

In Search of the Customer's Quality View

T. Stålhane*

SINTEF - DELAB

P.C. Borgersen

Software Innovation AS

K. Arnesen

Fellesdata AS

N-7034 Trondheim, Norway, p: +47 73593014, f: + 47 73532586 *

Abstract

This paper describes the work done in the PROFF project to get the customer's view of software product quality. The customers in this case are organizations that buy software for PCs and workstations. The main result is that the quality of service and follow-up activities are more important than the quality of the product itself. This observation should have an impact on the way we market and sell software products in the future.

Keywords

Customer quality view, product quality, service quality, price vs. quality requirements

1 INTRODUCTION

This paper describes the work undertaken by personnel working in the Norwegian software quality program PROFF, which has as its goal to improve the quality and competitiveness of Norwegian software industry.

The first activity in the project was to find the aspects of quality that are most important to the customers. In order to find an answer to this, we made a questionnaire based on the structure of the definitions of product and service quality in ISO 9126 and 9004 respectively.

This questionnaire was distributed to 1000 Norwegian companies that buy commercial software products for PCs and workstations, either off the shelf, as standard software packages or as adapted or tailor made solutions. The responders in our survey were asked to give us the relevant information for a product that they had bought recently. In this way, we hoped to get the responders to be quite concrete in their responses. In addition, this approach enabled us to also get information on price, application area and product category. Among the 90 questionnaires that were returned, six lacked all product related information, while eight only lacked information on product category. Thus, only 82 questionnaires were available for product category related analysis and 84 questionnaires could be used for analysis pertaining to product price.

The returned questionnaires were analysed and the results were presented to the Norwegian system development companies that are members of PROFF. These companies will use the results to better target their products against the market's expectations and to deliver products and services that are able to compete successfully in the European markets in the next five years.

2 QUALITY MODELS AND QUESTIONNAIRE STRUCTURE

The questionnaire was structured according to the structure of the factor - criteria models used in ISO 9126 (1991) and ISO 9004 (1987). These two models are shown in figures 1 and 2 respectively. For the service quality, we removed the criteria not related to the service provided by software producers. Typical examples of removed criteria are hygiene, comfort and aesthetics of the environment. Such criteria have little relevance here since software product service is mostly provided at the customer's premises or over the telephone.

As mentioned above, we also collected information on the price of the software. This was done in order to see if there was any significant difference in the quality requirements depending on the price of the software.

For each criterion, the responders were asked to indicate how important this quality criterion was for their decision to buy or not to buy the software product under consideration. The answers were given by using a four point scale, graded as follows:

- high importance - score 3;
- some importance - score 2;
- little importance - score 1;
- no importance or irrelevant - score 0.

In addition, we asked for price information. We split the responses according to price, using the following categories:

- category 1: 0 - 10 000 NOK, or approximately 0 - 1 500 USD;
- category 2: 10 000 - 50 000 NOK, or approximately 1 500 - 7 700 USD;
- category 3: more than 50 000 NOK or approximately 7 700 USD.

The complete questionnaire - in Norwegian - can be found in Arnesen (1995).

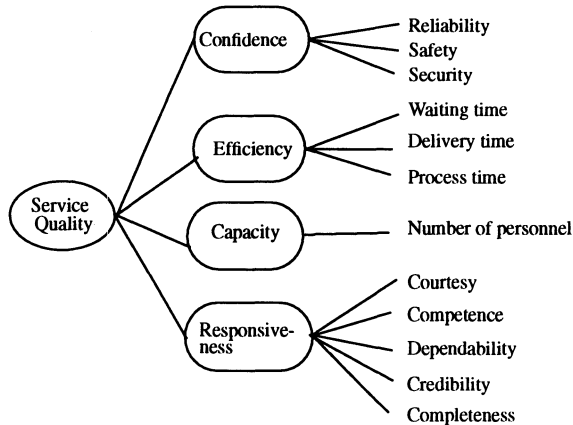


Figure 1. Reduced ISO 9004 model for service quality.

