quality leads to the conclusion that content is always the main source of value for the user (Hargittai et al., 2010; Kassim and Asiah, 2010).

In precise terms, any evaluation should show if the content is useful, presented in a fully integrated manner and true (Winkler, 2001), but also whether it corresponds to the target market (Nielsen, 2002). In addition, it has been shown that the importance of the criterion content quality is not the same for different user categories (Karkin and Janssen, 2014), which means that when assigning weight to the criteria the target market should be taken into consideration. Given the nature of bank operations, it is without question that content quality must be evaluated.

Trust in a website always depends on the perception of security and privacy. Security is better understood if taken together with privacy This is corroborated by the fact that in a number of papers the two criteria were always analyzed together (Long and McMellon, 2004). Many papers show that security and privacy are important determinants of online loyalty. (Ramanathan, 2011; Aldas et al., 2011). Only one study Gefen (2002) does not feature a statistically relevant connection between user

satisfaction and risk. Due to the nature of banking services, banks' website evaluation must include the criterion of safety and security.

Interactivity is a significant determinant of website quality and is assessed based on whether there are a sufficient number of communication channels through which clients can define their requirements and receive feedback (Achour and Bensedrine, 2005; Carlson and O'Cass, 2010). In addition, interactivity is also evaluated based on the manner and quality of information exchange between the website owner and the user. (Barrutia and Gilsanz, 2013).

Some authors suggest that webpage speed performance and download speed be standard website quality evaluation criteria (Kahraman and Kuya, 2011).

It is clear that, in addition to the selected criteria and sub-criteria, the validity of the assessment depends on the distribution of the weights in the total evaluation. For such purposes, the AHP method has been used multiple times (Tsai et al., 2010; Dincer and Hacıoğlu, 2013), and we have also opted for it when verifying weight coefficients.

2.2 Methodological approaches to website evaluation

Summing up the quantitative research on website evaluation Law et al., (2010) identifies 5 evaluation principles: counting, user judgment, automated, numerical computation, and combined methods. There are good and bad elements to all the methods.

The counting method has been used in numerous papers to determine how rich a website is in terms of content (Bai et al., 2006; Xiong et al., 2009). As can be seen from the very name, this method is used to identify whether certain methods are extant on a website, but their quality and the ease of their use are not determined.

The user judgment method is used to assess a website from the user standpoint, that is to

determine the level of the user's satisfaction, but without a numerical representation of the overall website evaluation. This method has been used in numerous studies to measure the user's satisfaction. (Hur et al., 2011; Belanche et al., 2012).

The automated method is used in many papers for automated, software supported evaluations with the maximum exclusion of the subjective opinion of the evaluators (Chan and Law 2006; Qi et al., 2008)). This approach is good because it ensures a consistency to the evaluation and testing technical characteristics of a website, but the drawback of the approach is that it neglects the user's perception.

One group of authors opted for the numerical computation method (Yeung et al., 2006; Miranda et al., 2006). This method has a mathematically-rooted background, and to determine the final numerically expressed assessment, lists of criteria and attributes should be defined, and their weight coefficients determined.

In order to avoid the downsides of the above methods, can be research based on combinations of two or more of the suggested methods.

The present research is based on a methodological framework which combines the counting and automated methods.

3. Research methodology, research tools and characteristics of the sample

As per the defined research objective, the authors started from the hypothesis that the same evaluation criteria are assessed differently by clients, bank employees and independent evaluators, that is, that there is a statistically significant difference in the average scores of the internal evaluation compared to those given by clients and independent evaluators.

There is a total of 14 banks operating in Montenegro and 3 different evaluation

procedures were carried out for all of them.

For these purposes, an extensive review of the literature in this field was undertaken and about 80 academic articles were consulted, which dealt with website evaluation in various branches of industry; following this, an analysis of the applicable models and their criteria and attributes performed.

According to the nature of the problem which is the subject of this paper, the 2QCV3Q model was selected. The reason for choosing this model is the fact that it is flexible and applicable for evaluating the quality of websites from different angles. The 2QCV3Q permits a multi-stakeholder approach that considers all perspectives of Web application designer, deployer, owner, and its users (Kumar et al., 2015).

This conceptual framework has been partially modified by introducing weight coefficients for the evaluation of various dimensions of a website, thus improving the quality of the methodology.

On the basis of the original 2QCV3Q a list of criteria and attributes for independent evaluation was developed and this later served to create surveys.

The first, independent evaluation was conducted by the authors of this paper.

The results of the evaluation performed by clients were obtained by processing the survey completed by a total of 610 clients.

The results of the internal evaluation, in which the survey was distributed to the employees in the marketing sectors, employees in the IT sector and Web designers, were obtained by processing data from 125 surveys.

After the average scores of the 3 different evaluations has been obtained, a single-factor analysis of variance (single-factor ANOVA) was carried out, which was aimed at determining whether there was a statistically significant difference between these results.

3.1 The process of independent evaluation

An independent evaluation of websites was conducted in such a way that on the basis of the original 2QCV3Q model (Table 1), which consists of 7 criteria (Identity, Content, Services, Contact, Maintenance, Usability, and Feasibility), a list of 69 attributes was defined. Then, the criteria and attributes were assigned weights. Since most multi-criteria decision-making models require that the criteria be assigned weights according to their importance, the authors did so using the experiential method, and then through the AHP model, specifically using the software Expert Choice, the index of inconsistency was calculated. Since the value of inconsistency was 0.07, which is less than 0.1, no additional corrections were made regarding the change in the weight coefficients.

