MEASUREMENT AND EVALUATION OF DETERMINANTS OF JOB SATISFACTION

Dissertation
in order to obtain the degree of
Master in Social Sciences Data Analysis
and Collection
of the University of Essex

JOAN MANUEL BATISTA-FOGUET

(University of Barcelona)

PREFACE

The present dissertation has been written with the assistance of several people . I would first like to mention Professor Willem E. Saris who supervised my work.

I am also greatful to X. Tort-Martorell and J. Ginebra from the "Departament de Tècniques Quantitatives de Gestiò. Universitat Politècnica de Catalunya" (Barcelona) who helped to clarify many ideas about the last section.

I would also like to thanks Martin Verwey from the Sociometric Research Foundation (Amsterdam) and Bas van den Putte from the Department of Methodology of the University of Amsterdam, for their useful comments on my work.

I have to emphasize that this research has been made possible by a grant of the "Fundaciò Jaume Bofill" provided to me for the whole period of seven months which I spent during the last two years in Amsterdam. The Sociometric Research Foundation (Amsterdam) provided me with the necessary facilities and the money for the data collection.

Last but not least I would like to thank the teachers of the "Essex Summer School" for their efforts to teach me the state of the art in the Social Sciences Data Analysis and Collection.

Without the support of all these people and organizations it would have been difficult to write my dissertation.

Table of Contents

pages

Part 1. Objectives and design of the data collection

1. Introduction 4
2. Preliminary 5
2.1 Objective versus subjective indicators 5
2.2 Life as whole and the specific life domains 5
2.3 The cognitive and affective components of the Well-Being 6
3. Objectives of this research 8
4. Measurement and Analysis problems 9
4.1 First measurement problem: Categorical versus continuous data 9
4.2 Second measuremnt problem: Variation in Response Function 10
4.3 The data analysis problem 11
5. The design of the experiment 13
5.1 The chosen aspects of Job satisfaction 13
5.2 The different instructions 14
6. The sample 18

Part two. Measurement aspects

7. The measurement model 21
8. The individual data analysis 24
8.1 Validity and reliability 24
8.2 Reduction in the Variation in Response Function 26
9. The consequences for cross sectional data analysis 29
10. Discussion of the results 33

Part three. The analysis aspect

11. The Job satisfaction experimental data analysis 36
11.1 Estimation of the four factors' effect on Job satisfaction. The use of design at two levels 37
11.2 Discussion of the results 41
12. The utility of using fractional factorial designs in questionnaires. The length reduction 43
13. Conclusions 46

References 50

PART 1. OBJECTIVES AND DESIGN OF THE DATA COLLECTION

1. INTRODUCTION

Since Andrews and Whitey (1974) paper " Developing measures of perceived life quality: results from several surveys", which actually represents a landmark for the assessment of perceived Life-quality, an enormous number of papers have appeared trying to cover more or less the same objectives.

According to Dienner (1984), psychological Well-being has become a major focus of applied and theoretical research, with over 700 studies published during the past 25 years. The topic has become so popular that political and social scientists jointly with psychologists publish a specialized journal, Social Indicators Research. Therefore it is a rather important topic and it makes sense to develop measurement instruments to evaluate the determinants of Life-quality (LQ) while trying to overcome some of the problems detected in this literature.

During the last decades substantial progress has been made in finding effective ways to measure people's sense of well-being (WB), in learning how measures of various aspects -domains- of LQ interrelate, and in evaluating how they contribute to the overall sense of WB.

In the literature concerning this topic, specific concepts and definitions have been developed which we will refer to before setting up the core of this study.

2. PRELIMINARY

2.1 Objective versus subjective indicators

Since concepts such as sense of WB or LQ are unobservable constructs indicators to measure them are required.

Traditionally the common division in subjective versus objective indicators was accepted. The latter are based on reports of events or situations which do not concern respondents own life, that is demographic or background variables. Nowadays this line of research has lost the popularity it had in the past because of the modest results. In Campbell (1981) words " If we try to explain the population's sense of WB on the basis of objective circumstances, we will leave unaccounted for most we are trying to explain".

At present, it is well known that these concepts should be measured by individual evaluation which is entirely subjective. The perceived overall sense of WB is not determined by individual's objective circumstances but mainly by the psychological perspective (values, expectations and personality traits) from which the individual perceives the circumstances (See Abbey and Andrews, 1985 and Wish, 1986).

2.2 Life as a whole and the specific life domains

The general sense of WB of the people is determined in a large extent by what the individuals themselves experience in the most important role-situations-domains (self; family life; work; friends; etc). As a consequence the LQ literature has devoted many pages to identify them, to find out how they are related, and to supply adequate indicators to tape these life domains.

Multivariate statistical techniques such as Factor Analysis (Smith et al. 1969; Kalleberg, 1972; Hodson, 1986) Principal Components analysis and cluster analysis (Andrews and Withey, 1976) or smallest space analysis (Levy and Guttman, 1975) have helped researchers to identify these domains from a large list of items as semi-independents concepts in the factors space.

Once these domains were stated they were used as predictors to evaluate their contribution to the overall sense of WB by means of structural equation models or regression techniques (See among others: Burt et al 1979 ,Campbell, 1981, Horley and Little, 1985 or Hodson,1986)

In this study we restrict the analysis to the Job-domain and four usual aspects (values): job-security; salary; atmosphere; and education agreement, are considered to evaluate their contribution to the job-satisfaction (See section 5.1). Since work itself constitutes a domain of the perceived LQ that is wholly subjective, it should be also measured by asking people their own opinion in assessing different job situations. Therefore the questionnaire used will constitute a set of subjective indicators to cover this domain. This questionnaire is described in section 5.2

2.3 The cognitive affective components of Well Being

Having established that we will deal with job specific subjective indicators, it is necessary to refer to the two major dimensions of the perceived WB, i.e., Happiness and Satisfaction, in order to set up the framework of our research.

In spite of the fact that these two components did not become regular features of the sociological and psychological abstracts until 1973 (See Larsen et all, 1985), since then, every article concerning this topic necessarily includes this distinction. specially after the McKennell's paper (1978) which clarifies the structure of the perceived global WB model into the cognitive and affective factors.

Campbell (1976) already noted that some of the measures, namely, those dealing with "satisfaction" with life, seem to reflect primarily the "cognitive" dimension of life. That is, responses to these measures reflect an intellectual process, in his own words "... the individual compares his perception of his present situation to a situation which he aspired to, expected or felt he deserved". On the other hand, other measures capture more the affective aspects of experience, i.e., the degree of "happiness".

Recent developments indicate that the cognitive and affective component may be somewhat independent of each other. In a review and reanalysis of pertinent findings by Zanjoc (1980) it is suggested that affect and cognition are under control of "parallel, separate, and partly independent systems in the organism".

Since measures of satisfaction reflect basically cognitive evaluation rather than affective states, satisfaction is normally considered as a more stable quality of life indicator than happiness, which is more susceptible to fluctuations of mood that may vary from day to day.

Due to this constancy on the people's level of satisfaction, it was the selected component. Since in our research the same questions about Job Satisfaction (JS) would be recorded in three different points in time (See section 4)

Before to concluding this introductory section it would be necessary to point out that, we accept the arguments given by Quinn and Mangione (1973) or Organ and Near (1985) that criticize the importance attached to this cognitive component in the evaluation of the Job Quality, because affective states of work were probably of much more interest than the cognitive ones. In spite of this fact the argument given in the previous paragraph explains why our research is focussed on subjective measures of this cognitive component.

3. OBJECTIVES OF THIS RESEARCH:

The present study aims to measure and evaluate determinants of J S. Therefore we have a double aim, namely, to develop a measurement instrument to collect data and to build a model to analyse these data in order to explain J S.

Sociological studies of J S concentrate on efficiency of work and productivity. They explain J S by occupational status, specialization of tasks, styles of supervision, opportunities for improvement or corporate structure than in the psychological line pointed in previous sections. (See Wannous and Lawler, 1972; Miller, 1980; or Hodson, 1986).

Given the large amount of work already done (Hodson, 1986), it might seem a naive objective to repeat what other researchers have done. However, we are going to consider just a minor part of Hodson's model. To be precise, the four factors of our model can easily be identified to what in this literature (Kalleberg, 1977) are called job rewards. In addition to these rewards Hodson's model includes background characteristics of the individual and the job as well as corporate characteristics such as organizational size, capital intensity, degree of subsidy and profit rate. This study does not pretend to shed new light in these discussions.

The new part in this study is that we use a different measurement instrument and secondly that we use an experimental design instead of a non experimental design. Both aspects are new and discussed below.

4. MEASUREMENT AND ANALYSIS PROBLEMS

Like many other Life quality researchers we are involved in searching a set of measures of perceived job-quality that could be used as social indicators, but before centering our attention on "What" explains JS we will discuss "How" it should be measured.

What we have seen in the literature is a proliferation of measures of perceived WB which normally provide 3,5,7,9 or 11 points answer scales (See Abbey and Andrews, 1985 or Larsen et al, 1985), that is, data collected in categorical scales which afterwards are used for the analysis as continuous scales. Thus, this will be the first aspect to deal with.

4.1 First measurement problem: Categorical versus Continuous data

In Psychophysics it is well known that the respondents are able to supply information on a continuous scale. Krantz et al (1974) showed that the information is stored by the individuals in a continuous scale, and there is no reason to believe that the respondents can only express their opinions in a categorical scale. Since on this scales, the subject is anchoring himself in a cell, the retrieved information will result at ordinal level at maximum.

In addition interval or ratio scales are normally required to satisfy for the assumptions of the statistical techniques commonly used in this field for the data analysis. Obviously,this assumption is violated when categorical scales are used to collect data.

Moreover, a frequent focus of interest in perceived WB, which concerns comparative studies (See Szalai and Andrews, 1980) is to look for constancies or-and differences among countries. A problem many authors found in the data used to carry out these comparisons, was that no standard category scale was available so they did not trust the measures provided by different studies

(See McKennell et al, 1980). Already in 1976 Andrews and Withey asked themselves:

"To what extent do the data produced by the various measurement methods indicate a person's true feelings about his life ?"

"How clear are the meanings of the categories used to describe people's feelings ?"

Stevens (1975), Saris et al (1977), Tursky and Lodge (1979), Saris et al (1980), or Saris (1987), show that respondents can answer in scales which suppose a higher level of measurement, for instance psychophysical scales. Likewise, in this research the mentioned subjective indicators of J S will be measured in psychophysical scales such as magnitude estimation and line production.

4.2 Second measurement problem: Variation in Response Function

In opinion research the opinions can not be directly measured. Since only answers to questions are available, it is implicitly assumed that respondents express their opinions in the same way in answers: that is the more similar the opinions the closer the answers given. This assumption, does not hold automatically. On the contrary, there are many reasons to question this assumption, as we will see below.

We borrow from Psychophysics the term Response Function (RF) to express the relationship between responses and opinions. The considered assumption is that the RF is the same for all respondents. Notice that the scores obtained for any item would be incomparable over persons if this assumption does not hold. If variation in RF (VRF) exists the responses can not be used as indicators for opinion.

In survey research most variables are measured by the answers to single questions. This makes it impossible to study the way a respondent expresses her/his opinion. In order to detect whether people use different RF when they give answers to questions Saris et al (1987) have carried out a series of experiments which show evidence of the existence of VRF. And they achieve four important conclusions that can be summarized as follow:

1. "If VRF was observed in these simple experiments it will exist anywhere"

2. "An important amount of variation in the responses which is normally interpreted as measurement error can be explained by individual response behaviour"

3. " Since VRF makes the answers of the respondents incomparable we can not trust any analysis done on data across respondents unless VRF is corrected or prevented"

4. " VRF is a problem independent of the answer modality used. Since categorical scales are equal distorted than psychophysical scales by VRF, and the former have many drawbacks magnitude estimation and lines production, will be the only response scales used in future experiments."

Therefore, before to built a definitive questionnaire to evaluate JS' determinants we will carry out an experiment to detect wether VRF exist in our data whether it can be prevented and what would be the consequences if we do not take into account this source of variability.

4.3 The data analysis problem

As has been mentioned in section 2.2 normally the variation in the overall life satisfaction is analysed by means of a regression model or a structural equation model which specifies a linear combination of the evaluations in the specific life domains.

Since the domains are considered more or less independent the analysis is limited to estimate additive linear main effects models, although some psychological research indicates that judgements are many times multiplicativily determined (See Minor et al, 1980).

Similar doubts concerning linear additivity of the models can be found among others in Mc. Kennell et al (1980) and Campbell (1981): "Even when domains seems to contribute additively to general life satisfaction a misfortune in one domain may as well have a corrector effect in the others".

In the experimental methodology we suggest to use, the effect of each factor on the general judgement of J S will incorporate the possibility of introducing interaction terms. Additionally, we will suggest in the last chapter other improvements for the data analysis and data collection derived from the application in opinion research of these techniques based on experimental designs.

5. THE DESIGN OF THE EXPERIMENT

The design of the experiment we are going to develop must consider the measurement and the analysis perspective we want to use. Hence it will be based on continuous data; it should be able to show that no VRF is included in the data; and finally it must specify a experimental design.

5.1 The chosen aspects of Job Satisfaction

Of course many different aspects could have been chosen to characterize a work situation. As was mentioned in section 2.2, we have chosen the following four aspects in this domain:

1. Job-Security; 2. Salary; 3. Atmosphere; 4. Education-agreement

For each of the four mentioned factors two levels have been specified, which represent the desirable and undesirable alternative. For instance the following specifications for job-security has been used:

(+) There is no chance to lose the job
(-) It is possible to lose the job

In the same way all 4 aspects are specified which leads to 2^4 (=16) combinations of these characteristics. All are represented in the first four columns of table 7. This display is called the design matrix.

