

Quality judgements : a general research design

Citation for published version (APA):

Dijkstra, L. (1991). *Quality judgements : a general research design*. (TH Eindhoven. THE/BDK/ORS, Vakgroep ORS : rapporten; Vol. 9104). Technische Universiteit Eindhoven.

Document status and date:

Published: 01/01/1991

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

**QUALITY JUDGEMENTS:
A GENERAL RESEARCH DESIGN**

Dr. Lieuwe Dijkstra

Report TUE/BDK/ORS/91/04

**Eindhoven University of Technology
Graduate School of Industrial Engineering and
Management Science**

QUALITY JUDGEMENTS: A GENERAL RESEARCH DESIGN

Introduction

In an earlier contribution, Govers (1992) evaluated the application of the vignette method in quality judgement research in a campsite project [1]. In this report we will concentrate on some general methodological and statistical aspects of the vignette method in customer research. The main point of interest is the application of the vignette method in Quality Function Deployment (QFD) [2].

The first part is about the objectives in quality judgement research. One objective is a descriptive quality profile of a product or a service. The other objective is to get insight into the factors which determine the profile. The vignette design is presented as an appropriate means for this last objective. In the second part, problems arising from the application in practice of the vignette design are discussed. In the third part three main versions of the design are outlined. In the fourth part some general methodological problems connected with the vignette design are discussed.

1. Evaluating Quality Judgement

Measuring the judgement of the quality of products and services is usually carried out by presenting a set of statements which describe the appraisal of a product. Respondents are invited to express their agreement or disagreement with these statements by choosing a response modality ranging from 'very bad' to 'excellent', and the like. Both design and level of measurement range from simple (one or a few statements; 5 points Likert scales) to complex (semantic profiles). Correspondingly, analysis and results range from simple (means and standard deviations) to complex representations of cognitive structures (e.g., factor or cluster models; multidimensional scaling).

The design outlined here can be extended to a 'needs-beliefs' model by adding the desired properties, attributes or 'performances' of product X in its most ideal form to the semantic profile of several concrete, existing brands of a product, e.g., detergents, walkmen, soft drinks.

In all these cases, the result is a more or less sophisticated quality profile, ranging from a simple 'good/bad' judgement to a description of specific aspects of the product, e.g., price, user friendliness, design, adjustment possibilities. This profile can even be related to qualities of the respondents, such as income or lifestyle.

From the point of view of product development, however, the problem of which factors or elements determine the judgement, is more interesting. As a consequence, the research attention shifts from product evaluation towards a 'customer wants' approach as starting point for product development. The question is no longer: what is the judgement of an existing product, but: which factors determine this judgement? What makes a judgement negative or positive, regardless of existence or nonexistence of a product?

The quality profile does not provide an answer, although the needs-beliefs approach comes close. In this paper a design based on the vignette method is offered which enables the elaboration of this problem. By using this method, the contribution of specific judgements to a general judgement can be determined. In figure 1 the two main types of research objectives are shown.

