Different Scales for Different Survey Methods: Validation in Measuring Quality of University Life

(Article begins on next page)

11 May 2024

ADVANCES IN QUALITY-OF-LIFE THEORY AND RESEARCH

and A. COSKUN SAMLI

Social Indicators Research Series

Volume 20

This new series aims to provide a public forum for single treatises and collections of papers on social indicators research that are too long to be published in our journal *Social Indicators Research*. Like the journal, the book series deals with statistical assessments of the quality of life from a broad perspective. It welcomes the research on a wide variety of substantive areas, including health, crime, housing, education, family life, leisure activities, transportation, mobility, economics, work, religion and environmental issues. These areas of research will focus on the impact of key issues such as health on the overall quality of life and vice versa. An international review board, consisting of Ruut Veenhoven, Joachim Vogel, Ed Diener, Torbjorn Moum, Mirjam A.G. Sprangers and Wolfgang Glatzer, will ensure the high quality of the series as a whole.

ADVANCES IN QUALITY-OF-LIFE THEORY AND RESEARCH

Edited by

M. JOSEPH SIRGY
Virginia Polytechnic Institute and State University, U.S.A.

DON RAHTZ
The College of William and Mary, U.S.A.

and

A. COSKUN SAMLI
University of North Florida, U.S.A.

TABLE OF CONTENTS

PREFACE

I am honored to be invited to write this preface. A few years ago when we started ISQOLS, I certainly did not dream of writing a preface of this magnitude.

This book has 14 chapters. These are very important culmination of very intense research efforts. They reiterate my conviction that quality of life research is a very broad multidisciplinary topic that needs to be explored in a multidisciplinary manner. The chapters of this book not only reiterate the conviction of QOL as a multi-disciplinary area of investigation, but also each chapter offers an aspect of explo-ration of this all-encompassing discipline. Each chapter not only offers some profound findings in different areas of QOL research but also makes a contribu-tion to understand that under similar circumstances the methodologies utilized in each chapter can be further utilized in other cultures, geographic areas, different socioeconomic and demographic groups as well as in different QOL domains.

I truly consider this volume a treasure of research and exploration. Chosen from numerous and valuable other submissions, the chapters of this volume have some major claims to make. First, they reiterate the need of knowledge and creativity to undertake QOL related research. Second, they identify the difficulties and impediments in undertaking social research. Third, each chapter makes a concrete contribution to the aspect QOL research it set out to do. Fourth, the more labo-rious and time consuming the research efforts, the greater the importance of findings and profundity of the conclusions. Finally, and perhaps most importantly, we, in different disciplines of social and behavioral sciences, all have a major contribu-tion to make to the knowledge pool therefore we must interact and communicate more through interdisciplinary activities such as the development of this volume.

I would like to congratulate ISQOLS putting together such a path breaking volume. I certainly hope that such an effort is not just going to be a one-time activity but a continuing tradition.

Young, serious minded, ambitious and devoted colleagues will have a picnic with this book. All I can say is "may the force by with you." We need more and even better efforts so help us.

A. Coskun Samli, Ph.D.
Research Professor
University of North Florida

F. MAGGINO AND S. SCHIFINI D'ANDREA

14. DIFFERENT SCALES FOR DIFFERENT SURVEY METHODS: VALIDATION IN MEASURING THE QUALITY OF UNIVERSITY LIFE

ABSTRACT: This chapter explores the assessment of subjective measurement instruments in terms of definition and selection of items and also in the identification of more suitable scales. The work presented here addresses the comparison of several scales in separate types of questionnaires (paper and/or CATI). It provides a conceptual framework and empirical analysis. Finally, it draws conclusions based on these analyses.

14.1. INTRODUCTION

The measurement of individual well-being and quality of life has appeared in many different forms, assessing well-being from different perspectives, and using different measures and different extensive scales to do so. The assessment of subjective measurement instruments needs special attention not only in the definition and selection of items but also in the identification of more suitable scales. This is particularly true in the measurement and assessment of the subjective perception of quality of life (in rating agreement levels, satisfaction levels, and so on), due to the disparities between areas, considered in studies found in the literature.

Synthetically, aspects involved in scale definition are:

1. Reference type (evaluation, preference, perception, image, judgment);
2. Scale type (expression of scale: verbal, rating, quantitative and graphical scale);
3. Range (number of levels for scale) in the sense of scale discriminate capacity.

The choice among all aspect combinations can mainly influence the construction and validation of indicators.

Since other elements could play important roles in this definition, such as investigated areas, semantic and cultural meanings, and survey methods, not all scales can be used in different situations. In reality, the problem of semantic definition and selection of items becomes complicated because of differences between questionnaire forms and leading survey types, such as paper questionnaire, presence of interviewers or not, Computer Assisted Telephonic Interviews (CATI), Web interviewing, and so on, particularly in the adaptation of items to more than one survey method.

The work presented here concerns the comparison of several scales in separate questionnaires (paper and/or CATI); this experience has allowed us to compare scales with:

1. Different reference (judgment vs. agreement, judgment vs. evaluation, image vs. agreement);
2. Different scale types (verbal vs. rating and/or graphical vs. numerical);
3. Rating scale with different ranges (0–10 vs. 1–7 and/or 1–7 vs. 1–4).

233

The goals of this study are to:

*   Cross-validate different questionnaires (paper and CATI);
*   Test the reliability of different scales;
*   Evaluate impacts of different scores and scale meanings in selection of Quality of Life indicators (in University context);
*   Compare individual levels of satisfaction and evaluation.

14.2. DIFFERENT SURVEY METHODS IN THE MEASUREMENT OF THE QUALITY OF UNIVERSITY LIFE

In order to reach our goals, we carried out surveys concerning the quality of university life, applying two different survey techniques (paper and telephonic) and three different questionnaires (paper and two telephonic, a and b), on three different samples of students (Table 14.1) of the Faculty of Economics at the University of Florence (Italy).

Table 14.1. Dimensions of the Sample.

Questionnaire		Sample dimension
Paper		300
CATI	a (498)	1015
	b (517)	

14.2.1. *The Conceptual Model*

The three questionnaires present the same conceptual model, consisting of two areas:

1. Quality of life, for which we identified three components:
    *   self-esteem
    *   general subjective well-being
    *   subjective well-being in living contexts
2. Quality of university life, for which we identified three components:
    *   personal motivation
    *   university career
    *   university environment

We identified a set of variables for each component, measured by means of scales with differentiated structures. The questionnaire structure also includes objective information like age, gender, birthplace, family background, etc. Table 14.2 shows the whole questionnaire structure.

Table 14.2. Questionnaire Structure.

EXTERNAL VARIABLES		o Gender
		o Age
		o University curriculum
		o Employment
		o Distance from University
INDIVIDUAL TRAITS AND DISPOSITION		o Self-esteem
		o Personal motivation towards study
ENVIRONMENT		o Family support
		o Friends support
VALUES		o Importance of particular ambits in one's life
SATISFACTION AND WELL-BEING PERCEPTION		o General Subjective Well-being (General Life Satisfaction)
		o Subjective Well-being in particular life ambits (Friendship, Family, Money, Free time, Health, Faculty, University career, University friendship)
		o Student Life Satisfaction
		o Happiness (at the present, one year ago)
UNIVERSITY LIFE	**Career Performances**	o Actual Performances (Successful Examination Number, Taking Examination Number, Marks Average, Proportion of successful exams towards requested standard, Course attendances at the present)
		o Perceived Performances (compared to other students, past expectations, future intentions)
		o Attitude towards Performances
	University evaluation	o Faculty Evaluations
		o Exam Perception

Scales for Quality of Life

To investigate Quality of Life area we considered three well-being levels:

- *Self-esteem:* We used the Rosenberg *ten-items scale*, which measures self-esteem as a personality trait and as having confidence in one's own worth and abilities. Students had to indicate their agreement on each statement regarding self-worth.
- *General subjective well-being:* The subjective well-being measurement has three identifiable components: emotional component, in the sense of *happiness*, cognitive component, in the sense of *general life satisfaction*, cognitive component for study satisfaction: in order to measure student judgment on his/her student life satisfaction.
- *Subjective well-being in living contexts:* We investigated subjective well-being in living contexts for different social domains such as friendships, family relationships, health, university career, economic status, etc. One of the goals was to synthesize these items in synthetic indexes, such as family and social relations, financial resources, and university involvement.

