The 2th Conference of the

European Survey Research Association

25 – 29 June 2007

Prague, Czech Republic

Subjective indicators:

a model for their construction and analysis in the perspective of data condensation

Filomena Maggino

Università degli Studi di Firenze, Italy

filomena.maggino@unifi.it

In order to meet the states of the scientific measurement, the measurement and the analysis of subjective data require the definition of a composite design consisting of two aspects:

- theoretic
- methodological

Theoretic aspects

1. the definition of the concept of **“subjective data”**,
2. the identification of a **reference-theory of measurement** that defines the theoretical characteristics that make the measurement “scientific”.

Theoretic aspects

1. "Subjective characteristics"

Traditionally we refer mainly to three content areas (Nunnally, 1978):

- Abilities
- Personality traits
    - o social traits
    - o motives
    - o personal conceptions
    - o adjustments
    - o personality dynamics
- Sentiments
    - o interests
    - o values
    - o attitudes
        - ✓ cognitive component (beliefs)
        - ✓ affective component (feelings, perceptions, …)
        - ✓ behavioral components (intentions and actions).

Theoretic aspects

2. Reference-theory of measurement

that defines the theoretical characteristics that make the measurement "scientific"; in other words, the reference-theory, by defining the concept of *measurement error*, allows to identify the models finalized to test:

objectivity, that is the capacity of a procedure to measure without alteration due to external factors and to be free from effects due to the observer;

precision, measured by controlling the coherence of the model of measurement → **reliability**

accuracy that is the capacity of the procedure to measure what we intend to measure (content) → **validity**

A procedure of measurement that meets these requirements not only gains a scientific relevance but can also be standardized.

Methodological aspects

Logical processes (modeling) concern

1. the measurement of subjective characteristics (in order to create subjective data),
2. the analysis of subjective data (in order to transform data into indicators).

Both logical processes (modeling) are based upon the definition of the

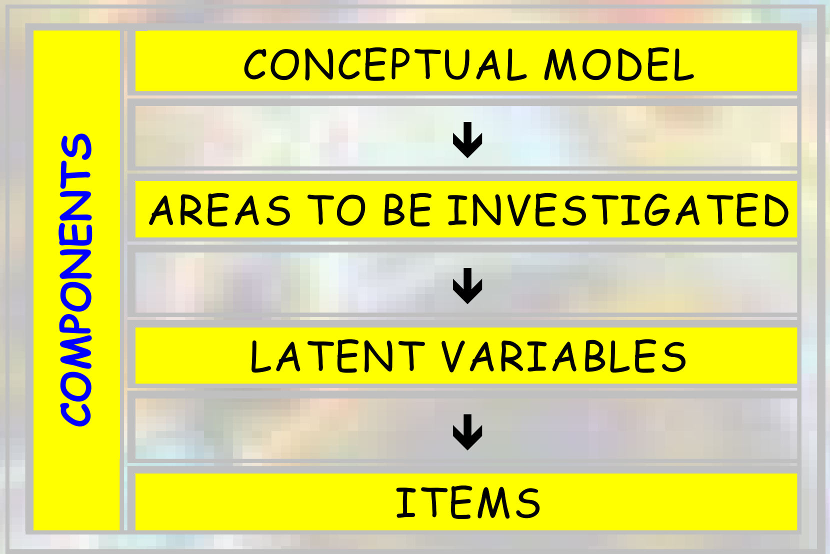
hierarchical design

The hierarchical design

The hierarchical design is defined through “consecutive components” (from definition of the conceptual model to definition of the single/elementary indicators).

In the hierarchical design, each component is defined and finds its meaning in the ambit of the preceding one.