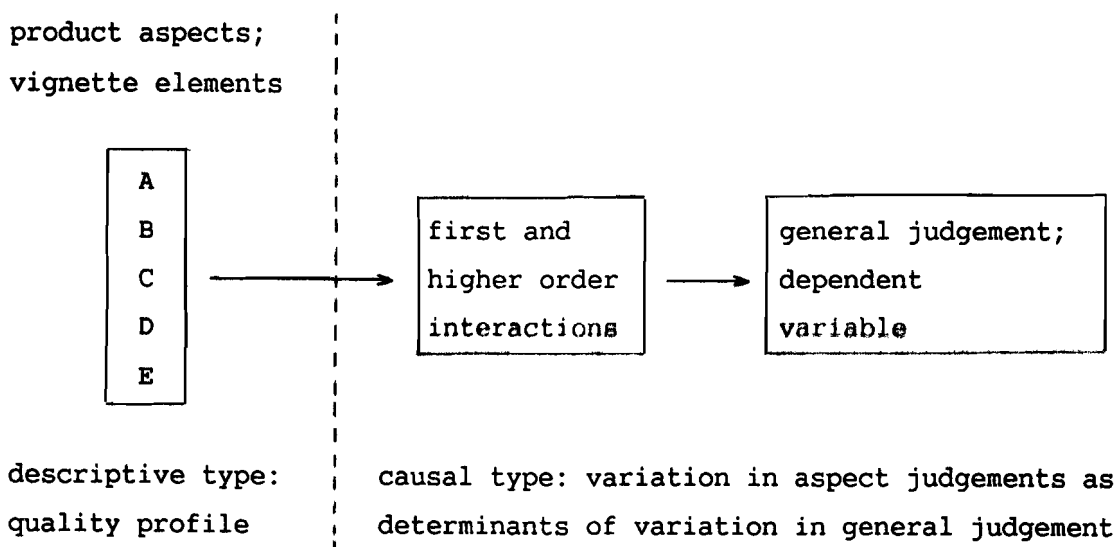


Figure 1. Two related types of product evaluation.

The vignette method was developed by Rossi (1982), in a series of research projects, mainly on the subjectively experienced fairness of income distribution and on the determinants of social status [3]. Rossi created fictitious households by describing age and educational level of husband and wife, their profession(s), income(s) and housing, the number, gender and age of children, and by putting questions about the experienced fairness of the family income or the social status of the family.

In other words, fictitious households were created (the vignettes). The vignettes could be based on real existing households and were judged on the perceived fairness of their incomes or their social status. The theoretical sociological background is not relevant for this paper, as we are primarily concerned with the method. This method comprises the advantages of an experimental design under realistic conditions.

The judgement given about the fairness of incomes or social status is the dependent variable here; the vignette elements are the explanatory independent variables. By randomization of the values of the vignette elements and by presenting a number of vignettes to respondents, a data set is obtained that can be analyzed by the methods of multivariate analysis, analysis of variance or multiple regression, for example. These models have in common that variation in a dependent variable, or in a set of dependent variables, is described in terms of variation in the independent variables.

This design can also be used to find which factors or elements determine a general quality judgement and to what extent.

2. Vignette construction in practice

In the application of the vignette design in quality research, two main parts should be distinguished. The first part consists of the identification of relevant vignette elements. The second part is the creation of vignettes, and the collection and analysis of the data.

Compared with sociological research, in quality research the identification of vignette elements occupies a special place, because now in every single case decisions have to be made about the concrete terms of the elements. In sociological research these concrete terms form part of a theory or a theoretical framework and are known in advance, as is demonstrated by the above mentioned vignettes describing households.

Furthermore, a distinction should be made between factors that can be influenced and factors that are difficult to control, such as natural

circumstances, transport, environmental conditions, or economic and political developments. If important factors are found which can be influenced by strategic policy, the cause-to-effect knowledge can be used in order to bring about a purposeful change in the quality judgement of existing products or to adjust a product to customer wants. Another application might be in Quality Function Deployment: a systematic place at the start of the process of product design and development.

In QFD 5 main levels can be distinguished:

- customer requirements;
- design requirements;
- component characteristics;
- manufacturing operations;
- product requirements.

QFD works from the top down and successive levels are connected by matrices; QFD can be seen as a chain of matrices, where the first matrix is the input to the second, the second is the input to the third, etc.

The identification of relevant vignette elements can take place in many ways, depending on the problem. Sometimes a few expert interviews may be sufficient. In other situations the identification problem may even require a research project of its own, when, for example, the cognitive structure of potential elements is needed to make decisions on terminology and formulations to be used in the vignettes. Adequate methods and designs are known in the fields of survey and small group research. In our pilot projects different solutions for the identification problem were explored.

In the campsite study, the question was: which factors determine the general judgement (the attractiveness) of a campsite? In this project, in a group session experts (campsite managers and professional workers in tourist organizations) rapidly agreed on the 5 aspects given below, to be presented to camping guests in vignettes:

- A guest reception pleasant/chilly
- B level of facilities simple/extensive
- C privacy limited/sufficient
- D staff behaviour correct/bad manners
- E maintenance of campsite rules strictly/loosely

In the bakery additives project the aim was to find which factors determine the general judgement of bakers about the service quality of a firm from which they order their additives. For that purpose, 17 firms' representatives were invited to list the elements which in their opinion played a part in the judgement of bakers.

It appeared, however, that the representatives even after an extensive introduction to the research design, still didn't have a clear idea of their task in making this inventory. Afterwards, the main reason was that in asking for the elements, it was not realized that the issues are mainly determined by the personal relationship between the representatives and their customers. A codification of terms did not exist.