Scales Quality of University Life

For the Quality of University Life area, we identified three related components:

- *Motivation toward study:* Personal motivation is one of the components on which, in our opinion, both individual career and environmental perception, depend. Personal motivation towards study in the University context was assessed by student agreement towards ten statements about e.g. future expectations, learning motivations vs. dropping-out inclination, self-evaluation capacities in relation to study achievements and curricula self-evaluation.

- *Career performances:* We measured this dimension through two approaches: (1) *real performances*, which used the following indicators: grade average, mean of taken exams, and proportion of successful exams towards requested standard; and (2) *perceived performances*, the measurement of which followed the Multiple Discrepancies Theory; we defined and identified comparisons with other students, past experiences, students' own potentials, and future expectations with respect to the number of exams taken and average grades.
- *University environment:* In order to identify aspects characterizing the relationship between students and the university environment, we defined a model that highlighted the cognitive-emotional judgment on the environment, measured through two approaches: (1) satisfaction in particular domains: we identified 16 living contexts; each student had to express his/her satisfaction level for each context on an eleven-point scale (only in paper-questionnaire), and (2) subjective evaluation: students have expressed their evaluations through semantic differential scales.

14.2.2. Different Item Approaches for Different Questionnaires

Because of different survey techniques, our questionnaires required different item approaches and definitions with regard to:

- scale reference
- scale type
- scale range

Let us examine these different approaches.

Scale Reference

The difficulty in adapting some particular items (such as graphical scales) in forms appropriate to telephonic interviews has been overcome by asking students about their agreement regarding some defined assertions (Table 14.3). In telephonic questionnaires, in place of semantic differential scales presented in the paper-questionnaire, we defined two different adjective groups concerning university: positive and negative. This procedure allowed us to verify the real polarity of adjectives.

Table 14.3. Different Scale References Chosen for Our Three Questionnaire.

Areas	Variables	Scale reference		
		Paper-Q.	Cati-Q.	
			a	b
University Evaluation	Faculty Evaluations	Image	Agreement (Positive adjectives)	Agreement (Negative adjectives)
Satisfaction and Well-Being Perception	Student Life Satisfaction	Judgment	Agreement	Agreement
	Happiness at the Present	Judgment	Evaluation	Evaluation

Scale type

We changed graphical (Face Scale, Self Anchoring Ladder Scale, Semantic Differential Scales) and labeled scales of paper-questionnaire into equivalent rating scales in telephonic interviews. For instance, in the paper-questionnaire, students evaluated their student life by the Ladder Scale (Cantril), in graphical form, while in the CATI-questionnaires we adopted a different approach: students had to refer their agreement regarding an assertion about their student condition (Table 14.4).

Notice that the only variable measured by a verbal scale in all questionnaires is Personal Motivation towards Study scale.

Table 14.4. Different Scale Types Chosen for Our Three Questionnaires.

Areas	Variables	Scale type		
		Paper-Q.	Cati-Q.	
			a	b
University Evaluation	Faculty Evaluations	Graphical	Numerical	Numerical
Satisfaction and Well-Being Perception	Student Life Satisfaction	Graphical (Self Anchoring Ladder Scale)	Numerical	Numerical
	Happiness at the Present	Graphical (Face Scale)	Numerical	Numerical
Individual Traits and Dispositions	Self-esteem	Verbal	Numerical	Numerical

Scale Range

One of the hypotheses raised regarding rating scale concerns the discriminate capacities for scales with different rating amplitude. In order to test this hypothesis, we defined different scale ranges for our three questionnaires by assigning different scale amplitude alternatively to questionnaires (Table 14.5).

Table 14.5. Different Scale Ranges Chosen for Our Three Questionnaires.

Areas	Variables	Scale range		
		Paper-Q.	Cati-Q.	
			a	b
Satisfaction and Well-Being Perception	General Life Satisfaction	0-10	0-10	1-7
	Subjective Well-Being in Particular Ambits	0-10	0-10	1-7
	Student Life Satisfaction	1-9	0-10	1-7
	Happiness at the Present	1-7	1-7	0-10
	Happiness One Year Ago	--	1-7	0-10
Values	Importance of Particular Ambits in one's Life	--	1-7	0-10
Individual Traits and Dispositions	Self-esteem	1-4	1-5	1-7
	Motivation	1-5	1-4	1-4

Table 14.6 allows us to summarize the experimental design showing comparison among scales we used in paper-questionnaire and CATI-questionnaires. For each variable, tables show:

- number of used items
- used reference (agreement, judgment, etc.)
- scale type
- scale range.

Table 14.6. Different Scale Reference, Type, and Range for Paper- and CATI-questionnaire.

Areas	Variables	Paper-Questionnaire				Cati-Questionnaire A				Cati-Questionnaire B			
		N. of items	Reference	Type	Range	N. of items	Reference	Type	Range	N. of items	Reference	Type	Range
UNIVERSITY EVALUATION	Faculty Evaluations	23	Image	Graphical (no numerical reference)	1-7	9	Agreement (Positive adjectives)	Numerical*	1-7	9	Agreement (Negative adjectives)	Numerical*	1-7
SATISFACTION AND WELL-BEING PERCEPTION	General Life Satisfaction	1	Evaluation	Numerical	0-10	1	Evaluation	Numerical*	0-10	1	Evaluation	Numerical*	1-7
	Subjective Well-Being in Particular Ambits	10	Evaluation	Numerical	0-10	10	Evaluation	Numerical*	0-10	10	Evaluation	Numerical*	1-7
	Student Life Satisfaction	1	Judgment	Graphical (Self Anchoring Ladder Scale)	1-9	1	Agreement	Numerical*	0-10	1	Agreement	Numerical*	1-7
	Happiness at the Present	1	Judgment	Graphical (Face Scale)	1-7	1	Evaluation	Numerical*	1-7	1	Evaluation	Numerical*	0-10
	Happiness One Year Ago					1	Evaluation	Numerical*	1-7	1	Evaluation	Numerical*	0-10
VALUES	Importance of Particular Ambits in one's Life					16	Judgment	Numerical*	1-7	16	Judgment	Numerical*	0-10
INDIVIDUAL TRAITS AND DISPOSITIONS	Self-esteem	10	Agreement	Verbal	1-4	10	Agreement	Numerical*	1-5	10	Agreement	Numerical*	1-7
	Motivation	10	Agreement	Verbal	1-5	10	Agreement	Verbal	1-4	10	Agreement	Verbal	1-4

Items marked by (*) are anchored verbally.

Number of items with the same scale amplitude differs among three groups because of the different number of items for questionnaire areas.

14.3. DATA ANALYSIS

The goals of data analysis, presented here, are to:

- Compare different performances of positive and negative references in individual evaluations. The analysis of distributions by graphical representations and statistical moments for quantitative data (from first moment to skewness index) allowed us to make these comparisons.
- Compare different scales for single items to evaluate different discriminate capacities. The graphical representations and statistical moments for quantitative data (from first moment to skewness and kurtosis indexes) of standardized scores allowed us to make these comparisons.
- Compare the impact of items presented in the different scales in selection of indicators. The Principal Component Analysis and the Additive Trees approaches have allowed us to explore, respectively, aggregations of items and aggregation process.
- Compare reliability of multi-item indicators with different scale types and scale ranges under different survey conditions; Internal Consistency Analysis allowed us to test reliability.