Respondents are requested to evaluate how satisfied they would be with each of the 16 situations. The situations are presented in random order. As an example, the description of the second situation is given bellow:

(+) There is no chance to lose the job

(-) The work is not in agreement with the received education,

(-) The atmosphere in which the work is done is as a whole, quite disagreeable,

(-) Inadequate salary in comparison with any other which has the same level of job.

5.2 The different instructions

Saris et al (1987), in a series of psychophysics experiments in which the existence of VRF was detected, suggest the introduction of reference points through the questionnaire instructions. Since this procedure worked pretty well in a phychophysical experiment we are willing to use it to detect the existence of VRF in our JS data and if it is necessary to prevent it with the introduction of reference points. The general question asked is:

HOW SATISFIED WOULD YOU BE IN THE FOLLOWING SITUATION ?

Six different response instructions were formulated for answering the questions. First of all, the respondents are requested to express their opinion in lines or numbers. Another difference was the number of reference points provided on the scale. We have chosen to present no, one or two references points. The whole set of instructions used is given below:

The zero standard instructions

i) The closer a work situation is to your ideal the larger the number you should give. The more the situation deviates from your ideal the smaller the number you should give. Imagine the following situation:

(here one of the situations is mentioned)

Express your judgement in a number: the first situation is given an arbitrary value

ii) The closer a work situation is to your ideal the larger the line you should drav. The more the situation deviates from your ideal the smaller the line you should drav. Imagine the following situation :

(here one of the situations is mentioned)

Express your judgement in a line: the first situation is given an arbitrary line

The instruction with one standard

iii) The evaluation of a work situation which you think to be neither good nor bad is expressed by the number 500.
Evaluate nov the following situation:

(here one of the situations is presented)

The better the situation the larger the number you should give.

iv) The evaluation of a work situation which you think to be neither good nor bad is expressed by the following line:

____

Evaluate nov the following situation;

(here one of the situations is presented)

The better the situation the longer the line you should drav.

The instruction with two reference points

v) The best work situation which you can imagine is evaluated by the number : 1000

The worst work situation which you can imagine is evaluated by the number : 0

Evaluate now the following situation

(here a situation is mentioned)

The better the situation the larger the number

vi) The best work situation which you can imagine is evaluated by the following line :

____

The worst work situation which you can imagine is evaluated by the following line :

_

Evaluate now the following situation

(here a situation is mentioned)

the better the situation the longer the line

By means of three questionnaires six answers from each respondent are obtained using each of the six instructions for each of the sixteen situations. That means that 96 answers are collected per respondent.

Data were collected, using a computer assisted interview procedure recently developed (Saris and Pijper, 1986), in three waves. In each wave line and number responses for one instruction were asked. The order of the instructions has been varied over the respondents in such a way that they can be gathered into three groups. The scheme of the procedure is shown in Table 1. This variation was made to reduce the order effect.

RESPONDENTS WAVES	**1**	**2**	**3**	**N**
GROUP				
1	0	1	2	15
2	1	0	2	13
3	2	0	1	6
Total				34

Table 1. *The sequence in which the instructions with 0, 1 and 2 reference points are presented to three different groups of respondents.*

6. THE SAMPLE

The data have been collected from a convenience sample of 72 respondents from the Dutch population. However, they were reduced to the 34 respondents that completed all 96 items.

These respondents can not be seen as representing any population. However, there is enough variation in their background variables (See table2: sex, age, education, employement, income) for performing tests of measurement procedures (See Verwey et al, 1986). This panel has been in existence for more than one year and they are used to answering interviews by computer, including the use of line-production

Sex		%
Male	18	52.9
Female	16	47.1

(a) Sex (N=34)

Age		%
< 20 years	4	11.8
20 - 30	5	14.7
30 - 40	6	17.6
40 - 50	8	23.5
50 - 60	9	26.5
60 - 70	2	5.9

(b) Age (N=34)

Income, if employed		%
<1000	5	15.2
1000-1500	4	12.1
1500-2000	4	12.1
2000-2500	8	24.2
2500-3000	5	15.2
3000-3500	3	9.1
>3500	4	12.1

(c) Income, if respondent has a job (N=33 out of 72)

Education		%
primary school	2	5.9
lowervocational	4	11.8
secondary school	8	23.5
middlevocational	4	11.8
college	4	11.8
higher vocational	8	23.4
university	4	11.8

(d) Education (N=34.)

Table 2. *Descriptives of the background variables of our respondents (The figures are similar to those corresponding to the overall panel).*

The collected information leads to three dimensional data cube with situations (k), respondents (i), and instructions (j) as the dimensions.

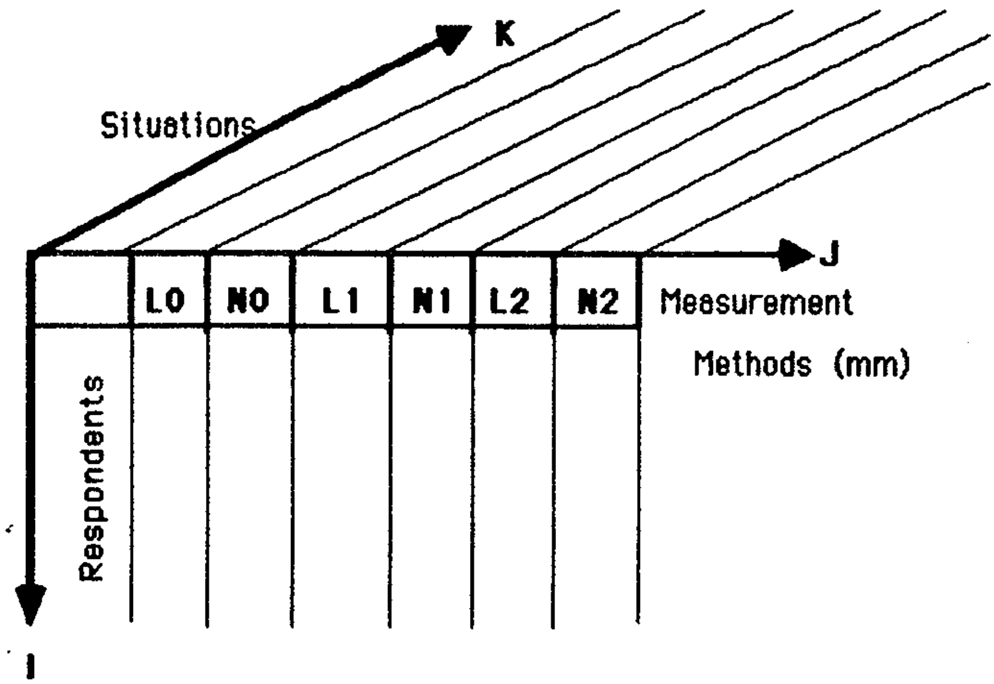

Fig 1. *The data matrix obtained by use of the above described experimental design, in which the answers subscript range, x_{ijk}, is determined as follows:*
Respondents (i = 1, ... , 34)
Situations (k = 1, ..., 16)
Measurement Methods or Instructions sets (j = L0,...,N2)

In the analysis one can choose to analyze for each respondent the data matrix with situations as cases and responses for the instructions as variables. Alternatively, one can analyse for each situation the data matrices with respondents as cases and the responses under different instructions as variables. The latter analysis is normally done. We will start with the former type of analysis and we will show afterwards what happens in the second type of analysis.

PART 2. MEASUREMENT ASPECTS

7. MEASUREMENT MODEL

In equation (1) the relationship between the responses for the j-instructions ($\mathbf{x}$) and the Opinion (ξ) is presented. Per each respondent six (j=1 to 6) RFs have to be specified. This leads to the following expression in matrix notation assuming linear relationships:

$$\underline{\mathbf{x}} = \underline{\alpha} + \Lambda\underline{\xi} + \underline{\delta} \qquad (1)$$

$$E(\underline{\xi}\,\underline{\delta}') = \underline{0} \qquad (2)$$

$$E(\underline{\delta}\,\underline{\delta}') = \underline{\Theta}_{\delta} \qquad (3)$$

$$E(\underline{\xi}\,\underline{\xi}') = \underline{\Phi} \qquad (4)$$

where $\underline{\mathbf{x}}$ (6x1) is a vector of the indicators

and $\underline{\alpha}$ (6x1) is a vector of the intercepts

$\underline{\Theta}_{\delta}$ (6x6) is a diagonal matrix with error variances

$\underline{\Phi}$ (3x3) is a symetric matrix of the correlations among factors

Λ (6x3) is the matrix of loadings, which is specified as:

$$\underline{\Lambda} = \begin{bmatrix} \lambda^{L}_{0} & 0 & 0 \\ \lambda^{N}_{0} & 0 & 0 \\ 0 & \lambda^{L}_{1} & 0 \\ 0 & \lambda^{N}_{1} & 0 \\ 0 & 0 & \lambda^{L}_{2} \\ 0 & 0 & \lambda^{N}_{2} \end{bmatrix} \qquad (5)$$

Figure 2 represents the same model in a path diagram

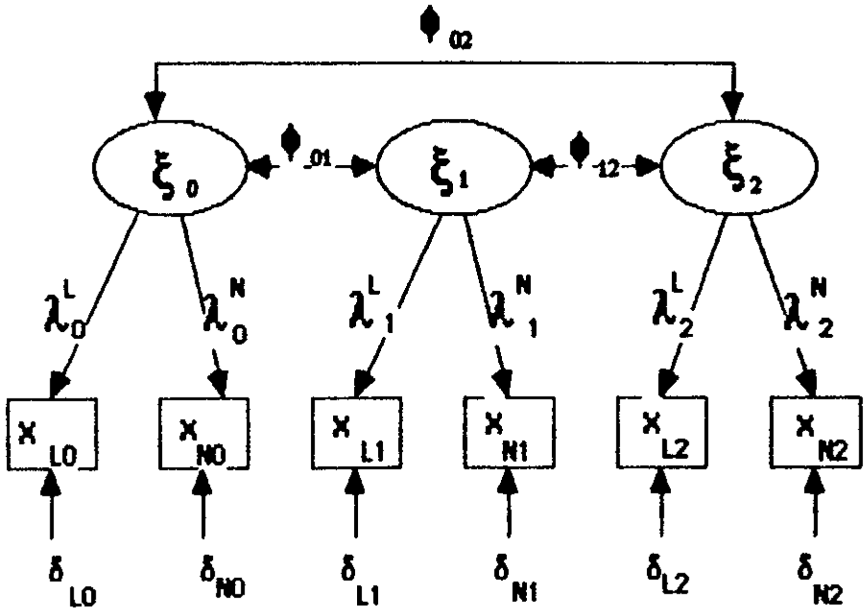

Fig. 2 Measurement model relating the unobserved opinion, ξ, and its two observable indicators, X.

In this case, unlike other experiments (See Saris et al, 1987) there are no metric stimuli available by which information about the RF could be obtained. Therefore the VRF has to be studied in the above specified measurement model.

We suppose that the RFs, characterized by the loadings and intercepts in each equation, will vary from person to person. But we hypothesize in line with Saris et al (1987) that the variation in coefficients is less in case of an instruction using two reference points than using one or zero reference points. This means that we expect less variation across persons in λ^L_2 and λ^N_2 than in λ^L_0, λ^N_0 and in λ^L_1, λ^N_1.

But before we can demonstrate this result we will first evaluate wether the variation in the instruction given affects validity and reliability of the different measures. The model of Fig.2 allows the evaluation of both validity and reliability of the proposed procedures.

Validity, means that the opinion is always the same although each respondent is requested to express her/his opinion in different ways. Reliability has to do with the correlation between the responses and the opinions across the measurement methods. With respect to the reliability we hypothesize that the used procedure has no effect on the reliability of the instruments. These hypotheses can be stated as:

$$H_0: \phi_{01} = \phi_{02} = \phi_{12} = 1.0 \qquad (6)$$

$$H_0: \lambda^L_0 = \ldots = \lambda^M_2 \qquad (7)$$

The first hypotheses suggests that the opinions measured are the same disregard the measurement procedure used. The second hypotheses suggests that for every instructions the reliability should be equally high. We start with the test of these hypotheses.

8. THE INDIVIDUAL DATA ANALYSIS

8.1 Validity and Reliability

Before showing the efficiency of the proposed method to reduce the VRF this experiment requires, as was mentioned in section 7, to asses the validity and reliability as well. In order to do so the model of figure 2 has been used. As data we used for each respondent the horizontal planes in figure 1 which represents the data matrix for a respondent. This means that the situations are the cases and the responses under the six different instructions the variables.

Table 3, displays a summary, computed over each of the 34 respondents, of the estimated parameters using the measurement model of figure 2, in which the variances of the latent variables were fixed to unity in order to avoid the identification problem of the model. The analysis was based on the correlation matrix and the EQS (Bentler, 1985) program was used for estimation.