The solution was a two step procedure. In the first step all the elements which had been mentioned in an open questionnaire were simply listed. In the second step this list was presented to the agents, who scored each item on a 5 points scale, ranging from 'very important' to 'very unimportant'.

A cluster analysis resulted in 6 clusters. The analysis was carried out somewhat unconventionally. The items were submitted to a procedure to unfold clusters which met the requirements of a Likert scale [4]. As a consequence, items in a cluster share a latent factor which is (in terms of this scaling model) the exclusive cause of systematic variation of the items in the cluster. After inspecting the clusters, each cluster was replaced by the subject matter that seemed to represent the best the items in the cluster. This resulted in the 6 vignette elements below (all in positive formulation):

- A partnership: company shows involvement; suggests technical and commercial renovations;
- B company provides uncalled information on consumer purchasing trends, materials and new products;
- C assortment: sufficient stock, clear instructions on recipes;
- D company always delivers at right time and to right place;
- E company is easy to communicate with; has a good knowledge of sector, materials and techniques; we speak the same language;
- F company is reliable, trustworthy; keeps its word;

This solution has a weak foundation. When it became evident that among the agents there was no consensus about a limited set of important items which could function as vignette elements, the judgements about the importance were forced interpreted as measures of closeness. Instead of this, the second step should have consisted of a cluster analysis of pairwise similarity/resemblance/communality of the listed items.

3. Choosing the design

Once the problem of identifying the vignette elements has been solved, a design can be chosen. Three main versions can be distinguished: the complete factorial design, the incomplete factorial design, and the randomized factorial design.

In the case of the **complete factorial design**, to each vignette element two semantically opposed values are ascribed. Thus, with k elements, the number of vignettes is 2^k . If the judgement is also dichotomous, an appropriate data model is the loglinear effect model, with the vignette elements as the independent variables [5]. The conditional frequency ratios of the classes of the dependent variable in the saturated model indicate single and interaction effects which can easily be interpreted and compared, as well as compared with the objective standard for no effects.

This also holds true if the dependent variable (the quality judgement) is measured on an ordinal scale. If the level of measurement is interval, analysis of variance, of course, is the appropriate data model.

In the campsite project, the positive/negative judgement ratios for single positive elements are listed below:

C 'privacy'	3.024
D 'staff'	2.191
E 'rules'	2.167
A 'reception'	1.986
B 'facilities'	1.698

The interpretation of the ratios is: if privacy is sufficient, the ratio positive/negative judgements is 3.024 : 1 (75% : 25%). If reception is chilly, the ratio positive/negative becomes 1 : 1.986 (33.5% : 66.5%).

The incomplete factorial design differs from the complete version because in this case some vignettes are excluded. For example, if a pilot study shows that judgements will be negative as soon as a vignette contains 3 or more negative elements, the design could be restricted to vignettes with just 1 or 2 negative elements. The design can still be complete within this restriction. When second order and higher interactions are of no interest an incomplete design is also preferable.

The most important problem with the incomplete design is the determination of single and interaction effects. A solution has been found in using an (ordinal) 6 points scale for the expression of the general judgement. Next, the frequency patterns of the vignettes are ranked with the help of the theta statistic, a measure of association for a nominal and an ordinal scale [6]. In the bakery additives project, this resulted in the ordering of the vignettes with just one negative element below:

negative vignette element	judgement						mean
	negative			positive			
	6	5	4	3	2	1	
D 'delivery'	3	3	19	9	0	1	3.91
F 'reliability'	2	1	10	14	6	3	3.17
C 'assortment'	0	0	5	15	16	0	2.69
B 'information'	0	0	4	9	15	3	2.45
A 'partnership'	1	1	1	9	17	5	2.38
E 'communication'	0	0	2	10	16	6	2.24
('D + F':	3	6	18	5	0	0	4.22)

Bringing in vignettes with two negative elements, for example 'delivery' and 'reliability', results in the frequency pattern 'D + F'. Here the cumulative effect of negative elements becomes clear.

The randomized factorial design consists of creating vignettes by combining randomly chosen values for elements. This holds true for continuous scales (e.g., age or income), and for multicategorical nominal scales (e.g., profession or religious affiliation).

There is a method for obtaining correlations between vignette elements, corresponding to known empirical relationships (e.g., educational level and income). Starting from the marginal totals of two or three vignette elements, the iterative proportional fitting algorithm [7] can be used to create a distribution with exactly the deviation from independence that is appropriate.