The list of attributes was created in accordance with the nature of the banking business.

Table 1. Criteria of the 2QCV3Q model (Mich et al., 2003)

Criteria	Attributes
Identity	Identification, Characterisation
Content	Coverage, Accuracy
Services	Functionalities, Control
Location	Reachability, Interactivity
Maintenance	Corrective Maintenance, Adaptive Maintenance
Usability	Assessability, Navigability, Understandability
Feasibility	Resources, Information and Communication Technology

From a total of 100 points for evaluating the quality of a website, 25 were intended to assess usability, 20 to assess content, service and feasibility each, and 5 points each to assess identity, contact and maintenance. Each weight coefficient was divided by the number of attributes and the maximum weights of the obtained attributes. In the case of the existence of an attribute, one point

was assigned and then multiplied by its maximum weight coefficient, and in the absence of an attributes, zero was assigned. The web pages of banks were tested by accessing them from one computer every day at the same time. Testing was repeated 3 times within 5 months, and the obtained results featured negligible differences. Therefore, the initial result was considered valid.

For various technical data such as speed of loading and the number of defective links two tools, Xenu and Watson, were used. The testing results were recorded in an Excel spreadsheet and thus the final evaluation for each bank was obtained.

3.2 The process of client evaluation

For the purposes of this part of the research the data from the clients was collected through a carefully developed survey. The conceptual framework for the creation of the questionnaire was the list of criteria and attributes that were used in the independent evaluation.

The respondents were asked to assess the quality of the webistes against the 7-point Likert scale (from “strongly disagree” to “strongly agree”).

In order to compare the received evaluations from various angles, the 43 questions were divided into 7 sections. Each section covered one criterion of the 2QCV3Q model. Each section contained a number of questions that were developed on the basis of the defined attributes for an independent evaluation.

Each criterion in this evaluation was allocated as many points as the criteria in the independent evaluation was. The maximum values for each criterion were divided by the number of questions that criterion covers and then the maximum value calculated for each defined question.

After that, all the responses (from 1 to 7) were converted via a formula into values within the range of the maximum values that can be obtained for the corresponding

question. The data processing was performed in Excel.

The respondents were recruited from among bank customers who are NetBanking users. The survey was placed on the webpages for access to net banking systems and conducted on the principle of voluntary participation.

In a 60-day period, the survey was completed by 687 respondents, with the data for 77 of them being discarded immediately. Out of these 77 discarded surveys, 56 were incomplete while in 21 cases, the respondents were disregarded because they had been net banking users for less than a year. A total of 71% of the respondents had used the Internet for more than 10 years, and 29% for between 5 and ten years. A sample of 610 users of NetBanking was representative for the purposes of this study.

From the total number of the processed surveys, 312 (51%) clients were male and 298 (49%) female. The largest number of the respondents were highly educated to univeristy level (around 73%), 26% were high-school educated, while 1% had a primary education.

3.3 Evaluation by the banks’ employees

In the process of internal evaluation, the banks’ employees participated. These employees included web designers, web analysts, SEO administrators, Web marketing managers, Internet PR specialists, experts in market research, experts in the security of electronic transactions and experts in digital media campaigns.

Out of the total number of 179 surveys sent (as questionnaires), 125 were returned duly answered.

The respondents were informed about the purpose of the research and were asked not to give answers which would favour a higher quality score for their website. As regards the structure of this survey, it differed from the one intended for clients in the first section only. Instead of demographic questions, this respondent category was asked about the

work post they occupied (occupation). As for the analysis of the customers' perception, Microsoft Excel was used to form the average score as well as its formula to convert the evaluation against the Likert scale into values within the maximum value range for the weight coefficient of the attribute being evaluated.

4. Results and discussion

Table 2 (Appendix) shows the results of the evaluation, that is, provides information on the average quality of Montenegro banks' websites for all the criteria, and information on the overall average scores.

These results show that the scores obtained by the internal evaluation are significantly higher than the scores obtained in the evaluation by customers and independent evaluators. In this case, the independent evaluation served as a confirmation of one of the two previous evaluations.

These findings confirm previous studies which demonstrated that internal evaluation may not always be correct, because it often relies too heavily on subjective feeling (Huizingh 2002; Quaddus and Achjari 2005), or simply on the use of ready-made indicative tools (Google Analytics, LiveSTATS, XSP, CMS400.NET). The disadvantage of this approach is raw data are gathered abundantly, but they can serve as a measure of the quality of a site (Quaddus and Achjari 2005; Stockdale et al. 2006).

For the analysis of the differences in average scores obtained through the three types of evaluation (internal evaluation, evaluation by clients and independent evaluators assessment) we used a one-way ANOVA. According to the hypothesis posed by this paper it is assumed that there is a statistically significant difference in the average scores of the evaluation done by the three groups. According to the ANOVA, the null hypothesis is that the mean value of the dependent variable in three or more groups are equal to each other, while the alternative

hypothesis suggests that at least one mean value is different from the others, but there is also the possibility that all mean values are different from each other. In other words, while the null hypothesis is unique, there are several alternative hypotheses.

The assumptions of an analysis of variance are that the populations are normally distributed, that the variances in the groups are homogeneous and that the data are expressed on an interval scale or ratio scale. The data in the analysis are expressed on a ratio scale, and what follows is the evidence for other two conditions for the use of the ANOVA. The normal distribution of all seven variables for these three groups of scores (7 attributes) was confirmed by the Kolmogorov-Smirnov and Shapiro-Wilk tests, whose results are presented in Table 3. In fact, since the signatures of the tests are greater than 0.05 (for content in the clients group it is greater than 0.01, as well as for contact in the owners group, for usability in the evaluators group in the Kolmogorov-Smirnov test, and for service in the clients group in the Shapiro-Wilk test) for all the attributes, the data meet the requirement of normal distribution for further analysis.