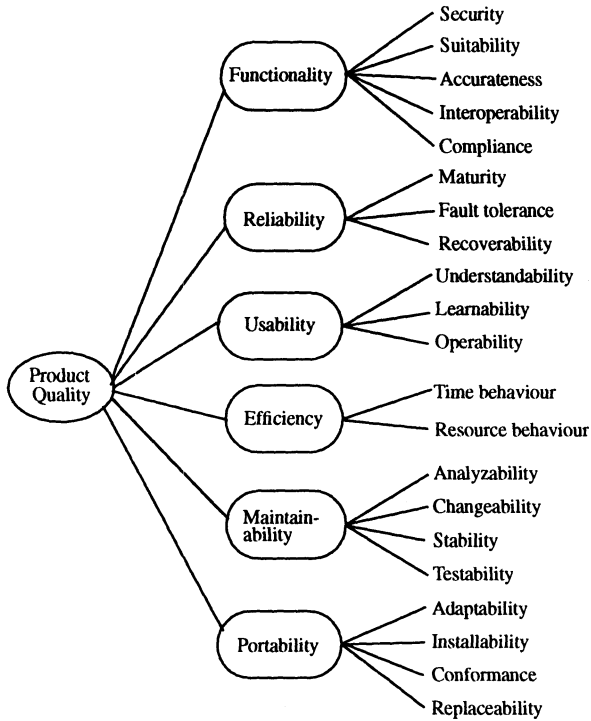


Figure 2. The ISO 9126 model for software quality.

By using the structure of the two quality models to structure the questionnaire, we got a questionnaire with the following structure:

Questions related to the software's price category, product category and application area

Questions related to product quality

1. Questions related to product quality factor 1
 - Questions related to criterion 1 for factor 1
 - ⋮
 - Questions related to criteria u for factor 1
2. Questions related to product quality factor 2
 - ⋮
 - Questioned related to product quality factor 2

Questions related to service quality

1. Questions related to service quality factor 1
- ⋮
- m. Questions related to service quality factor m

3 MAIN RESULTS

First of all, we looked at the overall score, i.e. the ranking of the factors that we got when pooling together the responses from all price and product categories. The score for each factor was computed as the average of the scores for all criteria related to this factor. This gave the following results:

Table 1. Scores for all products pooled together

Factor	Score
Service responsiveness	2.81
Service capacity	2.74
Product reliability	2.71
Service efficiency	2.65
Product functionality	2.60
Service confidence	2.60
Product usability	2.57
Product efficiency	2.46
Product portability	2.05
Product maintainability	1.89

At first glance, it might look strange that the functionality came in fifth place. This does, however, not imply that the functionality does not matter. Instead, it implies that if the software solves the problems at hand, the details of the functionality are not that important.

What surprised us was that the reliability came in third place. We found this so surprising that we conducted a follow-up review with four of our responders. The answers that we got can be summed up as follows:

Due to other software, such as for instance Windows, the machine crashes two to four times a week anyhow. As long as the product under consideration has a MTTF that is considerably higher than that, we do not care.

Small differences in scores in table 1 should not be used to rank one factor before another. A more detailed statistical analysis, built on the frequency of "high importance" scores is found in chapter 4.3

We split up the responses into three product categories according to type as follows:

- COTS - Commercial Off The Shelf - software
- Standard software packages
- Customized or tailor made software products

The number of responses for each category is shown in parentheses in the heading of table 2. The factor scores were computed in the same way as for table 1. We got the following results:

Table 2. Scores according to product category

Factor	COTS (36)	Standardized software packages (27)	Customized and tailored software (19)
Product functionality	2.55	2.65	2.62
Product reliability	2.65	2.77	2.74
Product usability	2.67	2.47	2.54
Product efficiency	2.47	2.35	2.58
Product maintainability	1.68	1.94	2.32
Product portability	2.01	2.04	2.09
Service confidence	2.55	2.59	2.67
Service efficiency	2.64	2.67	2.61
Service capacity	2.69	2.74	2.84
Service responsiveness	2.82	2.83	2.76

Table 2 shows that in all cases, the producer's service responsiveness and service capacity are considered to be of most important factors while product maintainability and portability are considered to be of the least importance. The only product factor that score consistently high is product reliability.