4. Discussion

Two main methodological phenomena should be mentioned here: the halo effect and the leniency effect [8]. Both have been found in our projects.

The halo effect occurs when specific judgements are deduced from one general judgement. This phenomenon is well known in social psychology. If my general opinion about Marcia is very positive, I will have a tendency to state that, more specifically, Marcia is good in maths, plays the piano very well, has excellent social skills, speaks French fluently and is a caring mother. On the other hand, a person who is labeled bad generally, will be surrounded by an aura of negative specific qualities.

The leniency effect or 'positive bias' is the tendency to express less negative feelings or judgements, the more the subject matter becomes important. If social skills are of vital importance to Marcia's profession, I probably hesitate in casting my doubt on her social skills; but if she doesn't need French at all, I will feel more free to suggest that her French might be bad.

On the halo effect we have an adequate check, consisting of the model of the Likert scale. The basic idea of this model is that a latent, unmeasured factor is the exclusive cause of variation in a set of measured variables. As a consequence, these measured variables should show some predictable characteristics concerning homogeneity (expressed in Cronbach's alpha) and internal consistency (expressed in item rest correlations) [4].

Both phenomena are illustrated by data from the bakery additives project. The halo effect is demonstrated by the item rest correlations and by Cronbach's alpha for 5 out of the 6 vignette elements. Respondents were also asked to judge the additives company on these vignette elements. The results are reported below:

vignette element	item rest correlation
A 'partnership'	.63
B 'information'	.70
C 'assortment'	.73
D 'delivery'	.68
F 'reliability'	.59
Alpha	.85

Alpha and item rest correlations are sufficiently large to accept the 5 elements as a subscale.

The leniency effect is demonstrated below by the judgements on single negative vignettes and by the judgements about the company, corresponding to these vignettes (ordering in conformity with the theta statistic):

vignette element	most (1) to less (6) negative vignette elements			
	judgement about company		single in vignettes	
	mean	rank number	mean	rank number
E 'communication'	2.67	(1)	2.24	(6)
B 'information'	2.42	(2)	2.45	(4)
A 'partnership'	2.46	(3)	2.38	(5)
C 'assortment'	2.30	(4)	2.69	(3)
F 'reliability'	2.28	(5)	3.17	(2)
D 'delivery'	1.94	(6)	3.91	(1)

It should be noticed that the halo effect is present in most general judgements, but not always visibly. In the vignette method the most can be made of this by adding up scores on specific judgements which fulfill the requirements of the Likert scale model. It can be proved that such sum scores are valid and reliable representatives of the latent factor (the general evaluation), much more than its component parts or the general judgement itself.

Furthermore, the leniency phenomenon holds the warning that the more vignette elements are negatively evaluated, the more respondents will be reticent in showing negative feelings relating to this aspect of a concrete object. This may cause an unjustified rosy impression and lead to the false conclusion that no steps have to be taken.

Notes and References

- [1] Govers, C.P.M., The Judgement of Service Quality, in: Kunst, P. and J.Lennink (Eds.), Quality Management in Services, Assen/Maastricht, 1992 (forthcoming).
- [2] Bossert, James L., Quality Function Deployment, A Practitioners Approach, ASQC Quality Press, Milwaukee, Wisc., 1991.
- [3] Rossi, P.H. and A.B.Anderson, Measuring Social Judgements, The Factorial Survey Approach, Beverly Hills, 1982. After several earlier publications by Rossi, this is a review of the method and its applications in sociological research.
- [4] Dijkstra, L., The Likert Attitude Scale: Theory and Practice, University of Technology Eindhoven, Report TUE/BDK/ORS/91/03. This report includes a computer program in Pascal, which enables cluster analysis, based on Likert scales. The subscales which are constructed on the basis of Alpha and item rest correlations are considered as clusters; the inter cluster distances are defined as the mean correlations between subscales.
- [5] Bishop, Y.M.M., S.E.Fienberg and P.W.Holland, Discrete Multivariate Analysis, MIT Press, Cambridge, Mass., 1975, Ch. 2.
- [6] Freeman, L.C., Elementary Applied Statistics, New York, 1965, Ch. 10, pp. 108-119.
- [7] The algorithm is described in Bishop et al., op. cit., pp. 73-102.
- [8] Freedman, J.L., D.O.Sears and J.M.Carlsmith, Social Psychology, Prentice Hall, 1978(3), pp. 84-88.