PARAMETER	MEAN	MED	Q_1	Q_3	SD	IQR/Q_1+Q_3	C.V.
Φ_{01}	.949	.965	.92	.99	.048	.0366	.0506
Φ_{02}	.939	.955	.90	.99	.064	.0476	.0682
Φ_{12}	.927	.950	.88	.99	.071	.0588	.0766
λ_{L0}	.948	.970	.93	.99	.062	.0313	.0654
λ_{M0}	.953	.983	.93	.99	.057	.0313	.0598
λ_{L1}	.953	.962	.94	.99	.049	.0259	.0514
λ_{M1}	.956	.962	.93	.99	.038	.0313	.0397
λ_{L2}	.936	.952	.91	.97	.055	.0319	.0588
λ_{M2}	.958	.966	.94	.99	.045	.0259	.0470

Table 3. Mean; median; first and third quartiles; standard deviation; non parametric measure of dispersion; coefficient of variation across all respondents of the estimated parameters of the model in fig. 2.

The values of the correlations among the latent variables (Φ) are so close to the unity that we are willing to accept the hypothesis that every measurement method is measuring the same variable. (Appendix 1 discuss the statistical test of this hypotheses). The values of the loadings (λ) are all approximately the same. Therefore we are also willing to conclude that the reliability of each instrument is equally high.

In summary, the magnitude of Φ and λ coefficients in the first four columns of table 3 reveals that the respondents do not change their opinion due to the instructions and that they are very consistent which implies small error variance in each indicator. The spread presented in the last three columns displays no appreciable differences neither due to different number of reference points.

NOTE: Since the Anscombe's paper, 1973 it is well known that the correlation coefficient has serious drawbacks if graphics do not accompany this single numeric measure. Hence an exploratory data analysis by means of scattergrams of line and numeric responses has been carried out for each individual. From these visual displays we derived the following conclusions:

1. In our questionnaire the points tend to be spread towards the maximum in both answer modalities. This fact is more acute in the numeric responses due to the wider range. Since the correlation coefficient is not robust measure of linearity we have observed that its value is frequently higher than it should be according to the linearity displayed in the scattergram.

2. Due to the fact that some respondents have used category scale patterns when magnitude estimation is used as response modality the correlation coefficient tends to be to some extent artificially inflated -not only because of the linear pattern in the scattergram.

The first problem specially affects those responses given with no or one reference points, since in them outliers are more frequently observed (See table 4). This problem should be taked into account in the future since reliabilities migth be overestimated.

8.2 Reduction in the Variation in RF

Although the validity and reliability of the measurement instruments have been stated this does not exclude the possibility of VRF among respondents.

If the same analysis described above, done on the correlation matrix, were carried out on raw data, the estimated unstandardized coefficients, α and λ, from equation (1) in the measurement model could be quite different for the different respondents. These coefficients represent respectively the score the respondent would give to the "worst situation" and the specific slope used by each individual to express the opinion in the answers.

The estimate unstandardized slopes which were obtained by fixing to unity the variance of the latent variables are presented in table 4. This table portrays the entire stem and leaf distribution (Tuckey, 1977) of the unstandardized slopes. We have added the standard deviation (SD) and the coefficient of variaton (CV) of each distribution.

	N0	N1	N2	L0	L1	L2
0*	000001					
T	2222222233333			3		
F						
S				67	67	67777
.				888888	888888	8888888888
1*				0001111111	00000111	00000001111
T				222233	2222222333	222233
F				4	45	44
S		7	67	77	6	
.	99		999	88	88	
2*	1	011	011	01	0	
T		2222223	2333	3	22	
F	55	44555	44445			
S	6667	66667	66666667	7		
.	899	88899	8899			
3*	00	00	0			
T		222	223			
F						
S			6			
.						
4*						
T	2					
F						
S						
.						
5*		01				
T						
F		5				
SD	13.1	8.5	4.5	5.11	3.99	2.02
CV	1.03	0.30	0.18	0.41	0.32	0.20

Table 4. Stem and leaf distribution of the individual unstandardized slopes of RF (outliers not removed). Unit(N): 1=10, Unit(L): 1=1 Statistics (SD, CV) of the distributions' spread.

From this table ve can dravn the following conclusions:

1. There exist VRF in this substantive topic too. These conclusions holds for magnitude estimation data as vell as line production data

2. The use of an instruction with two reference points leads to less dispersion in the data than zero and one reference points do. This also holds for number and line responses.

3. It can even be shown that the hypothesis that the slope coefficients come from the same normal distribution, can not be rejected when two reference points are given using Kolmogorov-Smirnov (Lilliefors) test. The same hypothesis has to be rejected for the two other types of data using a .05 significance level. This holds for number and line responses.

4. This table also shows that the dispersion is considerably less if lines are used as response modality than when numbers are used. This is probably due to the fact that the respondents have used the size of the screen as an upper bound of the answer scale, as a consequence they were not willing to give answers beyond it, while such a reference point does not exist for numbers.

These results are completely in agreement with our expectations with respect to the effect of the introduction of two reference points. Only the last point is a minor adjustment of our expectations. This means that two reference points can prevent incomparability of responses which exist in the other procedures for a large part. This point can also be shown in a different way.

9. THE CONSEQUENCES FOR CROSS SECTIONAL DATA ANALYSIS

Now that we have shown that variation in response functions exist in these data and can be reduced by introduction of more reference points we also want to shown what happens in normal analysis across respondents.

Hence now we choose an analysis for each situation. So, we take in Figure 1 a vertical plane which mean that the respondents are the cases and the responses to the six different instructions are the variables.

In equation (1) the relationship between Responses and Opinions is given. However, we have seen that the relationship can not be represented by one response function for all respondents. In the present case, as was shown in the last section the RF across respondents is not constant and requires a formulation allowing for variation in the parameters for each respondent. Therefore a subscript for the i^{th} respondent has to be used in this model. This leads to:

$$x_i = \alpha_i + \Lambda_i \xi_i + \delta_i \qquad (8)$$

However, since this variability among respondents is usually not taken into account, cross sectional data are analyzed by means of measurement models like the one of Figure 2. The results of such analyses are given in table5

PARAMETERS	MEAN	MED	Q_1	Q_3	SD	IQR/Q_1+Q_3	C.V.
ϕ_{01}	.770	.804	.53	1.00	.220	.307	.29
ϕ_{02}	.584	.517	.42	.71	.242	.257	.41
ϕ_{12}	.480	.499	.31	.61	.202	.326	.42
λ^{L}_{0}	.785	.906	.55	1.00	.243	.290	.31
λ^{N}_{0}	.289	.268	.23	.35	.113	.207	.39
λ^{L}_{1}	.726	.722	.53	.96	.220	.289	.30
λ^{N}_{1}	.778	.756	.65	.87	.136	.145	.17
λ^{L}_{2}	.721	.706	.63	.77	.111	.100	.15
λ^{N}_{2}	.874	.955	.76	1.00	.162	.136	.18

Table 5. Summary of the estimated parameters from the measurement model of fig. 2 considered across individuals: N=34. (Analysis based on standardized data)

As a consequence the estimated lambda coefficients actually are "averaged respondent lambdas" which make no sense at all. So, the estimated values of the parameters summarized in table 5 are lower than we expected because of the high quality of the measurement intrument as it was stated in last section.

The results from table 5 contradict the achieved conclusions in previous section because they reveal in some extent invalidity and unreliability of the used indicators. First of all, although it was proved that individual opinion is kept independent from the process of introducing more reference points in the answer scale, the correlations among opinion factors are far from unity. This is a surprising result regarding that respondents are always the same. The explanation of this fact can not be other than VRF.

Unreliability problems are larger for the numeric answer modality and for the procedures using no reference points or one reference point. The estimates achieved under two standards are better approximations of the actual reliability of the indicators because of the limited variation in RF. These coefficients also are the ones which have greater lambda values as well as less dispersion which should be the case as we have seen before.

Notice that, in agreement with previous results, the line response modality does not show spectacular differences among reliability coefficients because of the introduction of reference points.

Normally, in using measurement models people think that low lambda coefficient means lack of reliability of the used indicators. However, our explanation for the low lambda and phi coefficients in table 5 is what Kimball (1957) and later Tukey called "the third kind of error", i.e. the right answer is given to the wrong model. In the present case a measurement model is used without taking into account VRF in cross sectional data.

A consequence of this error is that the estimated values of the standardized λ coefficient goes down, since they actually represent an "average respondents lambda" and consequently confounds the interpretation of the indicator's reliability. This fact can not be explained by lack of accuracy of the measurement instrument or respondents inconsistency across instructions as was clearly demonstrated in the individual analysis. This means that the "measurement error" is actually due to Variation in Response Function.

In order to show that the VRF indeed affects the lambda estimates we have chosen out of the 34 respondents those 18 who had answered in numbers modality in every wave on the same scale, from zero to thousand. Table 6 provides a similar summary of the results for this subsample that table 5 does for the whole sample. When no reference points are used the mean of lambda estimates is much larger than the ones in table 5. This means that when the metric used is approximately the same in the sample, reliabilities are going up however VRF, can still be observed in the last two columns of table 6, unless two reference points were introduced through the instructions. Otherwise the conclusions are the same as for table 5.

	MEAN	MED	Q1	Q3	SD	IQR/Q1+Q3
Φ_{01}	.537	.662	.31	.72	.297	.398
Φ_{02}	.734	.744	.65	.98	.280	.202
Φ_{12}	.343	.401	.10	.69	.372	.746
λ^L_0	.692	.688	.53	.84	.208	.226
λ^N_0	.645	.624	.46	.86	.241	.303
λ^L_1	.757	.887	.54	1.0	.289	.299
λ^N_1	713	.783	.63	.93	.296	.192
λ^L_2	.656	.656	.50	.81	.236	.237
λ^N_2	.844	.907	.71	1.0	.162	.170

Table 6. A summary of the estimated parameters of the measurement model of fig. 2 across all respondents (N=18) with a metric scale between 0 and 1000. (Analysis based on the correlation matrix)

10. DISCUSSION OF THE RESULTS OF THIS CHAPTER

In section 8 it has been stated that no matter which instruction is given, the respondent is very consistent in her/his evaluation over the different instructions. The last section showed that this consistency is not observable if respondents are not forced to use rather similar response functions, as was illustrated in table 5.

In section 8 we showed that VRF of respondents can be prevented by a large part when two reference points were used in the response scale. This instruction "forces" respondents to express their opinions in a rather similar way while leaving the variation in opinion unaffected. This means that this procedure can make the responses comparable as it did in the last case. This was confirmed in the previous section, in which two reference points supplied the highest reliabilities because of the more similar answer patterns across respondents.

We have also seen, along the analyses that line judgements are distorted to a less extent by the problem of VRF than numeric judgements.

The lines judgements have an additional advantage. Since people are not accustomed to use this way of answering, apriori response patterns such as categorical answers which were frequently observed in magnitude estimation (See NOTE in section 8.1) do not appear at all in line production judgements. This is another argument to recommend the use of line' lengths as response modality.

The suggested solution using two reference points is not always possible. The word "ideal" is not in general applicable, for instance, how should the two reference points be chosen in the question: How interested are you in politics?. Further research on this line is needed.

Variation in response functions is not just a scale problem, that can be reduced by the introduction of two standards. If the RF is linear VRF will be prevented by the introduction of two standards. But if the RF is non linear this is not sufficient. In this case the procedure can still be improved by fixing an additional neutral point in the middle of the scale or by using even more reference points. But it will be difficult to specify such points in a way that these points have the same positions on the scale for all respondents.

The conclusions achieved so far allow us to define what the methodology ought to be in order to realize the substantive objectives of our research:

First, the only responses we are going to use from now on are those obtained with instructions specifying two references points. Secondly, the 2^k factorial design will be analyzed, to estimate the effects of the four mentioned factors in JS. Chapter 3 will end with the discussion concerning the possibility to reduce the length of this kind of questionnaires by the use of fractional factorial designs (See Box, Hunter and Hunter, 1978).

PART 3. THE ANALYSIS ASPECT

11. THE JOB SATISFACTION EXPERIMENTAL DATA ANALYSIS

Since the experimental control of social phenomena is in general not possible the non experimental design has been the main methodology followed by Social sciences researchers. Due to the complexity of the studied phenomena, multivariate statistical techniques such as simultaneous equation systems, factor analysis models, structural equation models, etc., have been the ones almost exclusively used to evaluate the proposed theories to model sense of WB or perceived LQ (See among others: Burt et al (1979a); Burt et al (1979b); Mc Kennell and Andrews (1980); Andrews and Mc. Kennell (1980); Headey et al (1985)).

All real problems have idiosyncrasies that should be taken into account before applying mechanically any methodology. However every multivariate statistical technique of common use nowadays have been developed in a specific field, to solve a particular problem: In Psychology, Factor Analysis was developed to solve the problem of intelligence measurement, Analysis of variance arise to analyse data in agriculture, Cluster Analysis appeared to solve the taxonomy problem in Biology, Least Squares in astronomy, and so on. But, it is well known that these multivariate statistical tools although they came from different areas can be successful and generally be used when their particular requirements hold in other data.

As has been suggested in the first chapter we are going to apply to our J S data instead of multivariate statistical techniques based on single or multi-item scales experimental techniques of analysis (a set of situations which form a 2^k factorial design -See section 5.1) that have a long tradition in Physical sciences but also in Psychology.

We are aware that the process of gathering information from our panel through questionnaires, we carried out, can not be considered to be done in laboratory conditions of control, as used to be the case in proper experiments. However, in collecting our data we have selected two "levels" or "versions" for each variable (factor) and responses are achieved for all possible combinations of them. In this sense we use the term experimental factorial design.