4 DATA ANALYSIS

4.1 Hypothesis to be Tested

In order to get a more in-depth understanding of the responses, we tested the following hypothesis:

- Ha: The requirements on product and service quality have the same important for COTS, standard software packages and customized or tailor made products.
- Hb: The requirements on product and service quality do not depend on the product's price.
- Hc: The importance ordering of the service and product quality factors is independent of product price.
- Hd: The importance ordering of the service and product quality factors is independent of product type.

In order to check the first two hypotheses, we reason as follows:

If a factor is more important in one product or price category than in another, then this factor will receive significantly more top scores - the rating "highly important" - than what would be the case if this factor was not considered so important.

In order to check the last two hypothesis, we reason as follows:

The ranking of importance is described by the score as it is assigned in tables 4 and 5 below. Thus, we can use the differences between the factor ranking for each category - price or product - as a measure of the independence of the factors across categories. If the differences are not significant, then the factor ranking are not different for different categories.

4.2 Methods for Analysis.

Since our responders graded each criterion separately, we used the average number of top scores per factor for our ranking process. This value was then divided by the number of responders in each group. The reason for only using the number of top scores, instead of adding the contributions from the other scores, is that it is not possible to assign numerical values to both “highly important” and “somewhat important” in such a way that these values mirror the absolute difference between these two categories.

When selecting statistical methods for testing the hypothesis, we decided to use non-parametric methods as far as possible. The reasons for this are that:

- The data that we have collected are categorical and thus only concerned with ranks.
- It frees us from the hard-to-validate assumption that our data are normally distributed. This - often unfounded - assumption plagues too many papers where statistical methods are used in software engineering research.

We based the tests for the hypotheses H_a and H_b on the Friedman statistics, described below. See for instance Lehman (1975). The method can be described by using the following table:

Table 3. Example of Friedman’s rank statistics

Factor	Category 1	Category 2	Category 3
F_1	R_{11}	R_{21}	R_{31}
:	:	:	:
F_N	R_{1N}	R_{2N}	R_{3N}
$R_{i.}$	$R_{1.}$	$R_{2.}$	$R_{3.}$

We see that if one of the categories consistently receive the best score then the rank sum for this column will be close to N , while a category which is consistently given bad scores will have a rank sum close to $3N$. If there are no differences, the score for all the columns will tend to be equal. The following formulas are used to compute the Friedman rank statistics:

$$R_{i.} = \frac{1}{N} \cdot \sum_{j=1}^N R_{ij}$$

$$Q = \frac{12N}{s(s+1)} \cdot \sum_{i=1}^s \left[R_{i.} - \frac{1}{2}(s+1) \right]^2$$

Here, N is the number of rows, while s is the number of categories (columns). As an approximation, we have that Q is Chi-square distributed with $s-1$ degrees of freedom. We will reject the hypothesis of no inter-column difference on the α -level if $Q > c$, where c is the α -percentile in the Chi-square distribution with $s-1$ degrees of freedom.

The testing of the two hypotheses H_c and H_d are based on the Spearman statistics, see for instance Lehman (1975). The idea is simple, as can be illustrated by using table 3 above. The statistics on which we will base our decision of hypothesis rejection or acceptance - here called D - is computed as follows:

$$D_{i,j} = \sum_{k=1}^N (R_{i,k} - R_{j,k})^2$$

The indices *i* and *j* are used to mark the columns that we test, while *N* is the number of rows in the two columns.

We see directly that if the factors are order in the same way in the two columns, we get *D* = 0. Thus, small *D* values indicate that the ranking of the factors in the two columns are dependent while a large *D* value indicates column independence.

For large *N*, the *D*s are approximately normally distributed with known mean and variance. This could be used to test the hypotheses. However, there exist tables that can be used to test the Spearman correlation for significance - for instance table *N* in Lehman (1975). This table is used here.

When we want to draw conclusions based on frequencies for individual quality factors, we will use a hyper-geometric model. The population is all companies that are members of DnD - The Norwegian Computer Union. If we assume that the *n*-sized sample we have is representative, we can estimate the population for category *i* as *N_i* = 1000 * sample size _{*i*} / total sample size.