The analysis includes the areas and variables as reported in Table 14.7.

Table 14.7. Variables and Respective Number of Items Included in Our Analysis.

AREAS	VARIABLES	Number of defined Items		
		Paper	Cati (a)	Cati (b)
UNIVERSITY EVALUATION	Faculty Evaluations	23	9	9
SATISFACTION AND WELL-BEING PERCEPTION	General Life Satisfaction	1	1	1
	Subjective Well-Being in Particular Ambits	10	10	10
	Student Life Satisfaction	1	1	1
	Happiness at the Present	1	1	1
	Happiness One Year Ago		1	1
INDIVIDUAL TRAITS AND DISPOSITIONS	Self-esteem	10	10	10
	Motivation	10	10	10
VALUES	Importance of Particular Ambits in one's Life		16	16

14.3.1. *Comparability among Groups*

In order to verify the real comparability of our groups, we test statistical significance of difference between samples using external variables and applying the proper statistical test for independent samples (parametric or non-parametric test depends on measurement scale). None of the variables has registered a significant difference at the defined α-value (0.01).

Here, we show some outcomes yielded by exploring some consequences of using different survey condition and different rating scales in:

- individuation of subjective University Evaluation dimensions through Semantic Differential Scales;
- discrimination of individual Happiness and Satisfaction Perceptions and in pointing out well-being dimensions;
- validation of individual traits measures.

For each presented analysis we show tables showing, according to the goals of specific analysis, frequency distributions, descriptive indexes, graphical representations, factorial loading matrixes and/or reliability analysis indexes.

14.3.2. *Semantic Differential Scales in Different Survey Conditions*

One of the most useful tools for measuring individual images and evaluations is the Semantic Differential Scale (SDS). However, as we know, the identification of real bipolar adjectives represents one of its limits. In order to verify the bipolarity of suitable adjectives in our context and overcome the difficult application of SDS in telephonic interviews we defined three different approaches:

- 23 Semantic Differential Scales with graphical scales (0–6 points) in paper-questionnaire
- 9 Stapel scales defined by 'positive' adjectives with agreement rating scales (1–7 points) in CATI-questionnaire (a)
- 9 Stapel scales defined by 'negative' adjectives with agreement rating scales (1–7 points) CATI-questionnaire (b).

One of our goals is to validate an efficient shape for SDS in subjective university evaluation in different survey contexts. The traditional SDS approach (Osgood, 1957), which requires a large number of graphical items, is not suitable in telephonic contexts that require a different approach from that used in the paper questionnaire, e.g., the Stapel technique (Alreck and Settle, 1985). According to Stapel technique, each item requires only one adjective. This causes a change in scale references and requires a choice to be made between positive or negative adjectives. This choice is not trivial and needs to take into consideration the presence of a real semantic bipolarity between them. Our goal is to verify the appropriate item definition by testing the real bipolarity of defined and selected items.

Since efficiency in telephonic contexts is mostly related to brevity, we selected a smaller group of items from 23 items presented in paper-questionnaire. We were able to select items by taking Principal Component Analysis results into consideration in order to reject items that were clearly irrelevant in our context or with the same semantic content of others, and to introduce selected items in CATI-questionnaires.

In our experimental design, we defined two different groups of adjectives for CATI-questionnaires: positive ones for a-questionnaire and negative ones for b-questionnaire (Table 14.8). We first tested the hypothesis of bipolarity for these adjectives and then we compared the factorial composition of student evaluations between samples.

Table 14.8. Positive and Negative Adjectives for the Two CATI-questionaires.

CATI-Questionnaires		
a	Item number	b
stimulating	1	boring
useful	2	Useless
organized	3	Disorganized
encouraging	4	Discouraging
dynamic	5	Inactive
innovator	6	Traditional
simple	7	Difficult
easy	8	Hard
rewarding	9	Disappointing

Testing Bipolarity of Adjectives

Bipolarity evidence for two adjectives can be derived from the observation of two specula distributions for each pair of adjectives. For this purpose, we observed both graphical representations and skew ness values (Table 14.9).[1] In this context, we use these statistical tools as bipolarity indexes.

Table 14.9. Positive and negative Adjectives for the Two CATI-questionnaires: Statistical Indexes.

		median	mean	standard dev	skewness
a	stimulating	5.0	4.6	1.4	-0.4
	Useful	6.0	5.6	1.3	-1.0
	organized	4.0	3.8	1.5	-0.1
	encouraging	4.0	3.5	1.4	-0.1
	Dynamic	4.0	4.0	1.5	-0.2
	innovator	4.0	4.2	1.6	-0.2
	Simple	3.0	3.1	1.5	-0.2
	Easy	3.0	2.9	1.5	0.4
	rewarding	5.0	4.6	1.5	-0.6
b	Boring	3.0	3.0	1.7	0.5
	Useless	1.0	2.0	1.5	1.8
	disorganized	4.0	4.3	1.8	-0.2
	discouraging	4.0	4.1	1.8	-0.1
	inactive	4.0	3.9	1.8	0.1
	traditional	4.0	4.8	1.8	-0.2
	Difficult	5.0	4.8	1.5	-0.5
	Hard	6.0	5.3	1.3	-0.7
	disappointing	3.0	3.0	1.8	0.7

The analysis reveals three different groups of items. A first group is composed of adjectives whose bipolarity is guaranteed by showing symmetrical distributions (Figure 14.1) and, as a result, skew ness values near zero (normal tendencies). We may notice that the opposite for two of these adjectives is obtained by the prefix "dis" in the Italian language, as well ("organized/disorganized" = "organizzata/disorganizzata"). This kind of adjective seems to assure perfect bipolarity.

A second group is composed of adjectives whose bipolarity is indicated by
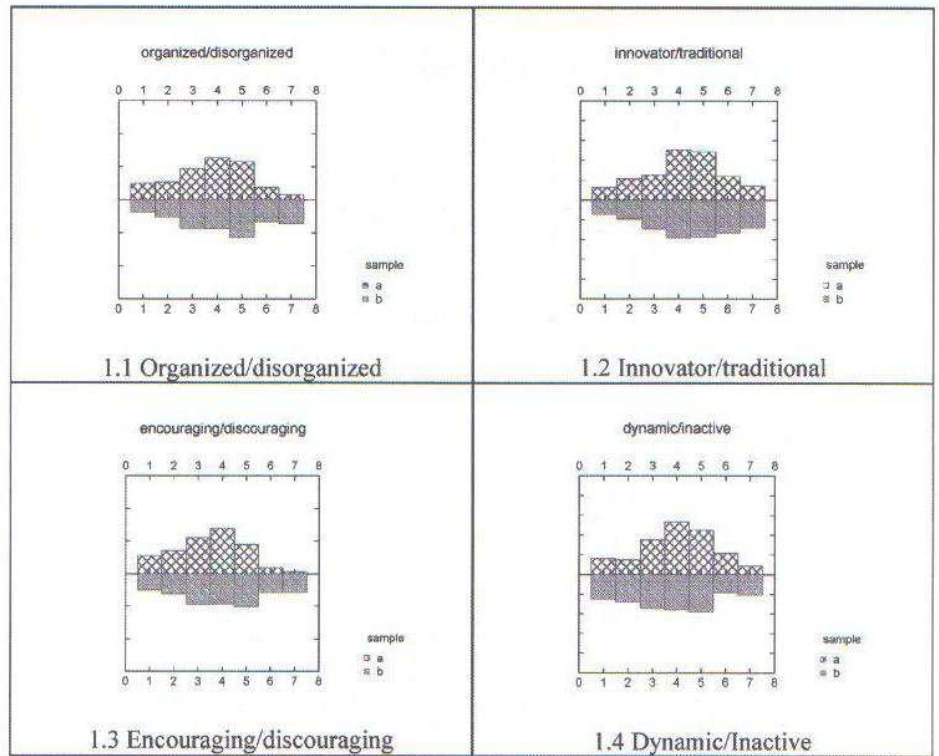
Figure 14.1. Bipolar Adjectives with Symmetrical Distributions.

showing opposite asymmetrical distributions (almost same skew ness values but of opposite signs). We can, however, also see that graphical representations (Figure 14.2) suggest a more prudent evaluation of students by negative adjectives (decreasing frequency values towards high scores) especially by 'useless' adjective.