On the other hand, one can critize that respondents are not pure "replicates" as normally they are assumed to be and we are going to use them. We agree with this objection, but in the exploratory data analysis of our data we have realized that in their answers respondents are sensitive indeed to a shift from low to high level in any factor, thus responses to situations (experimental conditions) could be pretty accurately ordered. In addition, the existence of consensus among our respondents was successfully tested as it is normally done (See Saris et al, 1987). Both results gave us confidence to follow the consequent analysis of our data.

11.1 Estimation of the four factors' effect on J S. The use of factorial designs at two levels

To define a general factorial design, after a number of levels of each factor has been selected, at least a response to each combination of these levels is required to carry out the estimation of the effects of the considered factors. However, if the relationship between response and factor are linear only data from two convenient points of the factor would be necessary to exactly determine the relation except for the error term. The simpler the relation the fewer the number of levels per variable would be necessary to collect data. As has been mentioned in section 5 here we will consider per each variable only two levels, which means that we assume linear relationship between general J S and each factor.

Before estimating the contribution of the Salary, Atmosphere, Education-agreement and job-security factors to general judgement it is necessary to remove the overestimation effect of the variance between individuals caused by the differences between optimists and pessimists, which respectively score high and low. This correction will not affect the estimated effects, only the variances (See Box et al, 1978). In order to do so we centered the data per individual, which means that from the 16 responses of each respondent we subtracted the mean, so that now indifference is equal to zero for all the respondents.

In Table 7 the average and the standard deviation for these centered data achieved with two standards are presented. This is our data matrix for the analysis, which summaries the values of our 16 experimental conditions (situations) measured using two reference points:

Situation	Level factors				Response			
	J	E	A	S	Mean (N2)	SD (N2)	Mean (L2)	SD (L2)
1	-	-	-	-	-347.7	76.2	-12.75	3.89
2	+	-	-	-	-171.8	122.4	-7.01	4.86
3	-	+	-	-	-240.9	79.0	-8.73	3.92
4	+	+	-	-	-43.6	112.2	-1.98	5.03
5	-	-	+	-	-176.2	106.5	-6.39	4.02
6	+	-	+	-	120.6	143.7	4.58	6.72
7	-	+	+	-	-24.7	138.1	-2.98	5.23
8	+	+	+	-	292.4	152.5	11.64	5.92
9	-	-	-	+	-228.9	61.6	-9.07	2.73
10	+	-	-	+	-.47.4	119.6	-1.01	6.21
11	-	+	-	+	-111.2	114.5	-5.69	5.00
12	+	+	-	+	118.6	134.8	3.52	2.56
13	-	-	+	+	-38.8	152.8	-0.79	6.80
14	+	-	+	+	307.3	105.9	12.27	4.22
15	-	+	+	+	109.0	170.0	4.46	7.64
16	+	+	+	+	547.4	111.6	22.63	3.54

Table 7. Design matrix: J *job-security;* E *educational background;* A *Atmosphere;* S *Salary. Data from a complete Job-satisfaction* 2^4 *factorial design*

For both centered responses (L2 and N2) the applicability condition of homoscedasticity was checked. To do so a plot of the means of the 34 respondents at the 16 situations (experimental conditions) versus the corresponding standard deviations suggests that the variability is not constant, as it can be seen in figure 3 it is higher in the middle and lower in the extremes. Efforts have been made to find a suitable transformation to stabilize the variance, but so far we have not succeeded. The commonly Box- Cox (1964) transformations do not help in this case (as an example, see the effect of log transformation in figure 4).

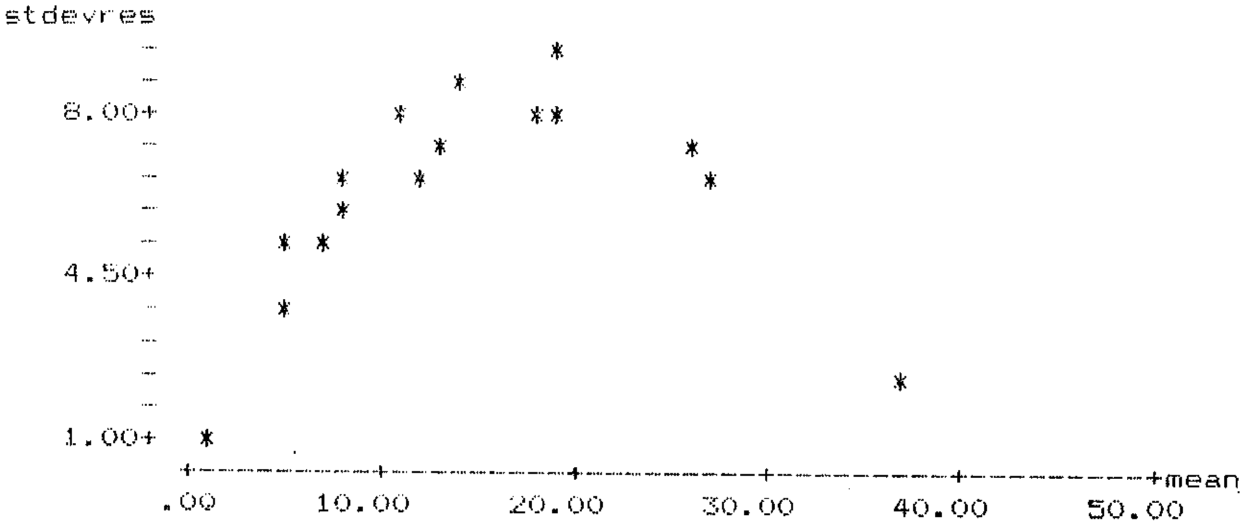

Fig.3 *Plot of the situations' mean versus their respective standard deviations (Answer modality: L2. Centered data)*

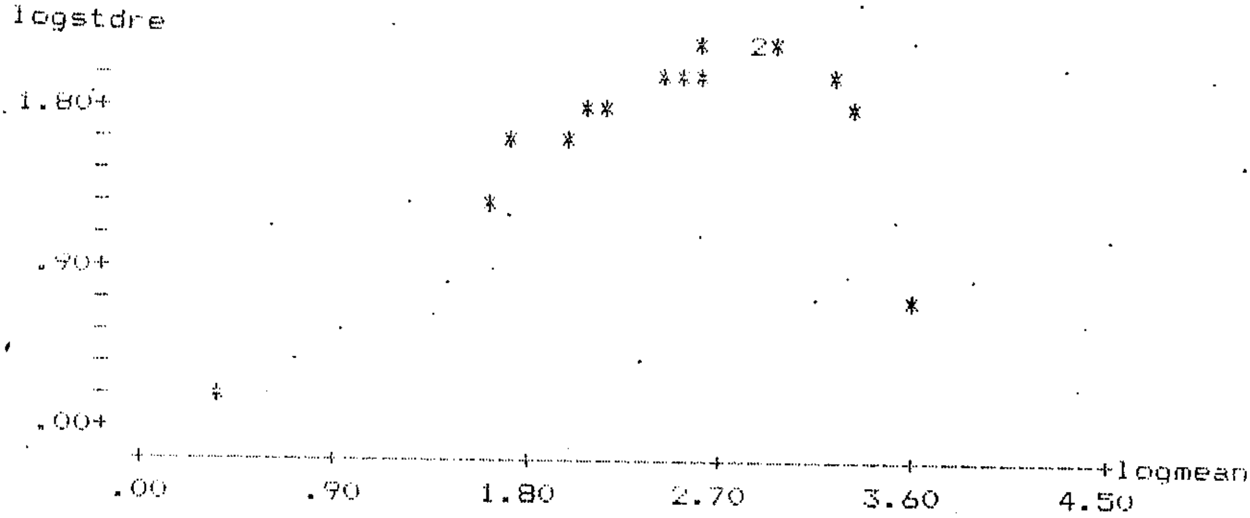

Fig.4 *Plot of the logarithm of the situations' mean versus the respective logarithm of their standard deviation (Answer modality: L2. Centered data)*

Regarding our data it seems logical to find this kind of heteroscedasticity, large dispersion in the middle of the range and lover in the extremes, since our data come from an instruction specifying the standards in the extremes of the answer scale. Because of these problems, Generalized Least Squares estimation method were used to calculate the effects since it corrects for heteroscedasticity.

With the gathered information it is possible to calculate the following 16 effects on J S: 4 main effects;6 two factors interactions;4 three factors interactions; 1 four factors interaction and the overall mean. Table 8 includes all estimates of these effects and their standard errors

	Est. Effect (N2)	Est. Effect (L2)	
Average	4.0	0.17	
J	272.9	10.82	*
E	153.7	5.38	*
A	276.2	11.02	*
S	156.0	6.24	*
JE	22.8	1.36	
JA	76.7	3.39	*
JS	26.1	1.30	
EA	24.1	1.14	
ES	14.2	0.50	
AS	22.2	1.69	
JEA	5.4	0.82	
JES	12.3	0.20	
JAS	16.5	0.11	
EAS	2.0	0.79	
JEAS	5.3	0.17	
SD	10.49	0.46	

Table 8. Analysis of the complete 2^4 factorial design: Estimates of effects. ( ) means significant effects.*

It is known that it is not possible to interpret the contribution of each factor on J S till the significance test has been done. This test can be carried out either by plotting the effects on normal probability plot (Daniel, 1976) or by considering the 34 respondents as "genuine" replicates of the experiment, hence assuming the combined variance among individuals as the noise of the system. Since we are not willing to accept the latter assumption significance was stated based on the first procedure which is illustrated in Figures 5 and 6.

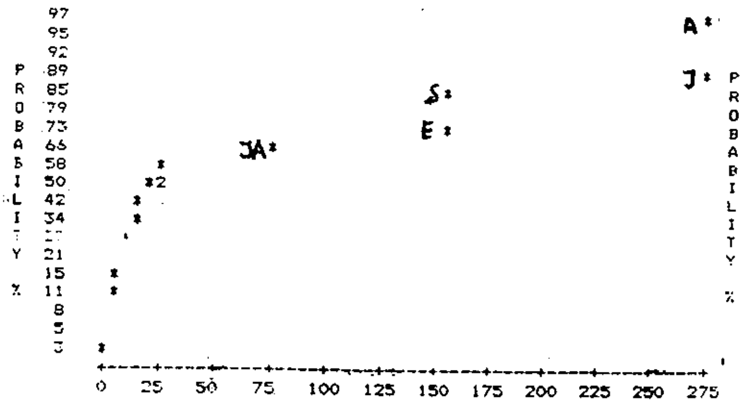

Fig. 5 Normal plot of the effects from complete 2^4 factorial design (magnitude estimation: N2)

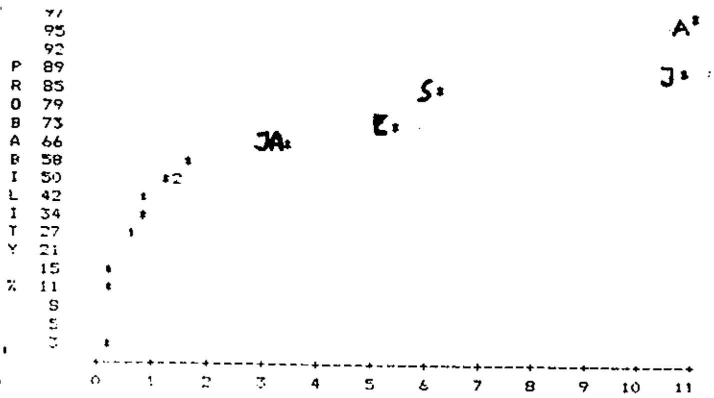

Fig. 6 Normal plot of the effects from complete 2^4 factorial design (lines production: L2)

11.2 Discussion of the results

From the above figures it is very clear that the four main effects, that is those from our four selected factors, are significant. Also the interaction between job security and atmosphere is significant. That means: that these effects are not easily explained by chance, since they are clearly distinguishable from the noise. Notice that the five mentioned effects are more than three SD (at the bottom of table 8) away from zero. Others are near to two SD but, we have already mentioned that this SD should be used carefully.

Before interpreting the main effect of each factor it is necessary to see whether these factors behave additively or whether they "interact". If true evidence of interaction effects are found they will be considered jointly with the main effect of the factor. In our case, as was mentioned in last paragraph, there exists an appreciable interaction between atmosphere and job-security, thus the effects of these two factors will be considered jointly.

Figure 7 illustrates the interpretation of this second order significant interaction: when the atmosphere is good the effect of job security on J S is larger than if the atmosphere is bad.

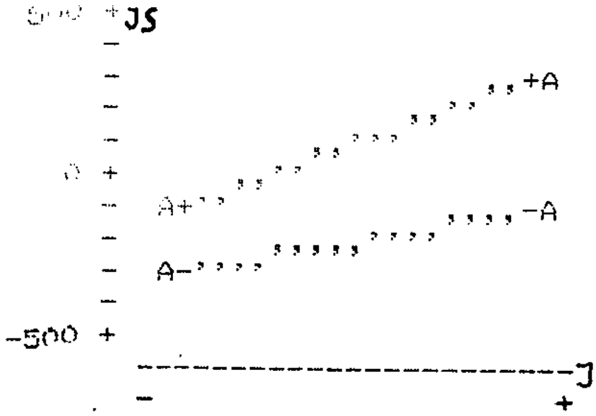

Fig. 7 The job-security- atmosphere interaction: plot of the job-satisfaction average (over levels of the other factors) at two levels of atmosphere and job-security (magnitude estimation: N2).