The homogeneity of variances is confirmed by the Levene Test (Table 4) whereby it is obvious that the null hypothesis of homogenous variances cannot be rejected if the level of significance is 5% and the level of feasibility is slightly lower (0.046).

The essence of the analysis of variance is to compare two types of variation, whereby the variation (variance) between groups is compared with the variation (variance) within a group, so as to evaluate the difference between the mean values. The measures of variations are obtained by "unbundling" the total variation into the variation that occurs as a result of the influence of the observed factor – factor variance (variation between groups) and the random variation - residual variance (variation within groups).

Table 3. Tests of Normality

	Evaluator	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
identity	Internal evaluators	0,171	14	,200*	0,91	14	0,155
	clients	0,146	14	,200*	0,935	14	0,359
	Independent evaluators	0,156	14	,200*	0,894	14	0,094
content	Internal evaluators	0,118	14	,200*	0,969	14	0,867
	clients	0,232	14	0,039	0,907	14	0,143
	Independent evaluators	0,14	14	,200*	0,936	14	0,37
service	Internal evaluators	0,154	14	,200*	0,92	14	0,22
	clients	0,194	14	0,159	0,858	14	0,028
	Independent evaluators	0,189	14	0,186	0,906	14	0,138
contact	Internal evaluators	0,245	14	0,023	0,919	14	0,212
	clients	0,182	14	,200*	0,935	14	0,357
	Independent evaluators	0,18	14	,200*	0,914	14	0,179
maintenance	Internal evaluators	0,128	14	,200*	0,971	14	0,89
	clients	0,126	14	,200*	0,92	14	0,217
	Independent evaluators	0,219	14	0,067	0,937	14	0,38
usability	Internal evaluators	0,125	14	,200*	0,957	14	0,672
	clients	0,122	14	,200*	0,93	14	0,304
	Independent evaluators	0,232	14	0,039	0,883	14	0,064
feasibility	Internal evaluators	0,128	14	,200*	0,984	14	0,993
	clients	0,163	14	,200*	0,94	14	0,422
	Independent evaluators	0,145	14	,200*	0,945	14	0,481

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Table 4. Test of Homogeneity of Variances

	Levene Statistic	df1	df2	Sig.
identity	0,271	2	39	0,764
content	1,937	2	39	0,158
service	0,311	2	39	0,735
contact	2,213	2	39	0,123
maintenance	1,839	2	39	0,173
usability	0,05	2	39	0,951
feasibility	3,334	2	39	0,046

Factor variance is calculated as the quotient of the sum of the squared deviations from the average and the number of degrees of freedom (the number of groups minus 1), respectively, that is:

$$V_F = \frac{\sum_{i=1}^r n_i (\bar{x}_i - \bar{x})^2}{r - 1}$$

where r is the number of groups, n_i the number of observations in the i -th group, \bar{x}_i the average value of the dependent variable in the i -th group, and \bar{x} the average of all

observations (all the study groups).

The residual variance is then:

$$V_R = \frac{\sum_{i=1}^r \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2}{nr - r}$$

where x_{ij} is the j -th observation in the i -th group, and n the number of observations in each group.

The F statistics is the quotient of the factor and residual variances:

$$F = \frac{V_F}{V_R}$$

and follows the F-distribution with $r-1$ and $nr-r$ degrees of freedom. If the F statistics is higher than the critical value of the F distribution with $r-1$ and $nr-r$ degrees of freedom, the null hypothesis is rejected, and

the conclusion is that the examined factor of influence does influence the mean values of the groups of data that differ significantly. On the contrary, a true null hypothesis implies that the average values do not differ, i.e., that the independent variable (factor of influence) has no effect on the dependent variable.

The hypothesis posed by the paper in this case is here the alternative hypothesis of the F test of the ANOVA model. The paper compares the average evaluation scores obtained from the three groups (internal evaluators, clients and independent evaluators) on all seven attributes of website evaluation (identity, content, service, contact, maintenance, usability and feasibility). The ANOVA results are given in Table 5.

Table 5. Results of ANOVA for 7 criteria

		Sum of Squares	df	Mean Square	F	Sig.
identity	Between Groups	8,757	2	4,378	161,997	0
	Within Groups	1,054	39	0,027		
	Total	9,811	41			
content	Between Groups	64,722	2	32,361	68,511	0
	Within Groups	18,422	39	0,472		
	Total	83,144	41			
service	Between Groups	20,553	2	10,276	14,624	0
	Within Groups	27,406	39	0,703		
	Total	47,959	41			
contact	Between Groups	1,02	2	0,51	8,987	0,001
	Within Groups	2,214	39	0,057		
	Total	3,234	41			
maintenance	Between Groups	0,344	2	0,172	4,203	0,022
	Within Groups	1,598	39	0,041		
	Total	1,943	41			
usability	Between Groups	154,531	2	77,265	59,706	0
	Within Groups	50,47	39	1,294		
	Total	205,001	41			
feasibility	Between Groups	38,638	2	19,319	78,23	0
	Within Groups	9,631	39	0,247		
	Total	48,269	41			

The test signature for each of the seven attributes is lower than 0.05 (Table 5), which means that the null hypothesis of the equality of the average score for all three groups in the surveyed 14 banks in Montenegro can be rejected already with a risk of error of 5%, in all 7 cases. In other words, this means that there is a significant difference between the mean values of the studied groups. However, the F test does not answer the question of whether a significant difference exists between all the mean values or only between some of them. In order to determine which pairs of mean values there is significant difference between, Tukey's HSD (honestly significant difference) "post hoc" test was used.