A third group is composed of adjectives with a uncertain bipolarity (Figure 14.3).

It may be interesting to verify, in the context of applying our questionnaires in other cultural contexts, to examine if the observed pondered use of negative adjectives represents a cultural and linguistic component only of Italian students.

3.1.2. *Comparing Component Structures in Positive and Negative Adjectives*

Comparing the factorial structures yielded by Principal Component Analysis (Table 14.10), we find an interesting difference between the two adjective groups: two components for positive adjectives, three components for negative adjectives, with almost the same level of total explained variance.

2.1 Stimulating/boring

2.2 Rewarding/disappointing

2.3 Useful/useless
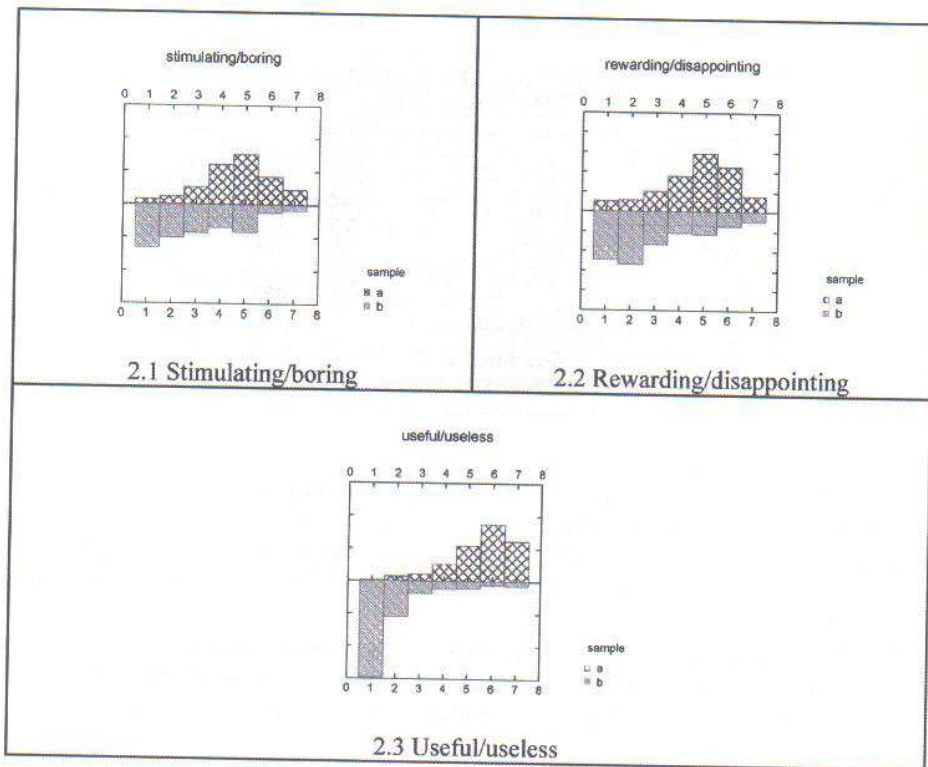
Figure 14.2. Bipolar Adjectives with Asymmetrical Distributions.

3.1 Easy/hard

3.2 Simple/difficult

Figure 14.3. Adjectives without Bipolarity.

Table 14.10. Factorial Structures in Positive and Negative Adjectives.

Rotated Loading Matrixes (VARIMAX)							
CATI-Questionnaire							
a Sample				b Sample			
1	2				1	2	3
0.780	0.034	Stimulating	Boring		0.236	-0.014	0.721
0.701	-0.030	Useful	Useless		0.032	0.012	0.798
0.684	0.080	Organized	Disorganized		0.823	-0.131	0.009
0.684	0.297	Encouraging	Discouraging		0.639	0.302	0.207
0.731	0.174	Dynamic	Inactive		0.699	0.070	0.324
0.741	0.038	Innovator	Traditional		0.681	0.165	0.232
0.066	0.914	Simple	Difficult		0.097	0.909	0.071
0.109	0.910	Easy	Hard		0.074	0.908	-0.048
0.691	0.078	Rewarding	Disappointing		0.360	0.059	0.655
3.612	1.799	"Variance" Explained by Rotated Components			2.239	1.795	1.795
40.133	19.985	%			24.883	19.949	19.948

The observation of component compositions allows us to say that the first component for positive adjectives corresponds to two dimensions for the negative adjectives; this distinction seems to be related with bipolar typologies (first dimension for negative adjectives is composed of all adjectives with symmetrical distributions).
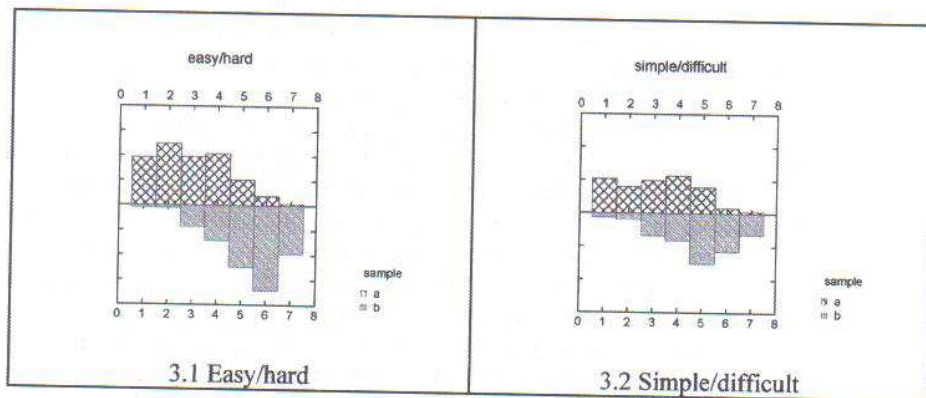
Moreover, the negative adjective solution seems to be more interpretable by showing a better discrimination among evaluation dimensions. We may label the three components respectively organizational environment (1), study (2) and psychological environment (3).

The application of the same analysis to the same adjectives in the paper-questionnaire yields same component solution of *a* group (63% of total explained variance).

This analysis seems to confirm the importance of using adjectives with a perfect bipolarity. Regarding Stapel scales (CATI-SDS), it is more useful, in order to meet individual evaluations, to apply negative adjectives. These negative adjectives seem to be used in a more meditated way. This observation needs a further investigation in order to verify its relation to cultural attitudes.

Moreover, we need to take into consideration the difficulty that occurs in converting these tools for cross-cultural surveys. The question is whether there is some relation between the cultural context and bipolarity of adjectives. Since in some cultural contexts, e.g. the Italian one, individual judgments are unlikely to be extremely positive or negative, it could be important to test the influence of cultural and linguistic factors of the respondent as regards bipolarity. This observation, also applicable to other subjective survey techniques, highlights a situation that might introduce a distortion (bias) that is difficult to evaluate, especially in cross-cultural comparisons. We come across this factor in the current research as well. The literal translation of our questionnaires into English does not always ensure a perfect semantic translation. Our intention is to plan new surveys involving different linguistic contexts.

14.3.3. *Life Satisfaction and Well-Being Perception Measures and Different Scale Ranges*

Variables defined in the conceptual model of our questionnaires, for the life satisfaction and well-being perception area, allow us to test the influence of different scale ranges in discriminating individual perceptions in single-item (general life satisfaction, student life satisfaction and happiness) and in highlighting well-being dimensions in multi-item scales.