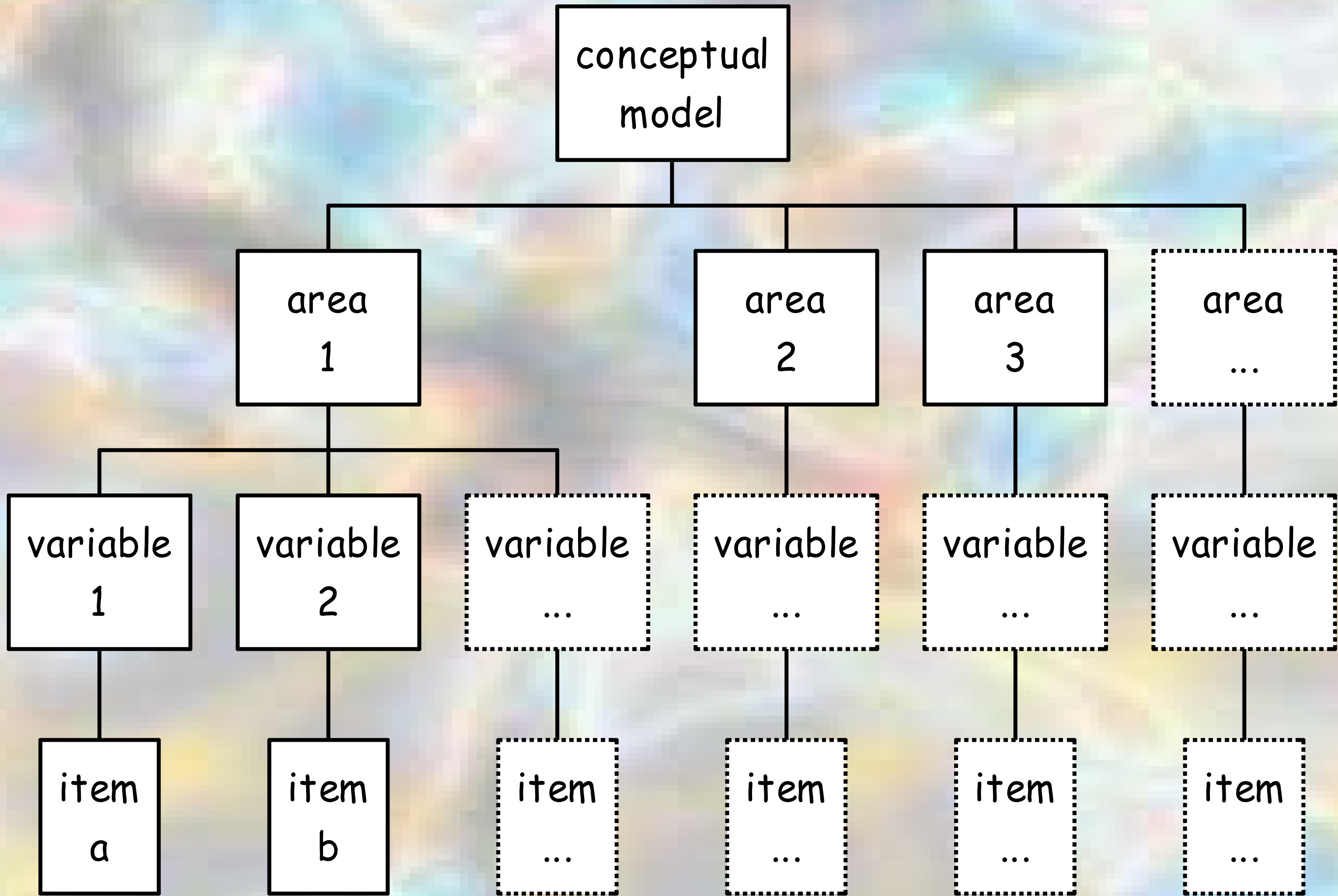
This allows to generating subjective data that are correct, consistent interpretable, and complex with reference to the complexity of the hierarchical design’s structure.



The hierarchical design

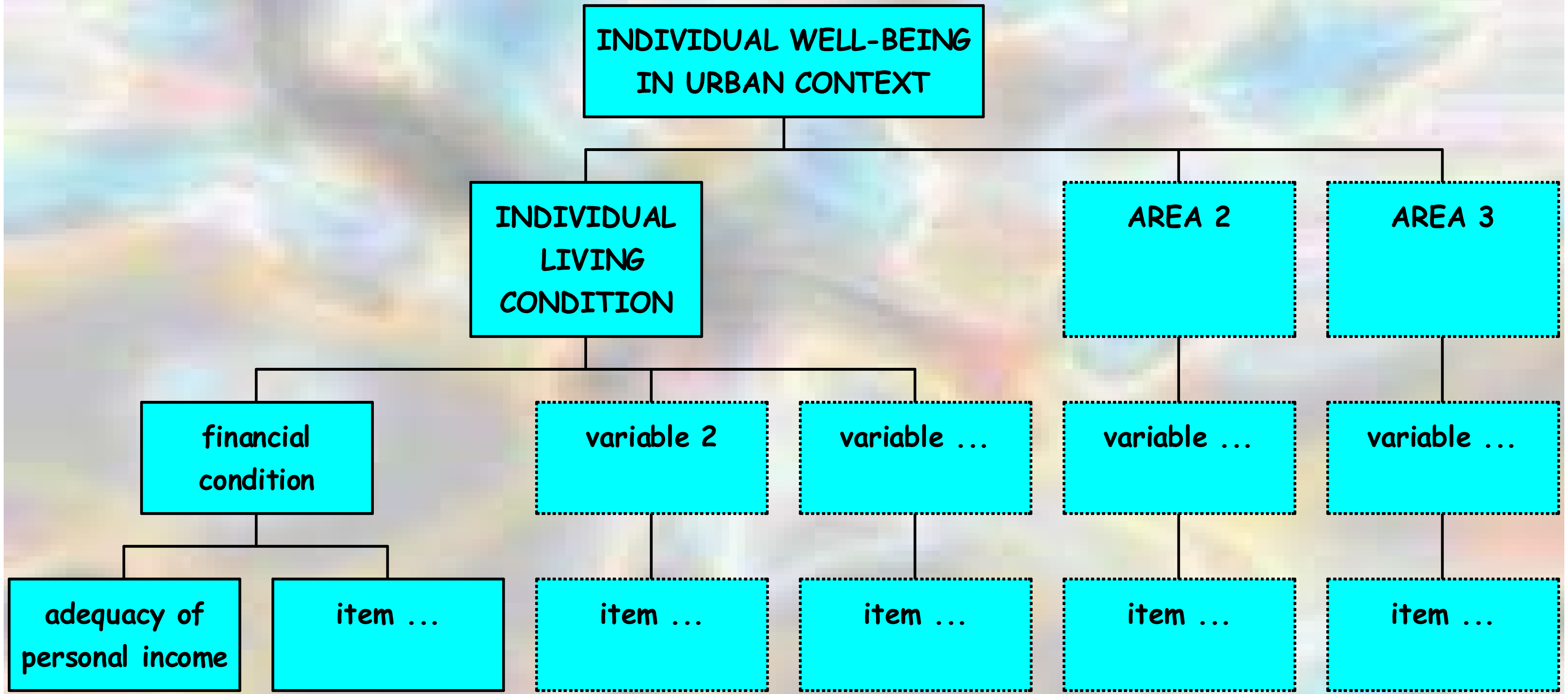
component		Question to which the component gives an answer	definition of the component
1	**conceptual model**	Which phenomena have to be studied	The conceptual model defines the **phenomena** to be studied and the domains and the general aspects that characterize the phenomena
↓			
2	**areas to be investigated**	Which aspects define the phenomenon	Each area represents each **aspect** that characterizes and defines the theoretical model
↓			
3	**latent variables**	Which elements have to be observed	Each variable represents each **element** that has to be observed in order to define the corresponding area. The variable is named **latent** since is not observable directly
↓			
4	**items**	In which way each element has to be measured	Each item represents **what is actually measured** for each variable and is defined by appropriate techniques and by a system that allows to evaluating and interpreting the observed value.