As our *H₀* hypothesis we will assume that the two frequencies *p_i* and *p_j* are equal. By using the normal approximation and rejecting *H₀* if $\Delta = p_i - p_j$ is large we reject *H₀* at the α level if

$$\frac{|\Delta|}{\sqrt{Var(p_i) + Var(p_j) + Cov(p_i, p_j)}} > u_{\alpha/2}$$

See for instance Bissell (1994)

If we compare frequencies within a single column, sample size and population size are the same for *p_i* and *p_j*. By using the fact that $[p(1 - p)]^{0.5}$ is close to 0.5 for all *p*-values of interest and assuming independence, we obtain

$$\Delta_{\alpha} > u_{\alpha/2} \sqrt{\frac{N-n}{2nN}}$$

Note that this is a quite conservative limit. Even a small correlation between the quality factors will lower the Δ -value considerably.

4.3 Results from the Analysis

We used the data lay-out in the tables below to test the hypotheses. Each table entry has the format <average number of top scores / rank within table row>. Next to this, we have inserted the column rank for each factor within each category.

If two or more table entries have the same value, they are first grouped together, given consecutive ranks and are then all given the rank equal to the average of all the ranks for this group.

For the product categories in table 4, we get *Q* = 0.11. With an α of 0.05, we find that we can not reject the hypothesis of no difference - *H_a*. We can thus pool the data from all the columns. If we use the frequencies based on the pooled data, we can also group the quality factors according to importance. In order to get a clear picture of the situation, we will order the quality factors on a frequency line as shown in figure 3 below. The acronyms are explained in table 4.

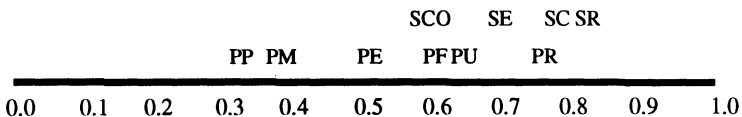


Figure 3. Distribution of "high importance" for quality factors

The Δ -value will in this case be 0.12. We can now group the quality factors into three groups with high, medium and low user importance respectively:

- High: service responsiveness, service capacity, service efficiency and product reliability
- Medium: service confidence, product usability, product functionality and product efficiency
- Low: product maintainability and product portability

We see that those who insist on equating product quality with product reliability have a strong case, at least from the users point of view.

Table 4. Factor importance versus product category

Factor	COTS (36)		Standardized software packages (27)		Customized and tailored software (19)		Pooled data (82)
Product functionality - PF	0.67 / 1	4.5	0.58 / 2	6	0.50 / 3	9	0.60
Product reliability - PR	0.67 / 3	4.5	0.79 / 2	2	0.81 / 1	2	0.74
Product usability - PM	0.74 / 1	2	0.51 / 3	7	0.63 / 2	6	0.63
Product efficiency - PE	0.51 / 2	7	0.43 / 3	8	0.60 / 1	7	0.50
Product maintainability - PM	0.29 / 3	10	0.37 / 2	9	0.52 / 1	8	0.37
Product portability - PP	0.32 / 3	9	0.35 / 1	10	0.33 / 2	10	0.33
Service confidence - SCO	0.40 / 3	8	0.74 / 1	4	0.73 / 2	3	0.58
Service efficiency - SE	0.66 / 3	6	0.73 / 1	5	0.71 / 3	4	0.69
Service capacity - SC	0.72 / 3	3	0.78 / 2	3	0.90 / 1	1	0.78
Service responsiveness - SR	0.85 / 1	1	0.83 / 2	1	0.65 / 3	5	0.80
Average column rank	2.3	-	1.9	-	1.9	-	-

In addition to variation over product types, we were also interested in the effect of product price.

Table 5. Factor importance versus product price

Factor	Price category 1 (20)		Price category 2 (22)		Price category 3 (42)	
Product functionality	0.62 / 3	5	0.79 / 1	2	0.73 / 2	5.5
Product reliability	0.60 / 3	6	0.82 / 1	1	0.77 / 2	2
Product usability	0.70 / 1	2	0.60 / 3	7	0.64 / 2	8
Product efficiency	0.43 / 3	8	0.57 / 1	8	0.51 / 2	9
Product maintainability	0.33 / 2	9.5	0.38 / 1	9.5	0.07 / 3	10
Product portability	0.33 / 3	9.5	0.38 / 2	9.5	0.76 / 1	3.5
Service confidence	0.65 / 3	3	0.67 / 2	6	0.76 / 1	3.5
Service efficiency	0.59 / 3	7	0.72 / 2	4	0.73 / 1	5.5
Service capacity	0.75 / 3	1	0.77 / 2	3	0.81 / 1	1
Service responsiveness	0.64 / 3	4	0.71 / 1	5	0.69 / 2	7
Average column rank	2.7	-	1.6	-	1.7	-

For the price categories in table 5, we get $Q = 7.4$. With an α of 0.05, we find that we can reject the hypothesis of no difference - H_b . A quick glance at the table also shows that the requirements on product and service quality is much lower for the products in the lowest price category.