Discriminate Capacity of Single-item Measures: General Life Satisfaction
Students referred their agreement on assertion concerning their general life satisfaction on a rating scale with different ranges

- 11 points agreement rating scale (from 0, *at all*, to 10, *completely satisfied*) in CATI-questionnaire, *a* (question 44), and paper-questionnaire (question 72)
- 7 points agreement rating scale (from 1, *at all*, to 10, *completely satisfied*) CATI-questionnaire, *b* (question 44)

Analyzing frequency distributions (Table 14.11) and descriptive statistical indexes (Table 14.12) we can observe, besides the high satisfaction levels expressed by almost all students, that

- Students with the longer rating scale did not use low score points
- Groups using the longer rating scale (paper-questionnaire and CATI-questionnaire, a) registered the same distribution shape (low concentration, long tails, revealed by low skew ness values)
- Different kurtosis values between groups using the longer rating scale (paper-questionnaire and CATI-questionnaire, *a*) may be attributed to survey effect

Table 14.11. Frequency Distribution of General Life Satisfaction Rating Scales.

FREQUENCY DISTRIBUTIONS					
CATI-Questionnaire				Paper-Questionnaire	
a		b			
RATING SCALE	PCT	RATING SCALE	PCT	RATING SCALE	PCT
0				0	
1		1	0.8	1	
2	0.2	2	1.5	2	0.7
3	1.2	3	2.3	3	2.0
4	1.2	4	6.6	4	2.4
5	3.0	5	23.2	5	7.1
6	8.7	6	38.7	6	13.9
7	25.7	7	26.9	7	30.4
8	35.4			8	22.0
9	15.8			9	14.5
10	8.9			10	7.1

Table 14.12. Descriptive Statistical Indexes for General Life Satisfaction Rating Scales (Standardized Data).

DESCRIPTIVE INDEXES	CATI-Questionnaire		Paper-Questionnaire
	a	b	
Mimimum	-4.2	-4.1	-3.3
Maximum	1.7	1.1	1.7
Median	0.2	0.2	-0.2
Skewness	-0.7	-1.3	-0.5
Kurtosis	1.5	2.3	0.6

- Group using the shorter rating scale (CATI-questionnaire, *b*) registered a compression of extremely high scores (high skew ness and kurtosis values).

These outcomes allow us to interpret the lack of extremely low values in longer rating scales as a clear positive group trend (nobody expressed a very low life satisfaction); the shorter rating scale does not allow us to reach the same conclusion even if students, using shorter rating, show the same trend.

In other words, longer rating scales are more useful in individual evaluation than the shorter rating scale, which seems unable to discriminate among extreme levels of satisfaction.

Discriminate Capacity of Single-item Measures: Student Life Satisfaction
In exploring respondent judgments about their student life satisfaction we used:

- Self Anchoring Ladder Scale by Cantril (9 steps), in paper-questionnaire (question 71)
- 11 points agreement rating scale, in CATI-questionnaire *a* (question 43)
- 7 points agreement rating scale, in CATI-questionnaire *b* (question 43).

Descriptive analysis (Tables 14.13 and 14.14) allows us to highlights, once again,

Table 14.13. Frequency Distribution of Student Life Satisfaction Rating Scales.

CATI-Questionnaire				Paper-Questionnaire	
a		b			
RATING SCALE	PCT	RATING SCALE	PCT	LADDER SCALE	PCT
0	3.6				
1	0.4	1	7.0	Step 1	0.7
2	2.3	2	10.8	Step 2	1.3
3	5.2	3	18.8	Step 3	4.0
4	8.2	4	21.7	Step 4	9.3
5	20.1	5	26.2	Step 5	24.3
6	17.4	6	9.0	Step 6	20.7
7	20.9	7	6.5	Step 7	26.3
8	14.2			Step 8	9.3
9	4.0			Step 9	4.0
10	3.8				

Table 14.14. Descriptive Statistical Indexes for Student
Life Satisfaction Rating Scales (Standardized Data).

DESCRIPTIVE INDICES	CATI-Questionnaire		Paper-Questionnaire
	a	b	Ladder Scale
Mimimum	-2.8	-1.9	-3.2
Maximum	1.9	1.9	2.0
Median	0.04	-0.01	0.04
Skewness	-0.6	-0.1	-0.3
Kurtosis	0.7	-0.6	0.1

the better capacity of longer scales in discriminating extreme agreement/disagreement levels.

As we can see, the extreme quartile group using the longer scale shows a greater dispersion among extreme scores; this could mean that students of the group using the shorter scale had to compress their attitude expressions, especially in low scores (higher frequency values for this group compared with low frequency values for other two groups).

Discriminate Capacity of Single-item Measures: Happiness
Students expressed their happiness level by one of the following approaches:

- Face Scale (7 expressions), in paper-questionnaire[2] (question 102)
- 7 points agreement rating scale, in CATI-questionnaire, *a* (question 65)
- 11 points agreement rating scale, in CATI-questionnaire, *b* (question 65)

In CATI-questionnaires students expressed their happiness level regarding both the present and past years.

Once more, a long-range scale reveals a better discriminate capacity; here this is more evident in low levels because of the strong concentration along high happiness levels registered for all students (Table 14.15). As we can see, *b* group distribution appears more concentrated (high kurtosis value) and with a long tail in correspondence with low happiness levels.

It is interesting to compare paper-questionnaire and a group distributions since they have the same range scale but different scale type (graphical vs. rating). They registered different kurtosis and very similar skew ness values (Table 14.16), revealing a less concentrated distribution for the *face-scale*. Since we cannot assume different psychological conditions between two groups, the *face-scale* seems to allow a better individual 'identification' of happiness perceptions. In other words, longer scales reveal a better discriminate capacity than shorter ones (scale-effect) and graphical scale outcomes suggest a better discriminate capacity than rating scales (survey-effect).

The better discriminate capacity of long-range scales is confirmed by observing distributions regarding happiness in the past year for CATI-questionnaires (Tables 14.17 and 14.18).[3]

Table 14.15. Frequency Distribution of Happiness Rating Scales.

Happiness level at the present					
CATI-Questionnaire				Paper-Questionnaire	
a		b			
RATING SCALE	PCT	RATING SCALE	PCT	FACE SCALE	PCT
		0	0.6		
1	0.6	1	0.4	Face 7	1.7
2	0.6	2	0.8	Face 6	2.7
3	3.0	3	0.8	Face 5	8.8
4	9.9	4	1.5	Face 4	18.6
5	29.1	5	6.0	Face 3	34.6
6	39.9	6	10.1	Face 2	25.4
7	16.8	7	27.1	Face 1	8.1
		8	36.9		
		9	8.3		
		10	7.5		

Table 14.16. Descriptive Statistical Indexes for Happiness Rating Scales.

	Happiness level at the present		
DESCRIPTIVE INDICES	CATI-Questionnaire		Paper-Questionnaire
	a	b	
Mimimum	-4.2	-4.7	-3.1
Maximum	1.4	1.7	1.6
Median	0.4	0.4	0.1
Skewness	-0.9	-1.3	-0.7
Kurtosis	1.6	3.8	0.5

Table 14.17. Frequency Distribution of Past Year Happiness Rating Scales.