The hierarchical design



The hierarchical design

Example



The hierarchical design

Single indicator approach

According to a simple and weak strategy, each latent variable is defined by a single item.

This strategy, very often applied because of its thrifty and functional capacity, requires the adoption of robust assumptions.

The adoption of single items presents a risk because can produce problems of precision and accuracy.

The hierarchical design

Multiple indicators approach

The presence of complex latent variables requires the definition of several elementary indicators by adopting the *multiple indicators approach* that considers the multiple indicators as *multiple measures*.

Each elementary indicator corresponds to one particular aspect of the latent variable.

The hierarchical design

The definition of the hierarchical design can be completed through the identification of the relations between:

- **latent variables**: these relations define the **structural model** and are hypothesized and identified in the ambit of the conceptual model;
- **latent variables and corresponding indicators (items):** these relations define the **model of measurement**, whose inspection allows to evaluate the reliability of the measurement;
- **items**: in this perspective, the defined relations can identify different states:
  - items relate to the same latent variable that means that contribute to the definition of same variable; in these case, the items are called **constitutive** and can be condensed;
  - items relate to different latent variables; in this case, the items are called **concomitant.**

The hierarchical design

The definition of the hierarchical design allows to identify

-the reference population,

-the sampling design

-the methodological approach to data collection.

1. MEASUREMENT OF SUBJECTIVE CHARACTERISTICS

Methodological aspects

1. Measurement of subjective characteristics

that requires the identification of the following models:

A. model for the construction of subjective data

B. model for the assignment of data values

A. Model for the construction of subjective data

A. Model for the construction of subjective data

A model is required in order to obtain from observation
an interpretable and analyzable information.

In other words, this model has to allow to transform

observation
(the collected information)

in

datum
(the information that can be analyzed)

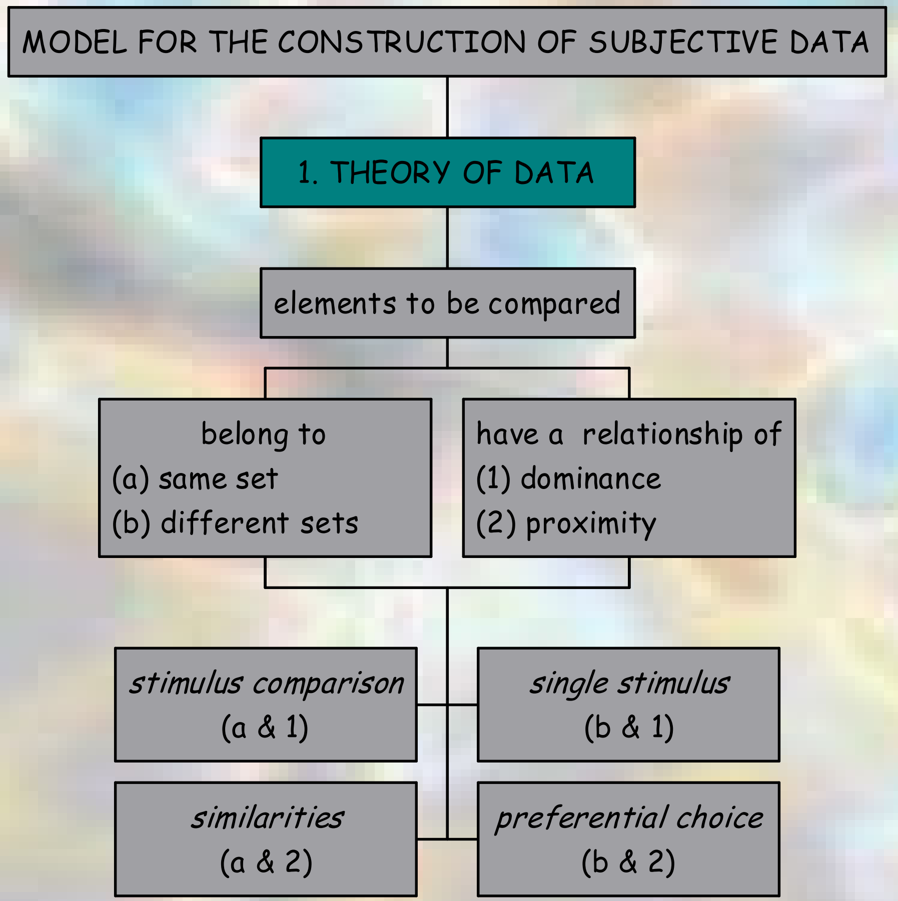
Therefore, data consist of portions of information extracted according to a reference model; in this sense, data represent a researcher’s construction and interpretation.