Tukey's HSD test compares the absolute difference of each pair of averages in the observed groups with the critical value:

$$TSD = q(\alpha, k, v) \sqrt{MS_E \frac{n_i^{-1} + n_j^{-1}}{2}}$$

where $q(\alpha, k, v)$ is Student's rank statistics, α is the risk of error, k is the number of averages being compared with v degrees of freedom and MS_E the mean square error from the ANOVA. Thus, for each pair of group averages whose absolute difference exceeds this critical value we can say that they are significantly different from each other, that is, that these differences led to the dominant influence of the observed factor on the observed phenomenon and the rejection of the null hypothesis in the ANOVA model. The results of Tukey's HSD are presented in Table 6 (Appendix).

In the second column of Table 6 differences are indicated that are at the level of significance of 5%. The conclusion which arises is that the average score of the internal evaluation differs significantly from the average scores given by the remaining two groups (clients and independent evaluators). In other words, internal evaluators assessed the quality of the site of their banks in a

significantly different way while clients and independent evaluators gave similar assessment for the same attributes, and their evaluation is more realistic. These results agree with previous findings (Granić and Marangunić 2011; Frøkjær and Hornbæk 2008) concerning whether there are differences in evaluation assessment between internal evaluators and clients. In the case of the evaluation of *maintenance* that the average score by the clients is somewhere between the average score given by the internal and by the independent evaluators, at such a distance that the difference is not statistically significant either when compared to the average score by the internal evaluation or compared to the assessment of independent evaluators.

The reasons for some of the greater discrepancies in the scores given by the employees and those obtained from the clients for 6 (out of 7) attributes, as well as the similar score for *maintenance*, can be numerous.

If we try to find the reasons for the deviations in these scores, these are various directions the discussion may branch into:

- The definition and perception of quality is certainly not the same for employees and customers. Even among the employees themselves there may be differences. It is logical that, when assessing attributes, the employees in the IT sector will place a greater emphasis on technical and security details, and that the employees in the marketing sector will be more focused on the visual attributes of a website.
- As hard as we tried to have objective assessment of a website, it was absolutely impossible to rule out the subjectivity of the employees and we are, therefore, of the opinion that clients give more objective assessment of the criteria, something which was confirmed through independent evaluation.

- The employees perceive the quality of the website on a much broader scale compared to customers. The clients evaluated the offered attributes only, while the employees probably evaluated them based on what benefits they can yield for them. It is also possible that the employees assessed a number of attributes on the basis of their potential contribution to greater advertising possibilities for other companies on the website of the bank.
- The employees perceive the website as a project and when assessing its quality they also think of the investment and the return on investment in the website and so on, while a customer sees only the tangible, visible output of the investment of others.
- When evaluating the website, the employees probably give better ratings because they are familiar with the technology behind the website, the integration of the web business technology with the existing information system, the IT expertise involved, and so on, while a client has no such considerations.

The similar scores obtained for *maintenance* probably stem from the fact that when assessing this criterion, the subjective feeling of the evaluator is least involved. There is an awareness within companies that users are easily able to observe whether a website is well maintained, and web technologies develop at such a fast rate that all companies, without exception, use readily available sophisticated tools for website maintenance (Brajnik, G. 2000).

5. Conclusion

Since the quality of a website depends on its ability to meet the needs of both customers and employees, it is logical that the evaluation should be carried out from

different perspectives.

This paper has shown that the same evaluation criteria are rated differently by clients and by bank employees, or rather that the perception of employees is significantly different from the perception of customers. The ANOVA confirmed that the average score given by the owner is statistically significantly different from the average score given by bank customers, for 6 out of 7 observed attributes of the quality of a site (with the exception being *maintenance*). The evaluation by independent evaluators served to confirm how realistic the scores given by internal evaluators are and showed that customers' ratings are still more realistic compared to the internal evaluation.

The paper has thus shown that the idea of the internal evaluation of a website is not good and such an evaluation may not result in the correct evaluation of the quality of a website.

When assessing website quality, the judgement of the marketing and IT sectors should be taken with caution, and thus should be verified by customers' ratings, and if need presents itself, by independent evaluators assessment as well.

In terms of theory, the results of this research are useful because they contribute to the debate on the importance of the evaluation of websites. In addition, this work could contribute to future research in terms of defining the criteria that can be used for website evaluation in the banking sector. While this research is limited to the banking sector, the proposed instrument for measuring website quality is flexible enough to be used for future research in evaluation across a range of industries.

This study complements previous research efforts in the field of web evaluation with the idea that differences in the evaluation scores produced by different categories of evaluators can be perceived.

The results offered here might serve as a basis for future comparative studies, that is, they can be used for comparative analyses of the banking sectors in countries that are at

the same or a similar level of socio-economic and technological development as Montenegro.

In practice, a study of this kind may increase awareness of the importance of formal website evaluation, and of the effectiveness of the methodologies by which a realistic assessment of the quality of a website can be obtained.

In addition, the obtained evaluation may be a good form of input for marketing experts on the future courses of action in the field of internet marketing. The analysis conducted in this paper can be used by employees in the marketing sector of banks to draw conclusions about the expectations and demands of clients in terms of the elements of a website and to direct their strategies towards meeting those expectations.