For the three columns used in the two tables 4 and 5 we find the following D values for the inter-column correlations:

Table 6. D values for Spearman statistics

Table number	Columns 1 and 2	Columns 1 and 3	Columns 2 and 3
Table 4 - Product category	23.10 *	96.50	26.00 *
Table 5 - Product price	82.00	101.00	68.00 *

From table N in Lehman (1975) we find that the correlations marked with an asterisk are significant at the 5% level. As we see, we can at least partly reject the two hypotheses H_c and H_d of no correlation. The correlation between neighbouring columns and the lack of correlation between the first and the third column can best be explained by the accumulation of small changes as we move from left to right in the table.

Once we had established the possible correlations, we found it fruitful to see what caused the correlation changes. In order to study this, we decide to look at which factors moved and which did not. Table 7 shows these movements over product categories. The differences shown in the table are computed as the differences between the leftmost and the rightmost categories (columns).

Table 7. Changes in relative top scores for product and price changes.

Factor	Score differences	
	Product categories	Price categories
Product functionality	- 0.17	+ 0.11
Product reliability	+ 0.14	+ 0.17
Product usability	- 0.11	- 0.06
Product efficiency	+ 0.09	+ 0.08
Product maintainability	+ 0.23	- 0.26
Product portability	+ 0.01	+ 0.43
Service confidence	+ 0.33	+ 0.11
Service efficiency	+ 0.05	+ 0.14
Service capacity	+ 0.18	+ 0.06
Service responsiveness	- 0.20	+ 0.05

As we move from COTS products to tailor-made products, product maintainability and service confidence get more importance, together with service capacity. The reason for this is that customers who buy tailored software do not expect the producer to put new, improved product versions on the market, except if they have a maintenance contract. Since this is a long term commitment, the requirements on service responsiveness decreases.

As the price increases, the customers increase their requirements on product reliability and portability. At least the latter of these two effects can be seen as a result of the customers' need to protect their investment in expensive software systems. They need to be able to port expensive software so that they do not have to buy new versions when the environments are changed.

The conclusion of this discussion is that software customers behave in an economically rational manner. Thus, their behaviours and preferences can be deduced from simple economic considerations.

5 RELATED RESULTS

There are two other survey results published at about the same time that are relevant to our results. One was published in *ComputerWorld Norway (1994)* and the other was published by a European provider of telecommunications software. The results from *ComputerWorld* are related to software on mainframe computers and are summed up in the table 8 below. For the sake of comparison, we have shown the overall results from our survey in the same table. Since the *ComputerWorld* survey used a scoring scheme different from ours, we have only shown the factors' ranking.

Table 8. Comparison of the PROFF and the ComputerWorld surveys on quality requirements

Our survey (PC and work station systems)	ComputerWorld survey (mainframes)
Service responsiveness	Product reliability
Service capacity	Product efficiency
Product reliability	Service responsiveness
Service efficiency	Operating system functionality
Product functionality	Product capacity

We see that for mainframe users, product reliability is rated as most important, with product efficiency in second place. This reflects, in our opinion, the much higher reliability of the supporting software on a mainframe than on PCs and work stations. In this case, the product reliability will get a much higher priority. In addition, the ranking also reflects the higher degree of professionalism of the main-frame sites.

The results pertaining to the telecommunication software provider had a different focus. The questions they asked their customers were "Would you buy from this company again?" and "Would you recommend others to buy from this company?". The results are shown in the tree in figure 3.

At a first glance it is rather surprising that the only case where we get a 100% yes on the question "Would you recommend others to buy from this company?" is if the customer has complained and got a satisfactory response. In the case of no complaints this question only got 87% yes.