A. Model for the construction of subjective data

In order to obtain subjective data from observation,
we need to define:

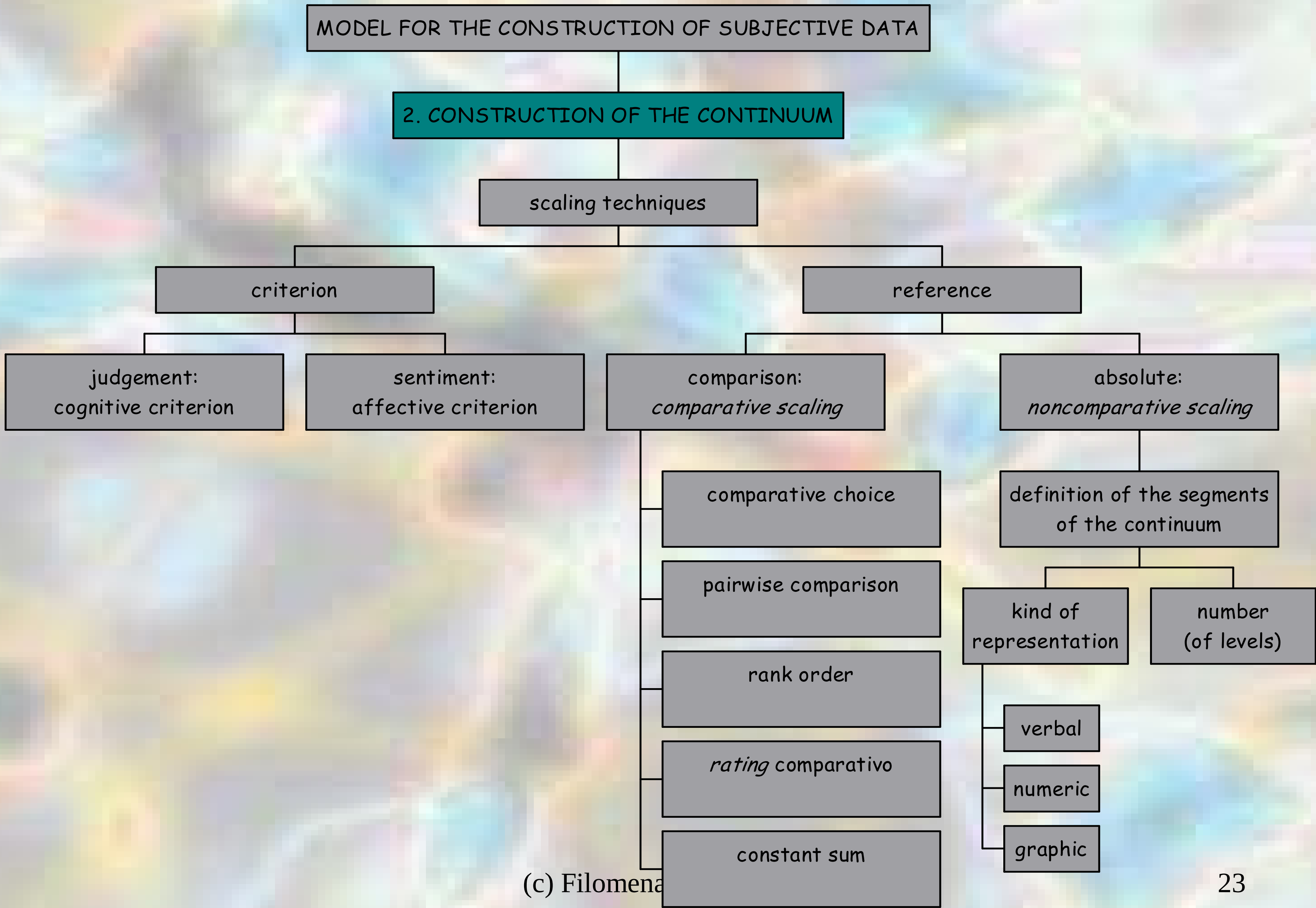
1. the **nature of data**, referring to an interpretative theory (**theory of data**)
2. a procedure finalized to the **definition and identification of the continuum** on which each individual case can be placed (**scaling techniques**), with reference to the observed characteristic
3. a system allowing for the **organization of data** (**data matrix structure**).