Since it is not possible to attach a physical meaning to the units used the interpretation of the effects is better done by comparison between their relative magnitudes. So, for these inhabitants of Holland Job security and atmosphere factors have the same effect and the effect is twice as large, considering the interaction effect as well, than the effect of the other two factors.

Finally, it can be seen that the other two factors are still significant, salary and education agreement, and have an equal contribution on the general judgement of J S. As it usually occurs all other higher order interaction are not significant.

12. THE UTILITY OF USING FRACTIONAL FACTORIAL DESIGNS IN THE QUESTIONNAIRES: THE LENGTH REDUCTION

Social and behavioural sciences data is mainly collected by interviewing people in a survey questionnaire. However in the set up of the questionnaire there exists a paramount problem, the length of the questionnaire. It is well known that the length of the questionnaire is related with the number of "missing values", with the number of incoherences in the responses, and respondents' fatigue which lead to diminish the reliability and validity of this measurement instrument.

However, when continuous measures are used the use of factorial designs can greatly simplify the analysis of the results as we have seen, we will show that in some cases also the design of the questionnaire to estimate the same number of effects can be simplified.

Two main problems arise when the proposed technique is used, and both appear when the number of factors (k) is large. The first one is that when k increases the number of questions (situations) increases geometrically, being 2^k (where 2 is the number of levels and k is the number of factors). However, this problem can be solved by the use of fractional factorial designs (See Box et al, 1978). This means that the number of questions can be drastically reduced if we can assume linearity and we are willing to reduce the number of effects that are of interest in the study.

Since it is very unusual to find with four factors, significant three or higher order interactions we can disregard them and concentrate the efforts on estimating the main effects and second order interactions. In fact there tends to be a certain hierarchy: main effects tend to be larger than two-factor interaction effect, which in turn tend to be larger than three-factor interactions and so on (See Box et al, 1978).

We are not going to discuss how to deal with fractional factorial designs as this can be found in the literature. In fact what we would like to show is that we could have used this technique and conduct a half-fraction of 2^4 factorial design which only would need $2^3 = 8$ questions to estimate the same effects, and our conclusions would have been very much the same, as is illustrated in Table 8, just for lines production responses

Situation	Level factors J	E	A	S	Response
1	-	-	-	-	-12.75
10	+	-	-	+	-1.01
11	-	+	-	+	-5.62
4	+	+	-	-	-1.98
13	-	-	+	+	-0.79
6	+	-	+	-	4.59
7	-	+	+	-	-2.98
16	+	+	+	+	22.63

Table 9. *Analysis of a half-fraction of the full 2^4 design: 2^{4-1} fractional factorial desing (Answer modality: L2)*

The design used is named a 2^{4-1}_{vi}, where 2^{4-1} indicates that 1/2 fraction of 2^4 complete factorial design has been used and vi is named the "design resolution" and indicates that in the estimated effects the main effects will be confounded with three-factors interactions and two-factors interactions will be confounded between them.

From this data, it is clear (Box et al, 1978) that the effects specified in table 10 can be estimated

	Est. Effect (L)	Confounding pattern
Average	.25	(Average + JEAS)
J	11.61 *	(J + EAS)
E	5.48 *	(E + JAS)
A	11.22 *	(A + JES)
S	7.06 *	(S + JEA)
JE	3.05	(JE + AS)
JA	3.89 *	(JA + ES)
JS	2.44	(JS + EA)

Table 10. Analysis of a half-fraction of the 2^{4-1} fractional factorial design: Estimates of effects (*) Means significant effects (Response modality: L2)

For example the value of 11.61 of J + EAS should be interpreted, as the estimated effect of job-security (J) and the three factor-interaction (EAS) together.

Notice that the only disadvantage with respect to the complete 2^4 factorial design is that the two-factors interaction JA is confounded with the two-factors interaction ES. But the other conclusion would be much the same as the ones we have drawn from the complete 2^4 factorial design in table 8.

The second problem we have pointed at is that if the number of factors, k, increases also the size of the question describing the situation does increase (See Verwey et al, 1986). This problem has no easy solution, and perhaps the best thing to do is to restrict its use to situations when k is lower than 6 or 7 unless visual or other types of aids can be used to clarify the way to present the situations in the questionnaire. In general respondent can not take as many factors into account any way.

13. CONCLUSIONS

This study has led to several conclusions. These are on the one hand Methodological and on the other substantive coming from the J S data analysis. Based on our analysis we will emphasize the following ones:

Our results in chapter 2 suggested that when it is possible data should be collected in opinion research by psychophysical scales. We have proposed lines production as the better response modality (see section 10).

Variation in Response Function makes the answers of the respondents incomparable, therefore it should be prevented in any cross sectional data collection by the introduction of at least two reference points in the response scale (see section 10)

Experimental factorial designs at two levels are easily implemented in the questionnaires and have at least two obvious advantages: interactions can be included in the models to explain overall judgements (see section 11); and likely if the number of predictor variables is not large the use of fractional factorial designs will allow to build shorter questionnaires (see section 12). Specific software is also available (JASS, 1985).

In this article we have only mentioned factorial designs at two levels but, if it is reasonable to assume that the response relation is quadratic, designs at three levels could be used (See section 11.1). In general all the techniques of experimental design can be applied, including such new ideas as the use of this technique to study variability (Taguchi,1980; Box and Meyer, 1984; Tort-Martorell, 1986).

The substantive analysis of J S data contains problems which are difficult to avoid. First, because of the size of the sample used (N=34) we can not elicit inferential conclusions for the Dutch population. However, this research might be seen as a pilot study to point out different lines to follow in future research.

Unlike previous studies (Kallenberg, 1977; Saris, 1982; Hodson, 1986) our results signal that nowadays job-security and atmosphere are much more important than salary or education agreement in explaining J S . This is not a surprising result since unemployement has become a major problem in recent years. A significant interaction between job-security and atmosphere has been detected as well.

Confirmatory studies of these findings in the Dutch population should follow this one. Comparative studies as well with other countries where for instances, salary is smaller, would be necessary to increase our understanding of the relative merits of the determinants of J S.

Secondly, because of the experimental framework of our research,not done in previous research, it is hard to carry out other comparisons between our results and results from others.

The main differences arise as a consequence of the data collection design. What other researchers (See every reference in this line from Andrews, 1974 to Hodsson, 1986) have been asking to the respondents is: How satisfied are you with your work? and sometimes other indirect measures such as whether the worker would recommend the job to a friend. Besides, other variables of the respondents are also measured and then used as predictors of J S in a statistical model which relates both sets of measures.

In our research instead of collecting information on J S and on their determinants separately, respondents were asked to evaluate themselves J S of the situations defined by the explanatory variables. Hence, comparisons to the results coming from these two different data collection processes would be somewhat risky.

In addition to the mentioned aspects, we have already emphasized those differences among the J S' predictors used by different researchers. Since the factors or predictors used are not independent the statement "what goes in is what comes out" holds also for our models. Their relative effects on J S will depend in a great extent on the factors considered as explanatory variables.

In this way, future research should include altogether with our four factors the age, that we think with Kalleberg and Loscocco, 1983, can explain large differences in J S, as well as the two "extrinsic dimensions" already used by Kallenberg, 1977 "chances of promotion and resources adequacy". Since the age is measured apart of the data collection design it does not increase our questionnaire, and these last two factors can be easily added to our questions (situations). Moreover, by using fractional factorial designs the length of the corresponding questionnaire would be reduced in this case (of six factors) to 32 questions.

APPENDIX

We have tried to verify the following hypotheses in the model of fig.2

$$H_0: \Phi_{01} = \Phi_{02} = \Phi_{12} = 1.0$$
$$H_0: \lambda^L_0 = \ldots = \lambda^N_2$$

Since for an excessive number of respondents the above model was rejected using the .05 level test (out of them those with higher lambdas values), the relationship between the power of the test and the lambda values was checked (Satorra and Saris, 1985). As an alternative hypothesis we specified a model with a deviation of .05 from unity for two correlations between latent variables:

$$H_0: \Phi_{12} = 1.0$$
$$H_1: \Phi_{01} = \Phi_{02} = .95$$

The λ coefficients varied from .75 to .99. The results of the sensitivety of the .05 level likelihood ratio test with respect to the power has been reported in table 11.

$\lambda^L_0 = \ldots = \lambda^N_2 = \lambda$	NON CENTR. PAR.	POWER
0.75	0.145	0.065
0.80	0.255	0.065
0.85	0.488	0.090
0.90	1.093	0.120
0.95	3.376	0.280
0.97	6.508	0.620
0.99	17.366	0.970

Table 11. Power of the test for different values of λ in fig.2

In this table we can see that when a misspecification of .05 in the correlations is introduced the increase "exponential" power with the lambda value will lead to rejection of too many individual models, particularly above .95. For example, 62% of the models with lambda coefficients around .97 would be rejected due to a misspecification of .05 in the correlations among factors. Since we think that this test is too powerful for the data we have and the correlations do not differ much more than .05 from 1 in general we have decided not to reject the hypothesis that the correlations among the factors are close to one.

REFERENCES

Abbey A., Andrews F.M., 1985
Modelling the psychological determinants of Life Quality
Social Indicators Research, 14, n1, pp. 1-34

Andrews F.M., Mc. Kennell A.C., 1980
Measures of self-reported well-being: Their affective, cognitive and other components.
Social Indicators Research 8, pp. 127-155

Andrews F.M., Withey S.B., 1974
Developing measures of perceived life quality: Results from several surveys
Social Indicators Research, 1, pp 1-26

Andrews F.M., Withey S.B., 1976
Social indicators of Well-Being: Americans' perceptions of Life Quality
Plenum Press, New York

Anscombe F.J., 1973
Graphs in statistical analysis
American Statistician, 27, pp 17-21

Bentler P.M., 1985
EQS, A structural equation program
BMDP Statistical Software, Inc.

Box G.E.P., Cox D.R.,1964
An analysis of transformations
J. Roy. Stat. Soc. Series B, 26, pp. 211

Box G.E.P., Hunter W.G., Hunter J.S., 1978
Statistics for experimenters: An introduction to design, Data analysis and model Building.
Wiley and sons

Box G.E.P., Meyer R.D., 1984
Analyzing two level fractional factorial experiments for possible dispersion effects.
University of Wisconsin M.R.C. Tech. Report n 2746

Bradburn N.M., 1969
The structure of psychological well-being
Aldine, Chicago

Burt R.S., Wiley J.A., Minor M.J., Murray J.R., 1979a
Structure of well-being: form, content and stability over time
Sociological methods and Research 6, n3, pp. 365-407

Burt R.S., Fischer M.G., Christman K.P., 1979b
Structure of well-being. Sufficient Conditions for identification as restricted covariance models.
Sociological methods and Research 8, n1, pp. 111-120

Campbell A., 1976
Subjective measures of well-being
American Psychologist 31, pp. 117-124

Campbell A., 1981
"The sense of well being in America". Recent patterns and trends
Mc. Graw-Hill

Daniel C., 1976
Applications of Statistics to Industrial Experimentation
Wiley

Diener E., 1984
Subjective Well-Being
Psychological Bulletin, 95, pp. 542-575

Doorn L., Saris W.E., Lodge M., 1983
Discrete or continous measurement: what difference does it make?
Kuantitatieve Methoden, 10.

Headey B., Holmstrom E., Wearing A., 1985
Models of well-being and ill-being
Social Indicators Research 17, pp. 211-234

Hodson R., 1986
Corporate structure and job satisfaction: A focus on employer characteristics
Sociology and Social Research, 70, pp.113-133

Hortley J., Little B.R., 1985
Affective and Cognitive components of global subjective well-being measures
Social Indicators Research 17, pp. 139-197

JASS, 1985
Software of Analysis of Experimental Design
Joiner Associates.

Kalleberg A.L., 1977
Work values and job rewards: A theory of job satisfaction
American Sociological Review, 42, pp. 124-143

Kalleberg A.L., Loscocco K.A., 1983
Aging, values, and rewards: Explaining age differences in job satisfaction
American Sociological Review, 48, pp. 78-90

Kimball , 1957
Errors of the third kind in statistical consulting
J.A.S.A., 57, pp.133

Krantz D.H. et al., 1974
Contemporary developments in Mathematical Psychology
Freeman

Larsen R.J., Diener ED., Emmons R.A., 1985
An evaluation of subjective Well-Being measures
Social Indicators Research, 17, pp 1-17

Levy S., Guttman L., 1975
On the multivariate structure of wellbeing
Social Indicators Research 2, pp. 361-388.

Lodge M., Tursky B.,1979
Comparison between Category and Magnitude scaling of Political Opinion employing SRC/CPS items
In American Political Sciences Review, 73, pp. 50-66

Mc. Kennell A.C., 1978
Cognition and affect in perceptions of Well-Being
Social Indicators Research, 5, pp 389-426

Mc. Kennell A.C., Andrews F.M., 1980
Models of cognition and affect in perceptions of Well-Being
Social Indicators Research, 8, pp 257-298

Mc. Kennell A.C., Andrews F.M., 1983
Components of perceived Life Quality
Journal of Community Psychology, 11, n2, pp 98-110

Mc. Kennell A.C., Atkinson T., Andrews F.M., 1980
Structural constancies in surveys of perceived Well being
In Szalai and Andrews (Eds.): Quality of Life: Comparative studies
Sage Pub.

Miller J., 1980
Individual and occupational determinants of job satisfaction
Sociology of Work and Occupations, 7, pp. 337-366

Minor M.J., Bradburn N.M., Schaeffer N.C., 1980
The structure of Life-Satisfaction: A comparative analysis across social groups
In Szalai and Andrews (Eds.): Quality of Life: Comparative studies
Sage Pub.