Only 14 banks operate in Montenegro, so the authors are aware that although the survey covered the whole banking market, the results of the research would be more relevant if there were more banks. This is the major limitation of the study.

Still, based on this very limitation, recommendations for future research can be defined. It would be interesting to conduct research that would show whether the perceptions of employees and clients differ to the same extent in all sectors of the economy.

In addition, since the research was conducted in Montenegro, a small and developing country, future research might show whether the findings differ depending on a country's size and level of development.

References:

- Achour, H., & Bensedrine, N. (2005). An evaluation of internet banking and online brokerage in Tunisia. In: *Proceedings of the First International Conference on E-Business and E-learning (EBEL)* (pp.147-158).
- Ahn, T., Ryu, S., & Han, I. (2007). The impact of Web Quality and Playfulness on user acceptance of online retailing. *Information & Management*, 44, 263-275.
- Aldas-Manzano, J., Ruiz-Mafe, C., Sanz-Blas, S., & Lassala-Navarré, C. (2011). Internet banking loyalty: evaluating the role of trust, satisfaction, perceived risk and frequency of use. *The Service Industries Journal*, 31(7), 1165-1190.
- Bai, B., Hu, C., & Jang, S.-C. (2006). Examining e-relationship marketing features on hotel websites. *Journal of Travel & Tourism Marketing*, 21(2/3), 33-48
- Barrutia, J. M., & Gilsanz, A. (2013). Electronic Service Quality and Value Do Consumer Knowledge-Related Resources Matter? *Journal of Service Research*, 16(2), 231-246.
- Belanche, D., Casaló, L. V., & Guinalú, M. (2012). Website usability, consumer satisfaction and the intention to use a website: The moderating effect of perceived risk. *Journal of retailing and consumer services*, 19(1), 124-132.
- Brajnik, G. (2000, June). Automatic web usability evaluation: what needs to be done? In *Proc. Human Factors and the Web, 6th Conference*.
- Burmaoglu, S., & Kazancoglu, Y. (2012). E-government website evaluation with hybrid MCDM method in fuzzy environment. *International Journal of Applied Decision Sciences*, 5(2), 163-181.
- Callaway, S. K. (2011). Internet banking and performance: The relationship of web site traffic rank and bank performance. *American Journal of Business*, 26(1), 12-25

- Carlson, J., & O'Cass, A. (2010). Exploring the relationships between e-service quality, satisfaction, attitudes and behaviours in content-driven e-service web sites. *Journal of services marketing*, 24(2), 112-127.
- Chan, S., & Law, R. (2006). Automatic website evaluations: the case of hotels in Hong Kong. *Information Technology & Tourism*, 8(3), 255-269
- Chiou, W. C., Lin, C. C., & Perng, C. (2010). A strategic framework for website evaluation based on a review of the literature from 1995-2006. *Information & management*, 47(5), 282-290.
- Davidavičienė, V., & Tolvaišas, J. (2011). Measuring quality of e-commerce web sites: Case of Lithuania. *Economics and management*, 16, 723-729.
- Dincer, H., & Hacıoglu, U. (2013). Performance evaluation with fuzzy VIKOR and AHP method based on customer satisfaction in Turkish banking sector. *Kybernetes*, 42(7), 1072-1085.
- Fang, X., Hu, P. J. H., Chau, M., Hu, H. F., Yang, Z., & Sheng, O. R. L. (2012). A data-driven approach to measure web site navigability. *Journal of Management Information Systems*, 29(2), 173-212.
- Ferreira, S. B. L., & Leite, J. C. S. P. (2002). Exemplificando aspectos de usabilidade em sistemas de informação. In *Anais ENANPAD*.
- Frøkjær, E., & Hornbæk, K. (2008). Metaphors of human thinking for usability inspection and design. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 14(4), 20.
- Gefen, D. (2002). Customer loyalty in e-commerce. *Journal of the Association for Information Systems*, 3(1), 27-51.
- Ghandour, A., Benwell, G., & Deans, K. R. (2011). Measuring the Performance of eCommerce Websites—An Owner's Perspective. *Pacific Asia Journal of the Association for Information Systems*, 3(1).
- Granić, A., Mitrović, I., & Marangunić, N. (2011). Exploring the usability of web portals: A Croatian case study. *International Journal of Information Management*, 31(4), 339-349.
- Hargittai, E., & Fullerton, L., Menchen-Trevino, E., & Thomas, K. Y. (2010). Trust online: Young adults' evaluation of web content. *International Journal of Communication*, 4(2010), 468-494.
- Hernando, I. & Nieto, M. (2007). Is the internet delivery channel changing banks' performance? The case of Spanish banks. *Journal of Banking & Finance*, 31(4), 1083-1099.
- Huang, W., Le, T., Li, X., & Gandha, S. (2006). Categorizing web features and functions to evaluate commercial web sites: An assessment framework and an empirical investigation of Australian companies. *Industrial Management & Data Systems*, 106(4), 523-539.
- Huizingh, E. (2002). The Antecedents of Website Performance. *European Journal of Marketing*, 36(11/12), 1225-1247.
- Kahraman, C., & Kaya, T. (2011). A fuzzy approach to e-banking website quality assessment based on an integrated AHP-ELECTRE method. *Technological and Economic Development of Economy*, 17(2), 313-334.
- Kalbach, J. (2007). Designing Web Navigation: Optimizing the User Experience. *Journal of Management Information Systems*, 21(1), 167-202
- Karkin, N., & Janssen, M. (2014). Evaluating websites from a public value perspective: A review of Turkish local government websites. *International Journal of Information Management*, 34(3), 351-363.