A. Model for the construction of subjective data

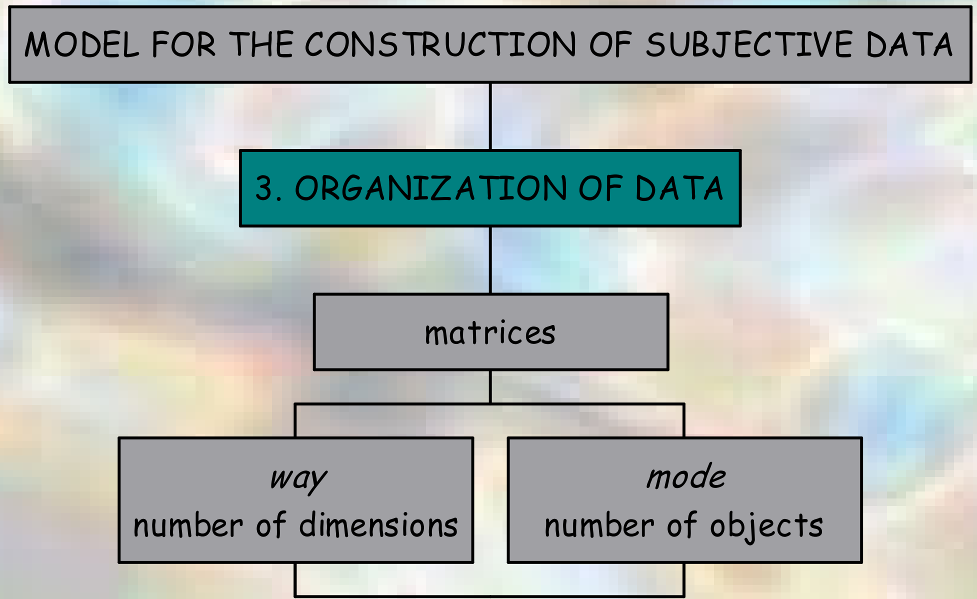


Coombs C.H. (1964) *A theory of Data*, Ann Arbor, MI: Mathesis.

A. Model for the construction of subjective data



A. Model for the construction of subjective data



B. Model for the assignment of data values

B. Model for the assignment of data values

This model allows to assign a value that makes the constructed data interpretable and that may be treated in operative terms.

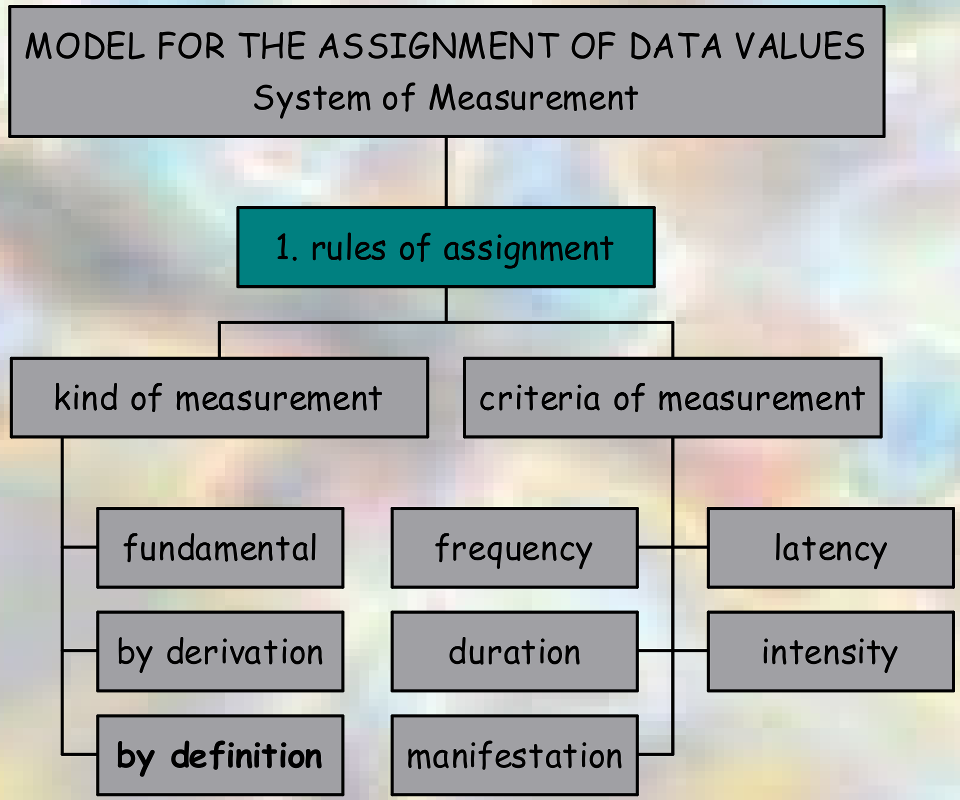
For this purpose we need to define the rules that clarify the procedure of correspondence and of assignment of a symbol to each identified level.