Organ D.W., Near J.P., 1985
Cognition vs affect in measures of job satisfaction
International Journal of Psychology, 20, pp 241-253

Quinn R.P., Mangione T.W., 1973
Evaluating weighted models of measuring job satisfaction: a Cinderella story.
Organizational Behaviour and Human performance 10, pp. 1-23

Saris W.E., 1982
Linear Structural Relationships: Measurement Models (Vol. 2)
Sociometric Research Foundation

Saris W.E. (Ed.), (press)
Variation in Response Behaviour: a source of measurement error
Sociometric Research Foundation

Saris W.E., Neijens P., Doorn L.V., 1980
Scaling Social Sciences variables by multimodality matching
M.D.N., 2, pp 3-22

Saris W.E., Pijper M., 1986
Computer assisted interviewing using home computers
Elsevier, European Research, 4, n3, pp. 144-151

Satorra A, Saris W.E., 1985
Power of the Likelihood Ratio Test in Covariance Structure Analysis
Psychometrika, 50, n1, pp. 83-90

Smith P.C., Kendall L.M., Hulin C.L., 1969
The measurement of satisfaction in work and retirement: A strategy for the study of attitudes
Rand Mc. Nally and Cia.

Stevens S.S.,1975
Psychophysics: Introduction to its Perceptual Neural and Social Prospect
Wiley and sons

Szalai A., Andrews F.M. (Eds.), 1980
The Quality of Life: Comparatives studies
Sage pub.

Taguchi G., Wu Y., 1980
Introduction to Off-Line Quality Control
Central Japan Quality Control Association

Tort-Martorell X., 1986
Disenos factoriales fraccionales, aplicaciòn al control de calidad mediante el diseno de productos y procesos.
Tesis doctoral (Univ. Politècnica de Catalunya)

Tukey J.W., 1977
Exploratory Data Analysis
Addison and Wesley

Verweij M.J., Saris W.E., Bonsel G., et al, (1986)
A measurement procedure for the evaluation of health states by the general public
Sociometric Research Foundation, R.M. 861016

Wish N.B., 1986
Are we really measuring the Quality of Life? Well-being has subjective dimensions as well as objective ones.
American Journal of Economics and Sociology, 45, n1,pp. 93- 99

Zajonc R.B., 1980
Feeling and Thinking: preferences need no inferences
American Psychologist, 35, pp 151-175

tr. propi

JOAN M. BATISTA-FOGUET, WILLEM E. SARIS, AND
XAVIER TORT-MARTORELL

DESIGN OF EXPERIMENTAL STUDIES FOR MEASUREMENT AND EVALUATION OF THE DETERMINANTS OF JOB SATISFACTION

560

(received April 30, 1988)

ABSTRACT. In this paper the authors examine the methodology of data analysis and collection usually applied in Job-Satisfaction studies and more generally in Well-being research. In this field, data is collected through questionnaires and analysed using multivariate statistical techniques, a process which is commonly known as non-experimental research. The article points out some of the problems in this procedure and proposes an alternative experimental methodology which is applied to evaluate Job-Satisfaction over a Dutch random sample. Essentially it consists of the implementation of factorial designs in questionnaire construction. The use of fractional factorials is also discussed.

1. INTRODUCTION

Since the Andrews and Withey (1974) paper "Developing measures of perceived life quality: results from several surveys", which in fact represents a landmark in the measurement and assessment of perceived life-quality, an enormous number of papers has appeared trying to cover more or less the same objectives.

According to Diener (1984), psychological well-being has become a major focus of applied and theoretical research, with over 700 studies published during the past 25 years. In recent decades substantial progress has been make in finding effective ways of measuring people's sense of well-being (WB), in learning how measures of various aspects -domains- of life quality (LQ) interrelate, and in evaluating how they contribute to the overall sense of WB. Interest in this topic was so widespread that political behavioural and social scientists published a specialized journal, called Social Indicators Research, to deal with it.

On the one hand, what can be seen in the literature is a proliferation of measures of perceived WB which normally provide 3, 5, 7, 9, or 11

Social Indicators Research **22**: 49–67, 1990.

point answer scales (Abbey and Andrews 1985; Larsen *et al.*, 1985), that is, data collected in categorical scales which afterwards are used for the analysis as continuous data.

On the other hand, for data analysis in this field exclusive use is made of multivariate statistical techniques such as Factor Analysis (Smith *et al.*, 1969; Kalleberg, 1977; Hodson, 1986) Principal Components analysis and Cluster Analysis (Andrews and Withey, 1976) or Smallest Space Analysis (Levy and Guttman, 1975).

These techniques have helped researchers first to identify domains from a long list of items (as independent concepts in the factor space) and then to use them as predictors to evaluate their contribution to the overall sense of WB, by means of structural equation models or regression techniques (Burt *et al.*, 1979; Campbell, 1981; Horley and Little, 1985; Hodson, 1986). That is, in this topic non experimental statistical analysis techniques are exclusively used to analyze data. In these studies the WB of the respondents was explained by other characteristics of themselves such as their work situation, age, income, etc.

In this paper we want to present a different approach. The respondents are confronted with a large number of situations and they have to indicate how satisfied they would be in these situations. The presented situations are under control of the designer of the questionnaire: one can specify the factors which are taken into account in the study. Such an experimental approach allows a different type of analysis which has advantages and disadvantages compared with the commonly used procedures. Below we want to illustrate this approach by an example and discuss the merits of the approach at the end of the paper.

2. EXPERIMENTAL RESEARCH OF JOB SATISFACTION

Out of the possible domains of WB we will chose the job-domain and four usual aspects (values): job-security, salary, atmosphere, and education-relatedness, are considered as explanatory variables.

Since work itself constitutes a domain of the perceived LQ that is wholly subjective, it should be also measured by asking people their own opinion in assessing different job situations. Therefore the questionnaire used will constitute a set of subjective indicators to cover this domain (See the arguments of Wish 1986, to consider also objective indicators in measuring LQ). This questionnaire is described in section 2.3.

In addition, as is well known from McKennell's paper (1978), which clarified the structure of the perceived global WB model by dividing it into cognitive and affective components, this distinction has become a regular feature in social and psychological studies (Andrews and McKennell, 1980; McKennell and Andrews, 1980). More recent developments indicate that the cognitive and affective components may be independent of each other Zajonc (1980).

Due to this fact, it is necessary to stress that our study will be restricted to the measurement of "satisfaction", job-satisfaction (JS), which basically reflects cognitive evaluation rather than affective states concerning the other dimension of "happiness".

Although we accept the arguments put forward by Quinn and Mangione (1973) or Organ and Near (1985) minimizing the importance attached to this cognitive component, it is not the aim of this paper to concern itself with that discussion.

Many sociological studies of JS usually concentrate on issues such as work efficiency and productivity. They explain JS by occupational status, specialization of tasks, styles of supervision, opportunities for improvement or corporate structure rather than along the subjective line referred to in the previous paragraph (Wannous and Lawler, 1972; Miller, 1980; or Hodson, 1986).

Like many other life quality researchers we are involved in searching for a set of measures of perceived JS that could be used as social indicators, but before centering our attention on "What" JS explains we will discuss "How" it should be measured.

2.1 *Categorical versus Continuous Data*

As we have mentioned above most studies use category scales to measure the subjective feelings. This leads to ordinal measurement level of these variables. However, interval or ratio scales are normally required to fulfill the assumptions of the statistical techniques commonly used in this field for data analysis. Obviously, this assumption is violated when categorical scales are used to collect data.

Moreover, a frequent focus of interest in perceived WB, which concerns comparative studies (Szalai and Andrews 1980) is the search for constancies or/and differences among countries. A problem many authors found in the data used to carry out these comparisons, was that no standard category scale was available (McKennell *et al.*, 1980).

In psychophysics it is well known that the respondents are able to supply information on a continuous scale. Krantz *et al.* (1972) showed that the information is stored by the individuals in a continuous scale, and there is no reason to believe that the respondents can only express their opinions in a categorical scale.

Stevens (1975), Saris *et al.* (1977), Tursky and Lodge (1979), Saris *et al.* (1980), Doorn *et al.* (1983) or Saris (1988), show that respondents can provide information on (log) interval level of measurement using continuous scales. Likewise, in our study the mentioned subjective indicators of JS will be measured by means of matching procedures (Stevens, 1971) such as magnitude estimation and line production.

2.2 *Experimental versus Non-Experimental Designs*

As has been mentioned normally the variation in overall life satisfaction is analysed by means of a regression model or a structural equation model which specifies a linear combination of the evaluations in the specific life domains.

These multivariate statistical techniques are specially appropriate for non experimental research, that is, either the data come from objective indicators (Fernández and Kulik, 1981) or come from the items concerning subjective indicators of a classical questionnaire (Campbell, 1981; Larsen *et al.*, 1985). However, if the researcher himself writes the items of the questionnaire to define subjective indicators there is no reason not to go a step further in order to introduce an experimental design for the data collection. A pretty similar approach has also been done in other fields (Bock, 1960; Bock and Bargman, 1966; Cronbach *et al.*, 1972; Mallenberg, 1979; Hox *et al.*, 1988; Louviere, 1988).

Since the domains are considered more or less independent the analysis is limited to estimate additive linear main effects models, although psychological research indicates that judgments are quite often multiplicativily determined (See Minor *et al.*, 1980; McKennell *et al.*, 1980; Campbell, 1981).

In the proposed experimental methodology the effect of each factor on the general judgment of JS can be studied but also interactions between the factors can be taken into account. In addition, we will suggest in the last section other improvements for the data analysis and data collection derived from the application in opinion research of these techniques based on experimental designs.

2.3 *Design of the Questionnaire*

The measurement instrument we are going to develop must consider the measurement and the analysis perspective we want to tackle. Hence it should provide continuous data and it must specify an experimental design to collect this data.

Of course many different aspects could have been chosen to characterize a work situation. As was mentioned, we have chosen the following four aspects: 1, Job-Security; 2, Salary; 3, Atmosphere; 4, Education-relatedness.

After the factors have been chosen, a decision has to be made about the number of levels. The simpler the relation the fewer the number of levels per factor would be necessary to collect data. Since we assume that the relationship between general JS and each factor is linear, only data from two convenient levels of the factor would be necessary to determine the relation exactly, except for the error term. Therefore we have chosen only two levels for each factor. For instance the following specifications for job-security have been used:

(+) There is no chance of losing the job
(−) It is possible to lose the job

In the same way all 4 aspects are specified which leads to 2^4 (= 16) possible combinations of these characteristics. All are represented in the first four columns of Table I. This display is called the design matrix.

Respondents are requested to evaluate how satisfied they would be with each of the 16 situations. The situations are presented in random order. As an example, the description of the second situation is given below:

(+) There is no chance of losing the job
(−) The work is not related to previous education,
(−) The working atmosphere is, as a whole, quite unpleasant,
(−) Inadequate salary in comparison with same position elsewhere.

The general question asked is:

HOW SATISFIED WOULD YOU BE IN THE FOLLOWING SITUATION?

and two different response instructions were formulated for answering the questions, so, respondents are requested to express their opinion in lines and numbers. Batista-Foguet and Saris (1988), as a conclusion to a series of psychophysical experiments (Saris 1988), have shown that the best way to collect information in opinion research is by the introduction of two reference points to fix the extremes of the continuous answer scale. In this study we followed this suggestion and used two psychophysical scaling procedures.

The first procedure was as follows:

The best work situation which you can imagine is evaluated by the number: 1000
The worst work situation which you can imagine is evaluated by the number: 0
Now evaluate the following situation
(here a situation is mentioned)
The better the situation the larger the number

The second procedure was:

The best work situation which you can imagine is evaluated by the following line:

____

The worst work situation which you can imagine is evaluated by the following line:

_

Now evaluate the following situation
(here a situation is mentioned)
the better the situation the longer the line

By means of this questionnaire two answers from each respondent are obtained for each of the sixteen situations. That means that 32 answers are collected per respondent. These data were collected, using a computer assisted interview procedure (Saris and Pijper, 1986).

The data have been collected from a convenience sample of 34 respondents that completed all 32 items. This panel has been in existence for more than a year and they are used for answering interviews by computer, including the use of line-production. This sample does not represent any population but the random selection of the respondents guarantees sufficient variation in background variables for methodological research. (See Verwey *et al.*, 1986).

3. DATA ANALYSIS

In this paper the data are collected based on a factorial design. Therefore analysis of variance can be used for analysis of these data. One can make the criticism that respondents are not pure "replicates" as they are normally assumed to be and we are going to use them. We agree with this objection, but in the exploratory analysis of our data we have realized that in their answers respondents are sensitive indeed to a shift from a low to a high level in any factor. Thus responses to situations (experimental conditions) could be pretty accurately ordered. In addition, the existence of consensus among our respondents was successfully tested as it is normally done (Batista-Foguet and Saris, 1988). Both results gave us confidence to continue with the consequent analysis of our data.

3.1 *Estimation of the four factors' effect on JS. The use of factorial designs at two levels*

Table I summarizes the data using the average and the standard deviation for each combination of these levels, 16 experimental conditions (Vignettes, in Rossi and Anderson, 1982). The data are centered, which means that from the 16 responses of each person we subtracted the mean, so that now indifference is equal to zero for all the respondents. By doing this the overestimation effect of the variance between individuals caused by the differences between optimists and pessimists (which respectively score too high and too low) is removed. This correction will not affect the estimated effects of the JS' factors, only the variances.