Happiness level one year ago				
CATI-Questionnaire				
a		B		Paper-Questionnaire
RATING SCALE	PCT	RATING SCALE	PCT	
		0	0.6	
1	1.8	1	0.8	
2	6.3	2	1.9	
3	6.9	3	3.1	
4	13.6	4	4.3	
5	25.9	5	9.1	Not Presented
6	30.1	6	17.7	
7	15.3	7	23.2	
		8	24.1	
		9	8.4	
		10	6.8	

Table 14.18. Descriptive Statistical Indexes for Past
Year Happiness Rating Scales (Standarized Data).

| DESCRIPTIVE INDICES | Happiness level one year ago | | Paper-Questionnaire |
| | CATI-Questionnaire | | |
	a	b	
Mimimum	-2.8	-3.6	
Maximum	1.3	1.7	Not presented
Median	-0.1	0.1	
Skewness	-0.8	-0.8	
Kurtosis	0.1	0.9	

Well-being Dimensions of Multi-item Measures with Different Rating Scales
The above analysis allows us to show the importance of scale length in individual evaluation; now we need to consider in what way scale length influences the construction of synthetic subjective indicators of well being.

In our surveys, students related their agreement on assertions concerning subjective well being in 10 particular ambits on the same rating scale as they used before for general life satisfaction:

- 11 points agreement rating scale in CATI-questionnaire, *a* (questions from 45 to 54), and paper-questionnaire (questions from 73 to 82, except 80)
- 7 points agreement rating scale in CATI-questionnaire, *b* (questions from 45 to 54)

The investigated ambits are:

- *friendship*
- *free-time*
- *family relationship*
- *personal health*
- *family health*
- *faculty*
- *family financial situation*
- *university career*
- *personal financial situation*
- *university friendship*

For the sake of brevity, we will not to show all the outcomes regarding descriptive indexes. This is because we observed the same results for each item, as pointed out for the general life satisfaction single item.

In exploring and testing the influence of different scale ranges in the construction of synthetic well-being indicators, we applied Principal Component Analysis (Table 14.19).

We can very clearly observe three dimensions of life satisfaction for all groups, even if each dimension accounts for a different proportion of variance in each group. We can define dimensions by the following labels (in brackets component number for each group):

Table 14.19. Well-being Ambits: Factorial Structures for the Three Questionnaires.

| | | Rotated Loading Matrix (VARIMAX) | | | | | | | | |
| | | CATI-Questionnaire a | | | CATI-Questionnaire b | | | Paper-Questionnaire | | |
		1	2	3	1	2	3	1	2	3
items	1	0.39	0.14	0.45	0.28	0.33	0.35	0.64	0.43	0.01
	2	0.18	0.11	0.73	0.11	0.22	0.67	0.57	0.04	0.26
	3	0.11	0.12	0.78	0.21	-0.03	0.80	0.52	-0.18	0.52
	4	0.11	0.77	0.27	0.79	0.02	0.32	0.31	0.05	0.81
	5	0.01	0.88	-0.00	0.83	0.00	0.08	-0.01	0.27	0.86
	6	0.13	0.56	0.15	0.51	0.35	-0.10	0.61	0.21	0.12
	7	0.06	0.13	0.70	-0.05	0.12	0.63	0.75	0.06	0.07
	8	0.75	0.18	0.07	0.03	0.71	0.17	0.10	0.85	-0.02
	9	0.75	0.17	0.14	0.03	0.74	0.11	0.18	0.75	0.28
	10	0.70	-0.11	0.16	0.13	0.74	0.08	Not presented		
Indexes	*	1.84	1.81	1.99	1.72	1.90	1.77	2.08	1.63	1.83
	**	18.37	18.06	19.86	17.15	19.03	17.65	23.06	18.14	20.37
	***	0.60	0.62	0.68	0.56	0.66	0.59	0.64	0.65	0.70

* Variance explained by rotated components
** Variance explained by rotated components (%)
*** ALPHA only for marked items (with reference to items with high factor loadings)

- university satisfaction (1: *a* group, 2: *b* group, 2: *paper-questionnaire* group)
- money satisfaction (2: *a* group, 1: *b* group, 3: *paper-questionnaire* group)
- satisfaction in relationships (3: *a* group, 3: *b* group, 1: *paper-questionnaire* group)

Free time is related with financial satisfaction (personal and family) in both CATI-questionnaires and with relationships in paper-questionnaire; in paper-questionnaire we also observed a relation between financial situation (family and personal) and family health that disappears in the component structure of CATI-questionnaires. Item concerning friendship satisfaction shows low component loadings in each dimension for *a* and *b* groups.

Furthermore, calculating the *alpha* values for each sub-dimension, taking into account only high factor loadings items for each component,[4] we can see that the paper-questionnaire and *a* questionnaire registered better values than the other one.

The outcomes of principal component analysis can be integrated by exploring the items aggregation process through the Additive Trees clustering method,[5] which yielded for each group a solution confirming preceding outcomes. The explained variance for three groups is, respectively, 87, 74 and 77 per cent. For the sake of brevity, we present only the graphical outcomes (tree diagrams – Figure 14.4).

In Figure 14.4, we also show the corresponding labels for each item for each questionnaire in order to appreciate three tree diagrams.

The Additive Tree analysis confirms the component structures of the a and b groups: the item aggregation process allows the identification of the same three dimensions. The item concerning friendship satisfaction (question 45) reveals its extraneousness as regards chosen variables.

In the paper-questionnaire group, Principal Component Analysis revealed some superimposition in two satisfaction dimensions (university satisfaction and relationship satisfaction); observing the tree diagram, the superimposition can be seen
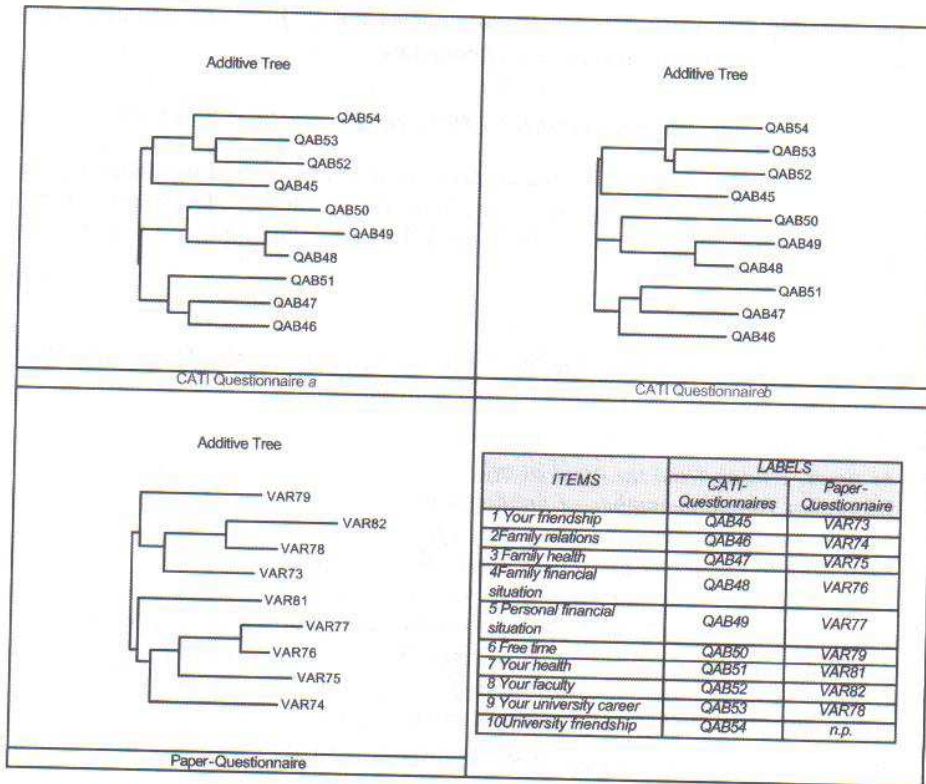
Figure 14.19. Tree Diagrams for Well-being Ambits.

more clearly; in particular, the items related with the first component join with items of the other two dimensions, which, on the other hand, can be clearly identified. The better appreciation of relations among variables, in terms of distances, allowed by additive tree analysis, highlights the potential insufficiency of Principal Component Analysis in the selection of variables for the construction of synthetic indicators purpose.