B. Model for the assignment of data values

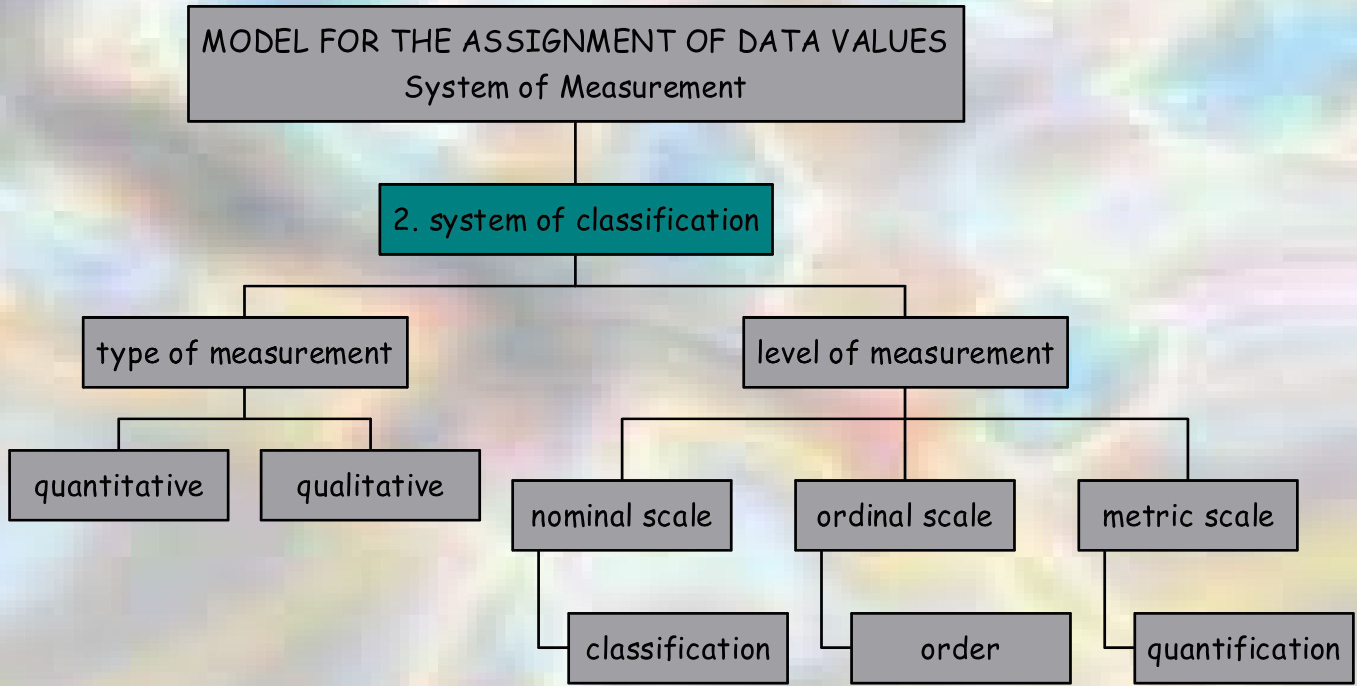
This requires the definition of a system of measurement that presents

1. **rules** that allow to assign numbers/symbols in a standard and uniform procedure (**kind and criteria of measurement**),
2. **a "system of classification"** that allows to assign to each case the status with reference to the measured characteristic (**type and level of measurement**).