- Kassim, N., & Asiah Abdullah, N. (2010). The effect of perceived service quality dimensions on customer satisfaction, trust, and loyalty in e-commerce settings: a cross cultural analysis. *Asia Pacific Journal of Marketing and Logistics*, 22(3), 351-371.
- Kesharwani, A., & Singh Bisht, S. (2012). The impact of trust and perceived risk on internet banking adoption in India: An extension of technology acceptance model. *International Journal of Bank Marketing*, 30(4), 303-322.
- Klaus, P., & Nguyen, B. (2013). Exploring the role of the online customer experience in firms' multi-channel strategy: An empirical analysis of the retail banking services sector. *Journal of Strategic Marketing*, 21(5), 429-442.
- Kumar, N., Dadhich, R., & Shastri, A. (2015). Quality Models for Web-based Application: A Comparative Study. *International Journal of Computer Applications*, 125(2).
- Ladhari, R., Ladhari, I., & Morales, M. (2011). Bank service quality: comparing Canadian and Tunisian customer perceptions. *International Journal of Bank Marketing*, 29(3), 224-246.
- Lai, J. Y. (2006). Assessment of employees' perceptions of service quality and satisfaction with e-business. *International Journal of Human-Computer Studies*, 64(9), 926-938.
- Law, R., Qi, S., & Buhalis, D. (2010). Progress in tourism management: A review of website evaluation in tourism research. *Tourism management*, 31(3), 297-313.
- Liang, C. H., Chen, J. J., & Wang, W. F. (2008). Does online relationship marketing enhance customer retention and cross-buying? *Service Industries Journal*, 28(6), 769-787.
- Liu, G. Z., Liu, Z. H., & Hwang, G. J. (2011). Developing multi-dimensional evaluation criteria for English learning websites with university students and professors. *Computers & Education*, 56(1), 65-79.
- Long, M., & McMellon, C. (2004). Exploring the determinants of retail service quality on the internet. *Journal of Services Marketing*, 18(1), 78-90.
- Manzari, L., & Trinidad-Christensen, J. (2013). User-centered design of a web site for library and information science students: Heuristic evaluation and usability testing. *Information technology and libraries*, 25(3), 163-169.
- Mich, L., Franch, M., & Cilone, G. (2003). The 2QCV3Q Quality model for the analysis of web site requirements. *Journal of Web Engineering*, 2, 105-127.
- Mich, L., Franch, M., & Gaio, L. (2003). Evaluating and designing web site quality. *MultiMedia, IEEE*, 10(1), 34-43.
- Miranda, F. J., Barriuso, C., & Cortés, R. (2006). Quantitative evaluation of e-banking web sites: An empirical study of Spanish banks. *The Electronic Journal Information System Evaluation*, 73-82.
- Molla, A., & Licker, P. (2001). E-Commerce System Success: An attempt to partially extend and respecify the Delone and Mclean model of IS success. *Journal of Electronic Commerce Research*, 2(4), 131-141.
- Molla, A., & Licker, P. (2001). e-Commerce systems success: an attempt to extend and respecify the Delone and Maclean of IS success. *Journal of Electronic Commerce Research*, 2(4), 131-141.
- Ngai, E. W. T. (2003). Selection of web sites for online advertising using the AHP. *Information & Management*, 40(4), 233-242.
- Nielsen, J., (2002). *Designing web usability: The practice of simplicity*. Indianapolis: New Riders Publishing.

- Qi, S. S., Leung, R., Law, R., & Buhalis, D. (2008). A study of information richness and downloading time for hotel websites in Hong Kong. In P. O'Connor, W. Hopken, U. Gretzel (Eds.), *Information and Communication Technologies in Tourism 2008* (pp. 267-278). New York: Springer-Wien.
- Qi, S., Ip, C., Leung, R., & Law, R. (2010, May). A new framework on website evaluation. In *E-Business and E-Government (ICEE), 2010 International Conference* (pp. 78-81).
- Quaddus, M., & Achjari, D. (2005). A Model for electronic Commerce Success. *Telecommunications Policy*, 29(2-3), 127-52.
- Ramanathan, R. (2011). An empirical analysis on the influence of risk on relationships between handling of product returns and customer loyalty in E-commerce. *International Journal of Production Economics*, 130(2), 255-261.
- Sénécal, S., Kalczynski, P., & Nantel, J. (2005). Consumer decision-making process and their online shopping behavior: A clickstream analysis. *Journal of Business Research*, 58, 1599-1608.
- Stockdale, R., Lin, C., & Liu, Y. (2006). The Perception of Website Effectiveness by West Australian Small Business Owners in a B2B Context. In *Pacific Asia Conference on Information*.
- Ghandour, A., Benwell, G., & Deans, K. (2011). Measuring the Performance of eCommerce Websites – An Owner's Perspective. *Pacific Asia Journal of the Association for Information Systems*, 3(1), 1-27.
- Tallon, P., & Kraemer, K. (2002). Executives' Perspectives on IT: Unraveling the Link between Business Strategy, Management Practices and IT Business Value. In America's Conference on Information Systems (ACIS2002). Dallas, TX, USA. *Telecommunications Policy*, 29(2-3): 127-152.
- Trujillo-Ponce, A. (2013). What determines the profitability of banks? Evidence from Spain. *Accounting & Finance*, 53(2), 561-586.
- Tsai, W. H., Chou, W. C., & Lai, C. W. (2010). An effective evaluation model and improvement analysis for national park websites: A case study of Taiwan. *Tourism Management*, 31(6), 936-952.
- Vaucher, S., & Sahraoui, H. (2010, September). Multi-level evaluation of web site navigability. In *Web Systems Evolution (WSE), 2010 12th IEEE International Symposium on* (pp. 93-100).
- Wang, X., & Huang, W. (2009). *Lund University Website Evaluation: Focus on homepage and English research pages* (Master thesis). Retrieved from <https://lup.lub.lu.se/luur/download?func=downloadFile&recordOid=1458603&fileOid=1647044>
- Winkler, R. (2001). *Portals-The all-in-one web supersites: Features, functions, definition, taxonomy*. SAP Design Guild (ed.3)
- Wolfenbarger, M., & Gilly, M. C. (2003). eTailQ: dimensionalizing, measuring and predicting e-tail quality. *Journal of Retailing*, 79(3), 183-98.
- Xiong, L., Cobanoglu, C., Cummings, P., & DeMicco, F. (2009). Website accessibility of U.S. based hospitality websites. In W. Hopken, U. Gretzel, & R. Law (Eds.), *Information and communication technologies in tourism* (pp. 273-284). New York: Springer-Wien.
- Xue, M., Hitt, L. M., & Chen, P. Y. (2011). Determinants and outcomes of internet banking adoption. *Management Science*, 57(2), 291-307.