The conclusion seems to be that for the customers it is more important with a satisfactory service response than having no problems at all.

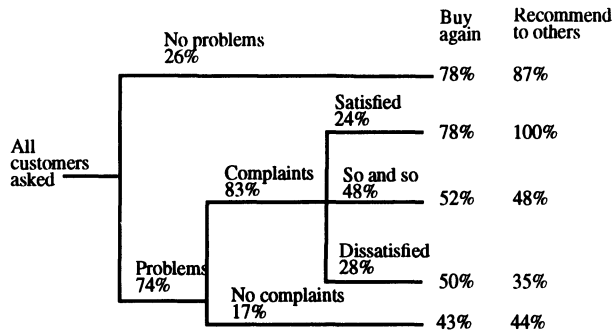


Figure 4. Customer reactions to software product problems

6 CONCLUSIONS

The conclusions are split into two parts, namely the conclusions based on the data analysis alone and the conclusions that are based on their interpretations.

6.1 Conclusions based on the Data Analysis.

The results below are based on the previous data analyses and discussions:

The first, and perhaps the most important, result of our survey is the strong focus on service quality. This can be seen already from the first summary of the results in table 1. The result is reinforced by the related surveys cited in the second half of chapter 5.

Next, we find that the customers' quality requirements increase with the price of the product. This is as should be expected - hypothesis Hb.

However, there are no significant difference in quality requirements between products of different categories - at least not as these are defined in our survey, namely COTS, standardized software packages, customized or tailor made - hypothesis Ha. This is reasonable since these categories describe how the product is marketed and sold and not the customers' quality needs or expectations.

Last, but not least, we found that there are a large amount of agreement over the relative importance of the quality factors between different product types and a somewhat lower degree of agreement between on the rating of the quality factors different price categories - hypotheses Hc and Hd.

6.2 Conclusions based on Interpretations of the Data Analysis

Firstly, software products are considered to be complex and having a large potential to cause problems that the average user has little competence in handling. For main-frame users this cause few problems since they have access to a competent service organization in-house. For other users, however, this has led to a strong focus on service quality.

Secondly, maintainability and portability are not considered important for the average software customer. They consider it the producer's responsibility to provide updated and enhanced versions of the software on the platform that they need.

For the producer on the other hand, maintainability and portability are important since they use these characteristics to provide the services that the customer requires.

Thirdly, the reliability of the software product is only important if its MTTF is comparable to the MTTF of the rest of the system. As long as the system's MTTF is much larger than that of the platform, reliability is not an issue for the average user.

Fourthly, software customers behave in an economically rational way when they set priorities on quality requirements. Their preferences can be deduced from simple economic relations. If the pricing policy or market mechanisms change, so will the importance of each quality factor.

In our opinion, the customers' strong focus on service quality is an opportunity and not a problem. The advice is: Do not just sell a product; sell a product plus service, provided through personal contacts. The personnel used to provide these services must know the customer's applications and his problems and needs. In addition, he must show a genuine interest in that the customer reaps the benefits he expected from the product that he bought.

7 FUTURE WORK

The next phase of work in the PROFF project is to decide on estimators for all important factors. We have already defined estimators for the product quality factors, based on data that are observable either at the developer's site or at the customer's sites.

We will later also define estimators for the service quality factors. In addition, we will define metrics related to product and development process that will help the developers to satisfy specified customer quality requirements. We hope to have achieved most of these goals by the end of 1995.

8 REFERENCES

- ComputerWorld (1994), ComputerWorld's guide to main-frames. ComputerWorld Norway, no. 30B, September 9, 1994 (in Norwegian)
- Lehman, E.L. (1975), Nonparametrics: Statistical Methods Based on Ranks. Holden - Day Inc.
- Bissell, D. (1994) Statistical Methods for SPC and TQM, Chapman & Hall Statistics Textbooks Series
- Arnesen (1995), Arnesen, K., Borgersen P. Chr. and Stålhane, T., Final Report from the Software Quality Survey, PROFF report no. 1, (in Norwegian)
- ISO/IEC 9126 (1991), Information technology - Software product evaluation - Quality characteristics and guidelines for their use, ISO
- ISO/IEC 9004 (1987), Quality management and quality elements - Guidelines for services