B. Model for the assignment of data values



B. Model for the assignment of data values



2. ANALYSIS OF SUBJECTIVE DATA

Methodological aspects

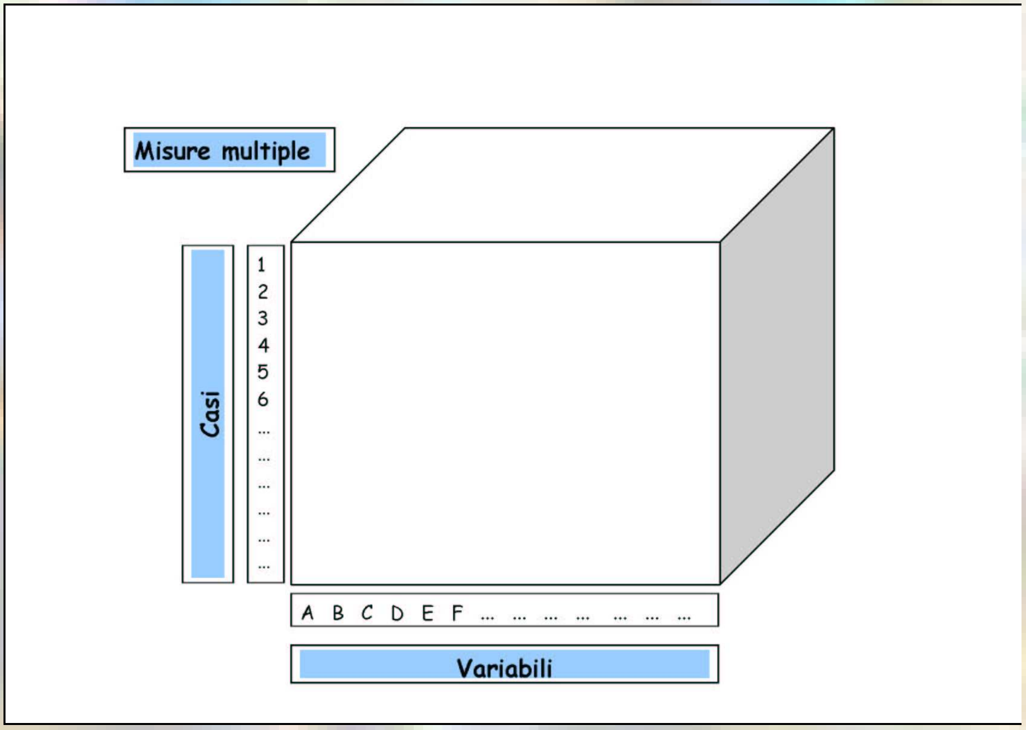
2. Analysis of subjective data

The model finalized to obtain subjective data presented above produces a complex data structure with reference to:

- *variables* → according to the hierarchical design, several variables are defined,
- *multiple measures* → identified for each variable (except those measured by single indicators),
- observed *cases*.

Methodological aspects

2. Analysis of subjective data



logical structure of data

Methodological aspects

2. Analysis of subjective data

In order to reduce this complexity, models have to be defined allowing to manage and to reduce the complexity of the measured data through a

condensing strategy

according to two different perspectives:

- *multiple measures* → **from elementary indicators to synthetic indicator** (A)
- *individual cases* → **from individual-points to grouping-point** (B)

Attempts exist in order to reduce the third dimension (*variables*).

They show some problems in both methodological perspective and interpretative sense.

A. From elementary indicators to synthetic indicator

A. From elementary indicators to synthetic indicator

The elementary indicators (multiple measures) are condensed in new synthetic values (synthetic indicators) in order to re-establish the unity of the described concept by running through the hierarchical design backwards:

latent variable

↓

elementary indicators

↓

aggregation

↓

synthetic indicator

the reducing procedure requires the implementation of a condensing model that is able to manage the data complexity

↓

scaling model

Scaling models allow to condense elementary indicators, considered multiple measures, according to the homogeneity criterion.

A. From elementary indicators to synthetic indicator

scaling model

Scaling model			Dimensionality	Nature of data	Scaling technique
Additive	Uni-dimensional		Uni	Single-stimulus	Not-comparative
	Multidimensional		Multi	Single-stimulus	Not-comparative
Cumulative	*Thurstone scale (differential scale)*		Uni	Stimulus comparison	Comparative (pair comparison or rank-order)
	Q methodology		Uni	Stimulus comparison	Comparative (rank-order or comparative rating)
	Deterministic	Guttman	Uni	Single-stimulus	Not-comparative
		Multidimensional Scalogram Analysis (MSA)	Bi		
		Partial Ordered Scalogram Analysis (POSA)	Bi		
	Probabilistic	Monotone (one or more parameters)		Single-stimulus	Not-comparative
Perceptual Mapping	*Multidimensional scaling*		Multi	Similarities	Comparative (pair comparison)
	Unfolding		Uni & Multi	Preferential choice	Comparative
Conjoint model			Multi	Preferential choice	Comparative (rank-order)

A. From elementary indicators to synthetic indicator

scaling model

Scaling model			Criterion for testing the model	Standard of measurement: final synthetic score assigned to
Additive	Uni-dimensional		Internal consistency	Cases
	Multidimensional		Dimensionality of the items	Casies
Cumulative	*Thurstone scale (differential scale)*		Metrics between items	Items
	Q methodology			Items
	Deterministic	Guttman	Scalogram analysis: reproducibility, scalability and ability to predict	Cases and items
		Multidimensional Scalogram Analysis (MSA)	Regionality and contiguity	Cases and items
		Partial Ordered Scalogram Analysis (POSA)	Correct representation	Cases and items
	Probabilistic	Monotone (one or more parameters)	• parameters estimation (maximum likelihood) • goodness of fit (*misfit* and residuals analysis)	Cases and items (without condensation)
Perceptual Mapping	*Multidimensional scaling*		Goodness of fit of distances to proximities (stress, alienation)	Items
	Unfolding		Goodness of fit of distances to ordinal preferences	Cases and items
Conjoint model			Goodness of fit of the model (part-worth) to the ranking	Items

B. From individual-points to grouping-point

B. From individual-points to grouping-point

The individual values (individual-points) are condensed in new synthetic values assigned to significant meaningful units (groupings) identified according to different perspectives (typological, geographical, etc.); in this perspective, the reducing procedure requires the:

1. identification of significant grouping (***condensing criteria***),
2. definition of weights to be assigned to each individual cases, whose values will be condensed into a grouping-point (***weighting criteria***),
3. adoption of technique allowing the aggregation of individual values - belonging to each grouping - in one or more representative values (grouping-point) that can be assigned to the grouping (***aggregating-over-individuals techniques***).