Yang, Z., Cai, S., Zhou, Z., & Zhou, N. (2005). Development and validation of an instrument to measure user perceived service quality of information presenting web portals. *Information & Management*, 42(4), 575-589.

Zhu, K. (2004). The Complementarity of Information Technology Infrastructure and E-Commerce Capability: A Resource-Based Assessment of Their Business Value. *Journal of Management Information Systems* 21(1), 167-202.

Biljana Rondović

University of Montenegro,
Faculty of Economics
Podgorica
Jovana Tomasevica 37,
81000 Podgorica
Montenegro
biljaro@ac.me

Julija Cerović

University of Montenegro,
Faculty of Economics
Podgorica
Jovana Tomasevica 37,
81000 Podgorica
Montenegro
julija@ac.me

Tamara Đuričković

University of Montenegro,
Faculty of Economics
Podgorica
Jovana Tomasevica 37,
81000 Podgorica
Montenegro
tamara.djurickovic@gmail.com

Boban Melović

University of Montenegro,
Faculty of Economics
Podgorica
Jovana Tomasevica 37,
81000 Podgorica
Montenegro
bobanm@ac.me

Appendix:

Table 2. Average quality scores of Montenegrin bank websites individually by criteria and the overall average scores

			Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6	Bank 7	Bank 8	Bank 9	Bank 10	Bank 11	Bank 12	Bank 13	Bank 14
			Bank 1	Bank 2	Bank 3	Bank 4	Bank 5	Bank 6	Bank 7	Bank 8	Bank 9	Bank 10	Bank 11	Bank 12	Bank 13	Bank 14
OVERALL WEBSITE QUALITY ASSESSMENT	IDENTITY (5)	OA	4,91	4,68	4,94	4,87	4,65	4,92	4,56	4,82	4,63	4,73	4,66	4,9	4,69	4,71
		CA	3,78	3,66	4,01	4,11	3,88	3,86	3,66	3,85	3,68	4,02	3,69	3,99	3,78	3,94
		IEA	3,75	3,59	3,99	3,86	3,76	3,19	3,71	3,76	3,63	3,93	3,71	4,03	3,66	3,86
	CONTENT (20)	OA	18,31	18,56	17,89	18,56	17,35	17,43	17,85	16,95	17,54	18,03	17,65	17,55	18,12	17,68
		CA	15,69	15,64	15,02	16,01	14,44	14,72	15,01	14,98	14,72	16,23	14,42	14,25	17,01	14,68
		IEA	15,47	15,31	15,31	16,13	14,31	15,14	14,93	15,02	14,56	15,96	14,26	14,35	16,83	14,81
	SERVICE (20)	OA	17,33	17,25	16,54	18,12	16,31	15,98	15,83	14,11	15,87	17,11	16,98	16,52	17,23	16,53
		CA	15,54	15,02	14,91	17,24	14,88	14,25	14,14	13,98	14,23	15,03	15,04	15,28	15,58	14,87
		IEA	15,46	15,43	15,01	16,99	14,69	14,43	15,01	14,2	14,28	15,43	14,94	15,09	16,01	15,06
	CONTACT (5)	OA	4,87	4,58	4,79	4,61	4,59	4,93	4,49	4,87	4,61	4,69	4,6	4,82	4,73	4,61
		CA	3,98	4,05	4,85	4,55	4,58	4,68	4,12	4,55	4,09	4,35	4,49	4,66	4,53	4,33
		IEA	3,59	4,12	4,79	4,26	4,49	4,44	4,11	4,29	4,38	4,27	4,38	4,57	4,61	4,41
	MAINTENANCE (5)	OA	3,99	3,89	4,11	3,98	3,69	3,87	3,86	4,03	3,85	3,72	3,66	4,21	3,95	3,84
		CA	4,03	3,86	4,05	3,97	3,82	4,03	3,68	3,84	3,56	3,71	3,53	3,46	3,51	3,61
		IEA	3,89	3,86	4,1	3,84	3,64	4,07	3,47	3,61	3,64	3,59	3,61	3,35	3,39	3,53
USABILITY (25)	OA	19,31	18,62	18,98	17,69	17,12	20,03	21,51	17,35	17,98	18,13	18,61	19,08	20,01	18,99	
	CA	14,26	13,98	12,87	13,33	13,01	15,24	16,05	14,58	15,02	15,65	15,58	16,03	16,21	14,85	
	IEA	14,2	13,76	13,01	13,41	13,36	15,89	15,99	13,98	14,87	15,77	15,67	15,96	15,67	14,69	
FEASIBILITY (20)	OA	14,01	13,69	13,86	13,51	13,46	14,23	14,52	13,31	13,67	14,09	13,92	13,87	14,11	13,96	
	CA	11,96	12,25	11,96	11,28	11,31	12,69	12,58	10,98	11,23	12,06	11,37	11,58	12,21	12,28	
	IEA	12,03	12,36	12,12	11,67	11,27	13,01	12,61	11,09	11,11	11,87	11,24	12,01	11,95	11,37	
Site quality estimate in total	OA	82,73	81,27	81,11	81,34	77,17	81,39	82,62	75,44	78,15	80,5	80,08	80,95	82,84	80,32	
	CA	69,24	68,46	67,67	70,49	65,92	69,47	69,24	66,76	66,53	71,05	68,12	69,25	72,83	68,56	
	IEA	68,39	68,43	68,33	70,16	65,52	70,17	69,83	65,95	66,47	70,82	67,81	69,36	72,12	67,73	