For both centered responses (L and N) the assumption of homoscedasticity was checked. To do so a plot of the means of the 34 respond-

TABLE 1
Design matrix

Level factors (Sit.) J E A S		Response Mean (N)	SD (N)	Mean (L)	SD (L)
1	− − − −	−347.7	76.2	−12.75	1.18
2	+ − − −	−171.8	122.4	−7.01	5.52
3	− + − −	−240.9	79.0	−8.73	5.04
4	+ + − −	−43.6	112.2	−1.98	6.75
5	− − + −	−176.2	106.5	−6.39	5.56
6	+ − + −	120.6	143.7	4.58	8.25
7	− + + −	−24.7	138.1	−2.98	7.81
8	+ + + −	292.4	152.5	11.64	7.60
9	− − − +	−228.9	61.6	−9.07	3.48
10	+ − − +	−47.4	119.6	−1.01	7.58
11	− + − +	−111.2	114.5	−5.69	6.42
12	+ + − +	118.6	134.8	3.52	8.07
13	− − + +	−38.8	152.8	−0.79	8.87
14	+ − + +	307.3	105.9	12.27	6.43
15	− + + +	109.0	170.0	4.46	9.41
16	+ + + +	547.4	111.6	22.63	2.20

J job security; E educational background; A Atmosphere; S Salary. Data from a complete Job-satisfaction 2^4 factorial design (Centered data; Sample size = 34).

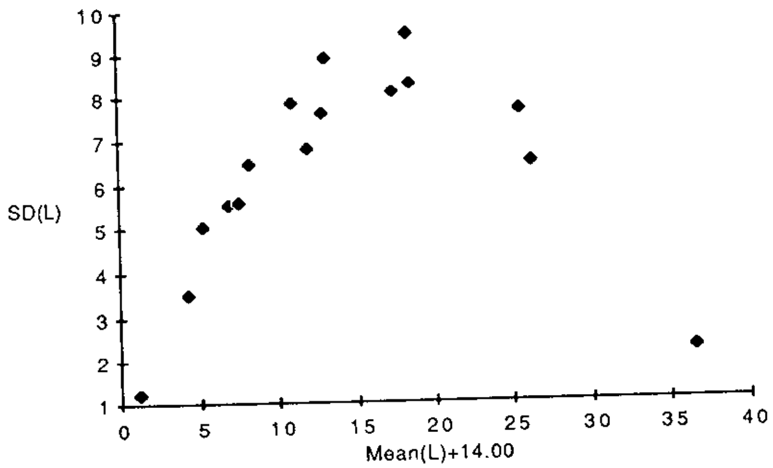

Fig. 1. Plot of the situations' mean versus their respective standard deviation (Answer modality: L. Centered data).

ents at the 16 situations versus the corresponding standard deviations suggests that the variability is not constant. As can be seen in Figure 1 it is higher in the middle and lower in the extremes. Efforts have been made to find a suitable transformation to stabilize the variance, but so far we have not succeeded. The commonly used Box-Cox (1964) transformations do not help in this case. As an example, see in Figure 2 the effect of the logarithm transformation on an adequate translation of the MEAN(L)

Regarding the data it seems logical to find this kind of heteroscedasticity — large dispersion in the middle of the range and lower in the extremes — since the data came from an instruction which fixed both extremes of the answer scale.

Because of this problem, Generalized Least Squares estimation was used to calculate the effects since it corrects for heteroscedasticity. Table II includes the estimates and the standard errors of the following 16 effects on JS: 4 main effects; 6 two factor interactions; 4 three factor interactions; 1 four factor interaction and the overall mean.

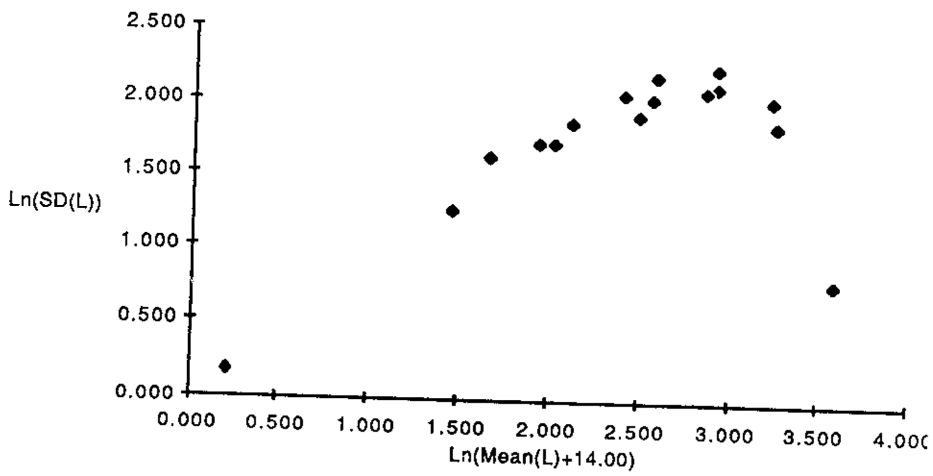

Fig. 2. Plot of the logarithm of the situations' mean versus the respective logarithm standard deviation (Response modality: L. Centered data).

TABLE II
Analysis of the complete 2^4 factorial design

	Est. effect (N)	Est. effect (L)
Average	4.0	0.17
J	272.9	10.82*
E	153.7	5.38*
A	276.2	11.02*
S	156.0	6.24*
JE	22.8	1.36
JA	76.7	3.39*
JS	26.1	1.30
EA	24.1	1.14
ES	14.2	0.50
AS	22.2	1.69
JEA	5.4	0.82
JES	12.3	0.20
JAS	16.5	0.11
EAS	2.0	0.79
JEAS	5.3	0.17
SD	10.49	0.46

* means significant effects.

3.2 *Testing and Interpretation of the results in the complete factorial design*

It is known that it is not possible to interpret the contribution of each factor on JS till the significance test has been done. This test can be carried out either by plotting the effects on normal probability paper (Daniel, 1976) or by considering the 34 respondents as "genuine" replicates, and testing the effects against the pooled variance over all situations. Since results were pretty much the same we only present a significance test based on the first procedure which is illustrated in Figures 3 and 4.

From the above figures it is very clear that the four main effects, i.e., those from the four selected factors, are significant. Also the interaction between job-security and atmosphere is significant. That means that these effects are not easily explained by chance, since they are clearly distinguishable from the noise.

Before interpreting the main effect of each factor it is necessary to

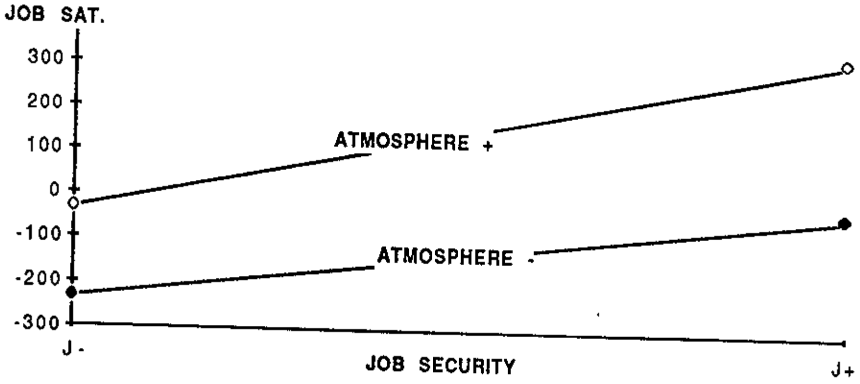

Fig. 5. The job security — atmosphere interaction: plot of the job-satisfaction average (over levels of the other factors) at the two levels of atmosphere and job security (magnitude estimation: N).

Figure 5 illustrates the interpretation of this second order significant interaction: when the security increases the satisfaction increases more when the atmosphere is good than if it is bad.

Since it is not possible to attach a physical meaning to the units used, the interpretation of the effects is better done by comparison between their relative magnitudes. So, for these inhabitants of Holland, job-security and atmosphere factors have the same effect and the effect is twice as large, considering the interaction effect as well, than as the effect of the other two factors.

Finally, it can be seen that the other two factors, salary and education relatedness are also significant, and make an equal contribution to the general judgment of JS. As usually occurs other higher order interactions are not significant. Actually, they are used as references in the testing procedure as the proper measure of the noise (See Figures 3 and 4).

4. EVALUATION OF THE USE OF THE EXPERIMENTAL DESIGN

Two main problems arise when the proposed technique is used, and both appear when the number of factors (k) is large. The first one is that when k increases the number of questions (situations) increases

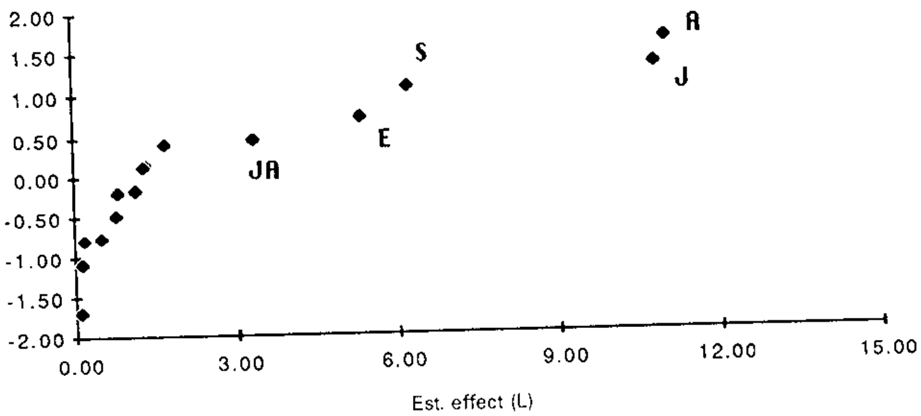

Fig. 3. Normal plot of the effects from a complete 2^4 factorial design (lines production: L).

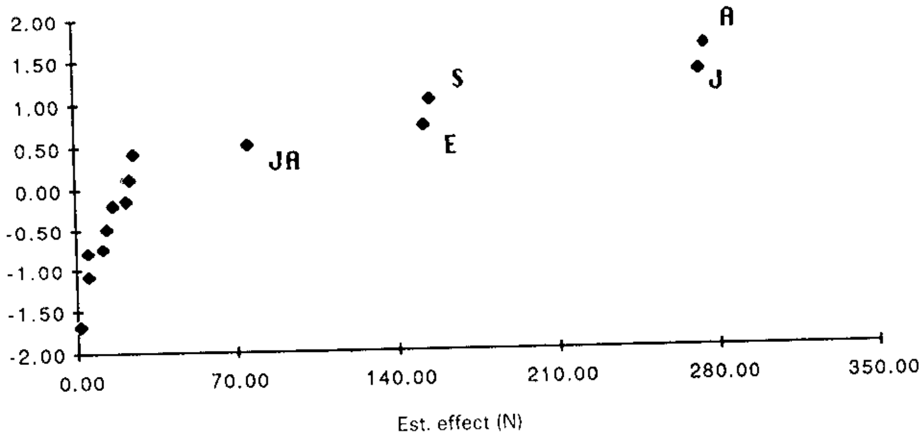

Fig. 4. Normal plot of the effects from a complete 2^4 factorial design (magnitude estimation: N).

see whether these factors behave additively or whether they "interact". If true evidence of interaction effects are found they will be considered jointly with the main effect of the factor. In our case, as was mentioned in the last paragraph, there exists an appreciable interaction between atmosphere and job-security, thus the effects of these two factors will be considered jointly.

geometrically, being 2^k (where 2 is the number of levels and k is the number of factors). The second problem is the length of the statements which have to be used to describe the different situations.

It is well known that the length of the questionnaire is related to the number of "missing values", to the number of incoherent combinations in the responses, and respondents' fatigue which diminishes the reliability and validity of this measurement instrument. It is also evident that the length of the stimuli provided to the respondents has a negative relationship with the quality of the responses (Schuman and Presser, 1981; Belson, 1981; Verwey *et al.*, 1986).

However, when continuous answer scales are used, the use of factorial designs can greatly improve and simplify the analysis of the results as we have seen, and we will show that also in some cases the number of considered items in the questionnaire can be simplified. In fact, the problem of the length of the questionnaire can be solved by the use of fractional factorial designs (See Box *et al.*, 1978). This means that the number of questions can be drastically reduced if we are willing to reduce the number of effects that are of interest in the study.

Since it is very unusual to find significant three or higher order interactions we can disregard them and concentrate the efforts on estimating the main effects and second order interactions. In fact there tends to be a certain hierarchy: main effects tend to be larger than two-factor interaction effects, which in turn tend to be larger than three-factor interactions and so on.

As it can be found in the literature (See Box *et al.*, 1978), we are not going to discuss here how to deal with fractional factorial designs. In fact what we would like to show is that we could have used this technique and conduct a half-fraction of 2^4 factorial design which would only need $2^3 = 8$ questions to estimate the same effects, and our conclusions would have been very much the same, as is illustrated in Table III, just for line production responses.

The design used named a 2_{iv}^{4-1} design, where 2^{4-1} indicates that half fraction of 2^4 complete factorial design has been used and, iv, is named "design resolution" and indicates that in the estimated effects the main effects will be confused with three-factor interactions and two-factor interactions will be confused between them (Confounding pattern).

From the data in Table III, it is clear (Box *et al.*, 1978) that the effects specified in Table IV can be estimated.