At this point, the registered differences between CATI outcomes and paper-questionnaire outcomes reveal a survey-effect rather than a scale-effect.

On the other hand, the outcomes obtained seem to suggest that the use of different scale ranges has no direct influence on the identification or construction of synthetic well-being indicators, but a partial impact related to reliability outcomes. The presence of a superimposition of both effects needs a follow-up analysis and evaluation.

Finally, the use of statistical multidimensional methods in selecting variables to construct synthetic indicators needs special attention. Even if all the methods can yield consistent and comparable results on the same data, each approach can produce

and reveal original images and representations of data. This suggests, and makes desirable, integration among different approaches.

14.3.4. *ScaleRreliability in Different Survey Conditions*

The measurement of self-esteem and motivation towards study dimensions by different scale approaches allows us to evaluate the influence of different survey conditions (different survey techniques and different type and range scales) on scale reliability.

Self-Esteem
The 10 items Rosemberg Scale was chosen to measure self-esteem. These Rosemberg Scale items were expressed in the following manner:

- *On the whole, I am satisfied with myself.*
- *At times, I think I am no good at all.*
- *I feel that I have a number of good qualities.*
- *I feel I do not have much to be proud of.*
- *I am able to do things as well as most other people.*
- *I certainly feel useless at times.*
- *I feel I'm a person of worth, at least on an equal plane with others.*
- *I wish I could have more respect for myself.*
- *I take a positive attitude toward myself.*
- *All in all, I am inclined to feel that I am a failure.*

We decided to define and adopt three different scales approaches:

- 4 labeled agreement levels, in paper-questionnaire (questions from 83 to 92)
- 5 points agreement rating scale, in *a* CATI-questionnaire (questions from 55 to 64)
- 7 points agreement rating scale, in *b* CATI-questionnaire (questions from 55 to 64)

Before observing the reliability of the Rosemberg scale under different survey conditions we can point out that, independently of survey conditions, we have observed the same kind of distribution in regards to each item.

Motivation Towards Study
In order to explore one particular aspect of student life, motivation towards study, related with other dimensions (students' performances, evaluations, and so on) we adopted a scale whose items were expressed in the following way:

- *I would go on with studying even if I would receive an offer of work of my liking.*
- *In the future I look forward to get a work consistent with my study.*
- *I believe I have the ideal characteristics to attend this degree.*
- *In the future I look forward to get a work which could be fulfilled.*

- At this moment study represents one of the most important thing for me.
- *My final preparation will be my main strength point.*
- *All exams in my curriculum are important for the aspiring career.*
- *I am up to any exam's level I prepare.*
- *Exam mark depends especially on me.*
- *I always have an interest in subjects I study.*

In the paper-questionnaire (questions from 61 to 70), we used five labeled levels: (5) strongly agree, (4) quite agree, (3) neither agree nor disagree, (2) quite disagree, and (1) strongly disagree.

In CATI survey (*a* and *b* questionnaires, questions from 33 to 42), we used the same scales, but eliminated the median-labeled level in order to verify the influence of the lack of the neutral position. This additionally allowed us to avoid the difficult presentation of a great number of labels in telephonic interviews.

Reliability Analysis of Self-esteem and Motivation towards Study Scales
Comparing Rosemberg scale items, we can observe that items of the *b* questionnaire generally show better performances than those of the *a* questionnaire and that paper-questionnaire items reach better performances than CATI-questionnaires items (Table 14.20). These outcomes are confirmed by total scale performance: the Rosemberg scale reaches a better internal consistency in paper-questionnaire than in other questionnaires. These outcomes allow us to conclude that whereas different results between CATI groups can be imputed to scale-effect, better paper-questionnaire performances can be explained by survey-effect.

Following the reliability analysis of motivation towards the study scale allows us to test the existence of survey-effect in the presence of the same scale type. Outcomes confirm the observed difference between paper-questionnaire and CATI-questionnaire (Table 14.21). In particular, it is very important to observe the low performance levels of CATI-questionnaire.

The low level of internal consistency for the CATI-group could be evaluated in

Table 14.20. Reliability Item Analysis of Rosemberg Scale for the Three Questionnaires.

	Paper-Questionnaire Labeled scores (4)				CATI-Questionnaire *a* Rating scales (1-5)				CATI-Questionnaire *b* Rating scales (1-7)			
	*	**	***	****	*	**	***	****	*	**	***	****
1	0.70	0.47	0.62	0.84	0.52	0.44	0.39	0.71	0.61	0.70	0.50	0.72
2	0.70	0.67	0.59	0.84	0.63	0.72	0.47	0.67	0.66	1.13	0.50	0.72
3	0.65	0.45	0.56	0.84	0.43	0.32	0.30	0.72	0.47	0.46	0.37	0.74
4	0.71	0.50	0.64	0.84	0.45	0.38	0.31	0.72	0.52	0.55	0.41	0.74
5	0.52	0.43	0.40	0.85	0.33	0.29	0.17	0.74	0.40	0.49	0.25	0.75
6	0.77	0.73	0.68	0.83	0.66	0.73	0.52	0.69	0.63	0.95	0.49	0.72
7	0.54	0.41	0.43	0.85	0.48	0.38	0.35	0.72	0.48	0.54	0.35	0.74
8	0.59	0.60	0.45	0.85	0.67	0.99	0.46	0.71	0.62	1.37	0.39	0.75
9	0.73	0.60	0.64	0.83	0.65	0.68	0.51	0.69	0.68	0.95	0.56	0.71
10	0.72	0.59	0.64	0.83	0.58	0.40	0.49	0.70	0.59	0.57	0.50	0.73
alpha	0.85				0.73				0.75			

* item-total r
** item-reliability index
*** item-total r excluding the item
**** alpha value excluding the item

Table 14.21. Reliability Item Analysis of Motivation towards Study Scale for the Three Questionnaires.

	Paper-Questionnaire				CATI-Questionnaires			
	Labeled scores (5)				Labeled scores (4)			
	*	**	***	****	*	**	***	****
1	0.72	0.95	0.60	0.79	0.43	0.45	0.18	0.55
2	0.67	0.68	0.58	0.79	0.45	0.33	0.29	0.52
3	0.67	0.66	0.58	0.79	0.45	0.33	0.29	0.52
4	0.58	0.59	0.47	0.80	0.38	0.21	0.25	0.53
5	0.59	0.62	0.47	0.80	0.53	0.52	0.32	0.51
6	0.73	0.80	0.64	0.79	0.51	0.48	0.31	0.51
7	0.54	0.64	0.40	0.81	0.51	0.52	0.28	0.52
8	0.57	0.62	0.46	0.80	0.39	0.36	0.18	0.55
9	0.55	0.74	0.40	0.81	0.36	0.36	0.12	0.57
10	0.57	0.65	0.44	0.79	0.49	0.46	0.28	0.52
alpha	0.82				0.56			

* item-total r   *** item-total r excluding the item
** item-reliability index   **** alpha value excluding the item

the context of the difficulty of use of the verbal scales in telephonic situations (survey-effect). It could thus be interesting to verify the possible superimposition of the scale-effect on survey-effect through supplementary studies.

14.4. CONCLUSIONS

The goal of the study essentially concerned methodological aspects involved in surveys. The presented analyses and outcomes show the importance of the choice of the scale under different survey methods. In fact, the outcomes show the different influence of survey-effect and scale-effect on quality of data.

A scale-range-effect exists independently from survey-effect. In particular, we can say that it is possible to reach better individual measurement and evaluation by wide-range scales. Longer scales reveal a better discriminate capacity than shorter ones (scale-effect). On the other hand, graphical scales suggest a better discriminate capacity than rating scales aside from scale range (survey-effect).