OA - Owners' assessment; CA - Clients' assessment; IEA - Independent evaluators' assessment

Table 6. Multiple Comparisons

Dependent Variable	(I) evaluator	(J) evaluator	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
identity	Internal evaluators	clients	,91143 [*]	0,06214	0	0,76	1,0628
		Independent evaluators	1,01714 [*]	0,06214	0	0,8658	1,1685
	clients	Internal evaluators	-,91143 [*]	0,06214	0	-1,0628	-,076
		Independent evaluators	0,10571	0,06214	0,218	-0,0457	0,2571
Independent evaluators	Internal evaluators	-1,01714 [*]	0,06214	0	-1,1685	-,8658	
	clients	-,010571	0,06214	0,218	-0,2571	0,0457	
content	Internal evaluators	clients	2,61786 [*]	0,25977	0	1,985	3,2507
		Independent evaluators	2,64857 [*]	0,25977	0	2,0157	3,2814
	clients	Internal evaluators	-2,61786 [*]	0,25977	0	-3,2507	-,1985
		Independent evaluators	0,03071	0,25977	0,992	-0,6022	0,6636
	Independent evaluators	Internal evaluators	-2,64857 [*]	0,25977	0	-3,2814	-2,0157
		clients	-,03071	0,25977	0,992	-0,6636	0,6022
service	Internal evaluators	clients	1,55143 [*]	0,31684	0	0,7795	2,3234
		Independent evaluators	1,40571 [*]	0,31684	0	0,6338	2,1776
	clients	Internal evaluators	-1,55143 [*]	0,31684	0	-2,3234	-,7795
		Independent evaluators	-,014571	0,31684	0,89	-0,9176	0,6262
	Independent evaluators	Internal evaluators	-1,40571 [*]	0,31684	0	-2,1776	-,6338
		clients	0,14571	0,31684	0,89	-0,6262	0,9176
contact	Internal evaluators	clients	,28429 [*]	0,09005	0,008	0,0649	0,5037
		Independent evaluators	,36286 [*]	0,09005	0,001	0,1435	0,5823
	clients	Internal evaluators	-,28429 [*]	0,09005	0,008	-0,5037	-,0649
		Independent evaluators	0,07857	0,09005	0,661	-0,1408	0,298
	Independent evaluators	Internal evaluators	-,36286 [*]	0,09005	0,001	-0,5823	-,1435
		clients	-,07857	0,09005	0,661	-0,298	0,1408
maintenance	Internal evaluators	clients	0,14214	0,07652	0,165	-0,0443	0,3286
		Independent evaluators	,21857 [*]	0,07652	0,018	0,0321	0,405
	clients	Internal evaluators	-0,14214	0,07652	0,165	-0,3286	0,0443
		Independent evaluators	0,07643	0,07652	0,582	-0,11	0,2629
	Independent evaluators	Internal evaluators	-,21857 [*]	0,07652	0,018	-0,405	-,0321
		clients	-,07643	0,07652	0,582	-0,2629	0,11
usability	Internal evaluators	clients	4,05357 [*]	0,42997	0	3,006	5,1011
		Independent evaluators	4,08429 [*]	0,42997	0	3,0368	5,1318
	clients	Internal evaluators	-4,05357 [*]	0,42997	0	-5,1011	-,3006
		Independent evaluators	0,03071	0,42997	0,997	-1,0168	1,0782
	Independent evaluators	Internal evaluators	-4,08429 [*]	0,42997	0	-5,1318	-,30368
		clients	-,03071	0,42997	0,997	-1,0782	1,0168
feasibility	Internal evaluators	clients	2,03357 [*]	0,18783	0	1,576	2,4912
		Independent evaluators	2,03571 [*]	0,18783	0	1,5781	2,4933
	clients	Internal evaluators	-2,03357 [*]	0,18783	0	-2,4912	-,1576
		Independent evaluators	0,00214	0,18783	1	-0,4555	0,4597
	Independent evaluators	Internal evaluators	-2,03571 [*]	0,18783	0	-2,4933	-,15781
		clients	-,00214	0,18783	1	-0,4597	0,4555

*. The mean difference is significant at the 0.05 level