For example the value of 11.61 of **J + EAS** should be interpreted, as the estimated effect of job-security **J** and the three factor-interaction **EAS** together. As can easily be proved, the estimated effects in the fractional factorial design (Table IV) coincide with the sum of the corresponding estimates in the complete design (Table II).

Notice that the only disadvantage with respect to the complete $\mathbf{2^4}$ factorial design is that the two-factor interaction **JA** is confused with the two-factor interaction **ES**. But the other conclusions would be

TABLE III
Analysis of a half-fraction of the full 2^4 design: 2^{4-1} fractional factorial design (Response modality: L)

Situation	Level factors J E A S	Response
1	− − − −	−12.75
10	+ − − +	−1.01
11	− + − +	−5.62
4	+ + − −	−1.98
13	− − + +	−0.79
6	+ − + −	4.59
7	− + + −	−2.98
16	+ + + +	22.63

TABLE IV
Analysis of a half-fraction of the 2^{4-1} fractional factorial design: Estimates of effects

	Est. effect (L)	Confounding pattern
Average	0.25	
J	11.61*	**(J + EAS)**
E	5.48*	**(E + JAS)**
A	11.22*	**(A + JES)**
S	7.06*	**(S + JEA)**
JE	3.05	**(JE + AS)**
JA	3.89*	**(JA + ES)**
JS	2.44	**(JS + EA)**

* Means significant effects (Response modality: L)

pretty much the same as the ones we have drawn from the complete 2^4 factorial design in Table I.

The second problem we have indicated is that if the number of factors k increases, the size of the description of the situation also increases. This problem has no easy solution. It is shown in the literature that people ignore information in such cases (Verwey *et al.*, 1986) or give unpredictable answers because they can not cope with the problem. Perhaps the best thing to do is to restrict the use of the experimental design to situations in which k is lower than 6 or 7 unless visual or other types of aids can be used to clarify the presentation of the situations in the questionnaire. Although this limits the possible application of the experimental design, we think that there are enough topics where this approach can be used in a fruitful way.

5. DISCUSSION

The main differences between the non-experimental and the experimental approach arise as a consequence of the data collection design. What other researchers (See every reference in this line from Andrews 1974 to Hodson 1986) have been asking the respondents is: "How satisfied are you with your work?" — and sometimes other indirect measures such as whether the worker would recommend the job to a friend. Besides, other variables of the respondents are also measured and then used as predictors of JS in a statistical model which relates both sets of measures.

In our research, instead of collecting information on JS and on their determinants separately (See for example, the "objective indicators" used for life satisfaction by Fernández and Kulik, 1981, or the "subjective indicators" for WB refered in Campbell, 1981, and in Larsen *et al.*, 1985), respondents were asked to evaluate their own JS in the situations defined by the explanatory variables.

On the one hand, this experimental approach provides us with more control over the whole design of the study and allows us to detect interaction effects in an easy way. These experimental designs are easily implemented and if the number of predictor variables is too large the use of fractional factorial designs will allow shorter questionnaires to be composed. Specific software is also available to analyze these data

(JASS, 1985). In this article we have only mentioned factorial designs at two levels, but if it is reasonable to assume that the response relation is quadratic, designs at three levels could be used. In general all the techniques of experimental design can be applied, including such new ideas as the use of this technique to study variability (Taguchi and Wu, 1980; Box and Meyer, 1984; Tort-Martorell, 1986).

On the other hand, there is the disadvantage that the provided stimuli become too long and too complex for the respondents. Perhaps the best thing to do is, as was mentioned, to restrict the use of the experimental design to situations in which k is lower than 6 or 7 unless visual or other types of aids can be used to clarify the presentation of the situations in the questionnaire.

So far it is not clear whether the results can be compared with the results from the previous studies. Our results, unlike previous studies (Kalleberg, 1977; Saris, 1982; Hodson, 1986), signal that nowadays job-security and atmosphere are much more important than salary or education relatedness in explaining JS. This is not a surprising result since unemployment has become a major problem in recent years. A significant interaction between job-security and atmosphere has been detected as well.

Confirmatory studies of these findings in the Dutch population should follow this one. Comparative studies with other countries where, for instance, salary is smaller, would also be necessary to increase our understanding of the relative merits of the determinants of JS.

In addition to the mentioned aspects, we have already emphasized those differences among the JS' predictors used by different researchers. Since the factors or predictors used are not independent the statement "what goes in is what comes out" holds also for our models. Their relative effects on JS will depend to a great extent on the factors considered as explanatory variables.

In this way, future research should include together with our four factors age, which like Kalleberg and Loscocco (1983) we think can explain large differences in JS, as well as the two "extrinsic dimensions" already used by Kalleberg (1977) "chances of promotion and resources adequacy". Since age is measured apart from the data collection design it does not enlarge our questionnaire, and these last two factors can be easily added to our questions (situations). Moreover, by using fractional factorial designs the length of the corresponding questionnaire (in this case of six factors) would be reduced to 32 questions.

ACKNOWLEDGMENT

(*) This research has been made possible by a grant to the first author of the article from "Fundacio Jaume Bofill" (Barcelona) and the facilities provided by Sociometric Research Foundation (Amsterdam) in the analysis and data collection.

BIBLIOGRAPHY

Abbey A. and Andrews F. M.: 1985, 'Modelling the psychological determinants of life quality', Social Indicators Research 14, 1—34.
Andrews F. M. and McKennell A. C.: 1980, 'Measures of self-reported well-being: their affective, cognitive and other components', Social Indicators Research 8, 127—155.
Andrews F. M. and Withey S. B.: 1974, 'Developing measures of perceived life quality: results from several surveys', Social Indicators Research 1, 1—26.
Andrews F. M. and Withey S. B.: 1976, Social Indicators of Well-Being: Americans' Perceptions of Life Quality, Plenum Press, New York.
Batista-Foguet J. M. and Saris W. E.: 1988, 'Two standards in an experiment with non-metric stimuli', in W. E. Saris (Ed.): Variation in Response Behaviour: a Source of Measurement Error, Sociometric Research Foundation.
Belson W. A.: 1981, The Design and Understanding of Survey Questions, Adelshot, Gower.
Bock R. D. and Bargmann R. E.: 1960, 'Components of variance analysis as structural and discriminant analysis for psychological tests', British Journal of Statis. Psichol, 8, 151—163.
Bock R. D. and Bargmann R. E.: 1966, 'Analysis of Covariance Structures', Psychometrika 31, 507—34
Box G. E. P. and Cox D. R.: 1964, 'An Analysis of Transformations', J. Roy. Stat. Soc. Series B, 26, 211—52.
Box G. E. P., Hunter W. G., and Hunter J. S.: 1978, Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building, Wiley, New York.
Box G. E. P. and Meyer R. D.: 1984, 'Analyzing Two Level Fractional Factorial Experiments for Possible Dispersion Effects', University of Wisconsin M. R. C. Tech. Report n. 2746.
Burt R. S., Wiley J. A., Minor M. J., and Murray J. R.: 1979 'Structure of Well-Being: Form, Content and Stability over Time', Sociological methods and Research 6, 365—407.
Campbell A.: 1981, 'The Sense of Well Being in America', Recent Patterns and Trends, McGraw-Hill.
Cronbach L. J., Gleser G. C., Nanda H., and Rajaratnam N.: 1972, Dependability of behavioural measurements: Theory of generalizability of scores and profiles, Wiley, New York.

Daniel C.: 1976, Applications of Statistics to Industrial Experimentation, Wiley, New York.
Diener E.: 1984, 'Subjective Well-Being' Psychological Bulletin 95, 542—75.
Doorn L. V., Saris W. E., and Lodge M.: 1983, 'Discrete or continuous measurement: what difference does it make?' Kwantitatieve Methoden, 10.
Fernández R. M. and Kulik J. C.: 1981, 'A multilevel model of life Satisfaction: Effects of individual characteristics and neighborhood composition', American sociological Review 46, 840—50.
Hodson R.: 1986, 'Corporate structure and job satisfaction: A focus on employer characteristics', Sociology and Social Research, 70, 113—33.
Hortley J. and Little B. R.: 1985, 'Affective and Cognitive components of global subjective well-being measures', Social Indicators Research, 17, 139—97.
Hox J. J., Kreff G. G., and Hermkens P. L. J.: 1988, 'Multilevel models for factorial surveys', Paper presented at the International Conference on Social Sciences Methodology. Dubrovnik, Yugoslavia: 30 May—3 June.
JASS (Version 1.2): 1985, Software of Analysis of Experimental Design, Joiner Associates.
Kalleberg A. L.: 1977, 'Work values and job rewards: A theory of job satisfaction', American Sociological Review, 42, 124—43.
Kalleberg A. L. and Loscocco K. A.: 1983, 'Aging, values, and rewards: Explaining age differences in job satisfaction', American Sociological Review, 48, 78—90.
Krantz D. H. et al.: 1972, 'A theory of magnitude estimation and cross-modality matching', Journal of mathematical psychology, 9, 168—99.
Larsen R. J., Diener ED., and Emmons R. A.: 1985, 'An evaluation of subjective Well-Being measures', Social Indicators Research, 17, 1—17.
Levy S. and Guttman L.: 1975, 'On the multivariate structure of well being', Social Indicators Research, 2, 361—88.
Louviere J. J.: 1988, Analyzing Decision Making. Metric Conjoint Analysis. Sage.
Mallenberg G. J., Kelderman H., Stijlen J. G., and Zondag E.: 1979, 'Linear models for the analysis and construction of instruments in a facet design', Psychological Bulletin, 86, 766—76.
McKennell A. C.: 1978, 'Cognition and affect in perceptions of Well-Being', Social Indicators Research, 5, 389—426.
McKennell A. C. and Andrews F. M.: 1980, 'Models of cognition and affect in perceptions of Well-Being', Social Indicators Research, 8, 257—98.
McKennell A. C. and Andrews F. M.: 1983, 'Components of perceived Life Quality', Journal of Community Psychology, 11, 98—110.
McKennell A. C., Atkinson T., and Andrews F. M.: 1980, 'Structural constancies in surveys of perceived Well being', in A. Szalai and Andrews (eds.): Quality of Life: Comparative studies, Sage.
Miller J.: 1980, 'Individual and occupational determinants of job satisfaction', Sociology of Work and Occupations, 7, 337—66.
Minor M. J., Bradburn N. M., and Schaeffer N. C.: 1980, 'The structure of Life-Satisfaction: A comparative analysis across social groups', in Szalai and Andrews F. M. (eds.): Quality of Life: Comparative studies, Sage.
Organ D. W. and Near J. P.: 1985, 'Cognition vs affect in measures of job satisfaction', International Journal of Psychology, 20, 241—53.
Quinn R. P. and Mangione T. W.: 1973, 'Evaluating weighted models of measuring job satisfaction: a Cinderella story', Organizational Behaviour and Human performance, 10, 1—23.
Rossi P. H., and Anderson A. B.: 1982, 'The factorial survey approach: an introduction', In P. H. Rossi and S. L. Nock (Eds.): Measuring social judgments. Sage.

Saris W. E. (ed.): 1988, Variation in Response Behaviour: a source of measurement error, Sociometric Research Foundation.
Saris W. E., Bruinsma C., Schoots W., Vermeulen C.: 1977, 'The use of magnitude estimation in large scale survey research', Mens en Maatschappy, 52, 369–95.
Saris W. E., Neijens P., and Doorn L. V.: 1980, 'Scaling Social Sciences variables by multimodality matching', Kwantitatieve Methoden 2, 3–21.
Saris W. E., and Pijper M.: 1986, 'Computer assisted interviewing using home computers', Elsevier, European Research, 4, 144–51.
Schuman H. and Presser S.: 1981, Questions and answers in attitudes surveys, Academic Press.
Smith P. C., Kendall L. M., and Hulin C. L.: 1969, The measurement of satisfaction in work and retirement: A strategy for the study of attitudes, Rand McNally and Cia.
Stevens S. S.: 1971, 'Issues in psychophysical measurement', Psychological Review, 78, 426–50.
Stevens S. S.: 1975, Psychophysics: Introduction to its Perceptual Neural and Social Prospect, Wiley, New York.
Szalai A. and Andrews F. M. (eds.): 1980, The Quality of Life: Comparatives studies, Sage.
Taguchi G., and Wu Y.: 1980, Introduction to Off-Line Quality Control, Central Japan Quality Control Association.
Tort-Martorell X.: 1986, Diseños factoriales fraccionales, aplicaciòn al control de calidad mediante el diseño de productos y procesos, Unpublished Phd. dissertation. Polytechnic of Catalunya.
Tursky B. and Lodge M.: 1979, 'Comparison between Category and Magnitude scaling of Political', Opinion employing SRC/CPS items, American Political Sciences Review, 73, 50–66.
Verweij M. J., Saris W. E., *and Bonsel G.: 1986, 'A measurement procedure for the evaluation of health states by the general public' Sociometric Research Foundation, R.M. 861016
Wannous J. P. and Lawer III E. E.: 1972, 'Measurement and meaning of Job Satisfaction', Journal of Applied Psychology, 56, 95–105.
Wish N. B.: 1986, 'Are we really measuring the Quality of Life? Well-being has subjective dimensions as well as objective ones', American Journal of Economics and Sociology, 45, 93–99.
Zajonc R. B.: 1980, 'Feeling and Thinking: preferences need no inferences', American Psychologist, 35, 151–75.

Faculty of Psychology,
Univerzidad de Barcelona,
Zona Universitaria, 08028,
Barcelona,
Spain.