The formation of satisfaction dimensions and the aggregation processes of single-items in the construction of synthetic indicators seem not to be directly influenced by scale-range.

We also observed partial scale-effect and survey-effect in each scale item in regards to reliability and a combined scale- and survey-effect in whole multi-item scale reliability. In particular, labeled-scales do not seem to fit telephonic questionnaires; in fact, labeled-scale items in paper-questionnaire show better reliability performances than labeled-scale items in CATI-questionnaires.

Further study, such as a WEB-survey combining paper and CATI-questionnaires characteristics, are needed in order to evaluate the superimposition of both effects on reliability levels.

The reference-scale effect must be considered as it relates with item definition and cultural factors. In our experience, this effect allowed us to identify a particular outcome in judgments expressed by positive and negative references. This cultural-effect needs to be tested further in comparison with other cultural contexts.

These tests could be of great interest and could introduce important observations and considerations about the use of international comparison analyses.

Outcomes confirm our hypothesis about the importance of the measurement instrument under different survey conditions. This is especially true for the tele- phonic approach, frequently considered, erroneously, a simple survey method, for which the definition of the measurement instrument is less important. Unfortunately, in many cases, CATI-questionnaires are derived from a simple and direct adapta- tion of a paper-questionnaire, i.e. following the same criterion in expression and wording of the questions.

In our opinion, the theme is very delicate and complex since it also involves cultural components. In such cases, the generalization of individual experiences to other cultural contexts may not be allowed, in trying to provide a truly accurate data analysis.

14.5. NOTES

[1]   Kurtosis values are not considered crucial to identify bipolar adjectives.

[2]   Notice that *face scale* expressions show an inverted direction as regards the rating scales of CATI- questionnaires. In order to compare distributions, we reversed the *face-scale* codes.

[3]   We cannot compare the three groups since this item was not presented in the paper-questionnaire.

[4]   A follow-up analysis will define component scores by using factor scores.

[5]   As we know the Additive Trees Clustering Method, unlike Hierarchical Clustering methods, uses tree branch length to represent distances between objects. Objects within a cluster can thus be compared by focusing on the horizontal distance along the branches connecting them.

14.6 REFERENCES

Alreck, P. L. and Settle, R. B. (1985). *The Survey Research Handbook.* Homewood Illinois: Richard D.Irwing Inc.

Andrews, F. M. and McKennel, A. C. (1980). Measures of self-reported Well-Being: their affective, cognitive, and other components. *Social Indicators Research*, **8**.

Arcuri, L. and Flores D'Arcais, G. B. (1974). *La misura degli atteggiamenti.* Firenze: Martello – Giunti.

Baker, T. L. (1994, 2nd ed.). *Doing Social Research.* New York-London: McGraw-Hill.

Bolasco, S. (1999). *Analisi multidimensionali dei dati. Metodi, strategie e criteri di interpretazione.* Roma: Carocci.

Carmines, E. C. and Zeller, R. A. (1992). *Reliability and Validity Assessment.* Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-017. Newbury Park, CA: Sage.

Converse, J. M. and Presser, S. (1991). *Survey Questions: Handcrafting the Standardized Questionnaire.* Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-063. Newbury Park, CA: Sage.

Dautriat, H. (1988, 6th ital. ed.). *Il questionario.* Milano: Franco Angeli Editore.

Del Vecchio, F. (1995). *Scale di misura e indicatori sociali.* Bari: Caccucci.

De Sandre, P. (1972). *Classificazione e misura nella ricerca sociale.* Padova: CLEUP.

De Vellis, R. (1991). *Scale devolopment. Theory and Applications.* Applied Social Research Methods Series, vol. 26. London: SAGE.

Edward, A. (1957). *Techniques of Attitude Scale Construction.* Englewood Cliffs, NJ: Prentice-Hall.

Edward, W. and Newman J. R. (1982). *Multiattribute Evalutation.* Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-026. Newbury Park, CA: Sage.

Flament, C. (1976). *L'analyse booléenne de questionnaire.* Paris-La Haye: Mouton Editeur.

Fordyce, M. W. (1988). A review of research on the happiness measures: a sixty seconds index of happiness and mental health. *Social Indicators Research,* 4.

Ghiselli, E. E. (1964). *Theory of Psychological Measurement.* New York-London: McGraw-Hill.

Gilbert, N. (1993). *Researching Social Life.* London: Sage.

Guilford, J. P. (1936). *Psychometric Methods.* New York-London: McGraw-Hill.

Jacoby, W. G. (1991). *Data Theory and Dimensional Analysis.* Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-078. Newbury Park, CA: Sage.

Larsen, E. R., Diener, J., and Emmons, R. A. (1985). An Evaluation of Subjective Well-Being Measures. *Social Indicators Research,* 17(4).

Lyubomirsky, S. and Lepper, H. S. (1999). A Measure of Subjective Happiness: Preliminary Reliability and Construct Validation. *Social Indicators Research,* 46(2).

Maggino, F. (2000, 2nd ed.). *Il questionario. Aspetti metodologici, informatici e statistici.* Firenze: C.E.T.

Maggino, F. (1999). Metodi di validazione di strumenti soggettivi. In E. Aureli, F. Buratto, L. Carli Sardi, A. Franci, A. Ponti Sgargi, and S. Schifini D'Andrea (eds.), *Contesti di qualità della vita* (pp. 185–213). Milano: F. Angeli.

McIver, J. P., and Carmines, E. G. (1979). *Unidimensional Scaling.* Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-024. Newbury Park, CA: Sage.

Nunnally, J. C. (1978). *Psychometric Theory.* New York-London: McGraw-Hill

Oppenheim, A. N. (1966). *Questionnaire Design and Attitude Measurement.* London: Heinemann.

Osgood, C. E., Suci G. J., and Tannenbaum, P. H. (1957). *The Measurement of Meaning.* Urbana: University of Illinois Press.

Osterlind, S. J. (1983). *Test Item Bias.* Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-030. Newbury Park, CA: Sage.

Saris, W. E. (1990). *Computer-Assisted Interviewing.* Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-080. Newbury Park, CA: Sage.

Schifini, D'Andrea S. (1999). Qualità della vita: misure, teorie e modelli. In E. Aureli, F. Buratto, L. Carli Sardi, A. Franci, A. Ponti Sgargi, and S. Schifini D'Andrea (ed.), *Contesti di qualità della vita* (pp. 15–66). Milano: F. Angeli.

Schifini D'Andrea, S. and Maggino, F. (1999). Qualità della vita universitaria: validazione di strumenti soggettivi. In E. Aureli, F. Buratto, L. Carli Sardi, A. Franci, A. Ponti Sgargi, and S. Schifini D'Andrea (eds.), *Contesti di qualità della vita* (pp. 155–184). Milano: F. Angeli.

Spector, P. E. (1990). *Research Designs.* Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-023. Newbury Park, CA: Sage.

Spector, P. E. (1992). *Summated Rating Scale Construction. An Introduction.* Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-082. Newbury Park, CA: Sage.

Sullivan, J. L. and Feldman S. (1981). *Multiple Indicators. An Introduction.* Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-015. Newbury Park: CA:Sage.

Tesi, G. (1983). *Note per la ricerca mediante questionario.* Firenze: Università degli Studi.

Torgerson, W. S. (1958). *Theory and Methods of Scaling.* New York, London, Sydney: John Wiley & Sons, Inc.

Traub, R. E. (1994). *Reliability for the Social Sciences – Theory and Applications.* Measurement Methods for the Social Sciences series, vol. 3. London: SAGE Publications.

Zumbo, B. D. (ed.) (1999). Validity Theory and the Methods used in Validation: Perspectives from Social and Behavioural Sciences. *Social Indicators Research,* 45(1–3).

14.7. AFFILIATIONS

Filomena Maggino, Università di Firenze – Italy

Silvana Schifini D'Andrea, Università di Firenze – Italy