B. From individual-points to grouping-point

1. condensing criteria

identifying the significant grouping

Two condensing criteria can be defined:

A. **Homogeneity**: the values are condensed if the individual cases that turn out to be homogeneous with reference to the indicators of interest;
   ➔ This allows the comparison between the identified groups (**typologies**) by other contextual and background variables ;

B. **Functionality**: the values are condensed if the individual cases belong to groupings that do not require any analytic procedure for their identification;
   ➔ This allows for comparisons and differential evaluations with reference to the defined indicators:
   - **groups** (social, generational, etc.);
   - **areas** (geographical, political, administrative);
   - **time-periods** (years, decades, etc.).

B. From individual-points to grouping-point

2. **weighting criteria**

defining the weights that have to be assigned to each of the individual cases, whose values will be condensed into a grouping-point

The assignment of a certain weight to each individual case in order to condense the individual scores in one grouping-point occurs particularly when data come from sample-surveys.

For this reason the matter is dealt directly through statistical approaches related to inference methods and sampling techniques.

B. From individual-points to grouping-point

3. aggregating-over-individuals techniques

adopting a technique allowing the aggregation of the individual values of each grouping in one or more representative values (grouping-point) that can be assigned to the grouping.

The aggregation of individual scores is a well-known issue in many scientific fields, like economics and informatics, where it is dealt with the application of particular analytic approaches (like probabilistic aggregation analysis).

In econometrical fields, particular empirical methodologies have been developed, allowing the explanation of systematic individual differences (*compositional heterogeneity*) that can have important consequences in interpreting aggregated values (Stoker, 1993).

B. From individual-points to grouping-point

3. **aggregating-over-individuals techniques**

adopting a technique allowing the aggregation of the individual values of each grouping in one or more representative values (grouping-point) that can be assigned to the grouping.

Grouping identified through **homogeneity criterion**

aggregation → few statistical problems:

the homogeneity, in fact, allows to condense by applying simple statistical averaging techniques, univariate (mean, median) or multivariate (centroid).

B. From individual-points to grouping-point

3. **aggregating-over-individuals techniques**

adopting a technique allowing the aggregation of the individual values of each grouping in one or more representative values (grouping-point) that can be assigned to the grouping.

Grouping identified through **heterogeneity criterion**

aggregation → problematic:

the application of the traditional statistical averaging techniques does not allow to pointing out the distributional characteristics of each grouping.

Consequently, no comparison between grouping is permitted.

Concerning this, attempts exist oriented to weight average values by different criteria (Kalmijn, 2005; Veenhoven, 2005).

B. From individual-points to grouping-point

3. aggregating-over-individuals techniques

adopting a technique allowing the aggregation of the individual values of each grouping in one or more representative values (grouping-point) that can be assigned to the grouping

The interpretation of the information obtained through the procedure of segmentation in not easy.

Assigning a certain level of subjective satisfaction to a certain grouping (i.e. a geographical area) can lead to attribute that value uniformly to the whole members of the grouping (stereotype) even if the it is not necessarily so (*ecological fallacy*).

C. Beyond condensation: the creation of composite indicators

C. Beyond condensation: the creation of composite indicators

A further development in the treatment of subjective indicators is represented by the possibility to create composite indicators, yielded by the aggregation of elementary and/or synthetic indicators (objective and subjective).

By definition, a composite indicator aggregates indicators that conceptually refer to different latent variables (*heterogeneity criterion*). A distinction can be made between:

- <u>composite indicator</u>: the aggregation is obtained by indicators (elementary and/or synthetic) that are related but not necessarily in a statistic sense
- <u>comprehensive indicator</u>: the composite indicator is constructed with the intention to be exhaustive with reference to a certain QOL construct or reality

C. Beyond condensation: the creation of composite indicators

latent variables / multidimensional latent variable	
↓	↓
single and/or synthetic indicators	single and/or synthetic indicators
↓	↓
aggregation	aggregation
↓	↓
composite indicator	**comprehensive indicator**

With reference to the construction of subjective indicators, this distinction is from the

-theoretical point of view → acceptable

-applicative point of view → not always valid and applicable.

C. Beyond condensation: the creation of composite indicators

latent variables / multidimensional latent variable	
↓	↓
single and/or synthetic indicators	single and/or synthetic indicators
↓	↓
aggregation	**aggregation**
↓	↓
composite indicator	**comprehensive indicator**

↓

Serious problems → in constructing and interpretation of data

The condensation process involving subjective data can be conducted to a limited level (Maggino, 2007).

C. Beyond condensation: the creation of composite indicators

The creation of composite indicators requires the adoption and the application of particular technical and analytical approaches (Nardo, 2005; Sharpe, 2004; Tarantola, 2000) related to data management; the approaches, finalized to obtain composite indicators not only meaningful but also interpretable, allow to

1. defining the importance of each indicator to be condensed (***weighting criteria***),
2. identifying the technique for condensing the indicators values into the composite indicator (***aggregating-over-indicators techniques***),
3. assessing the robustness of the composite indicator in terms of capacity to produce correct and stable measures (***uncertainty analysis***, ***sensitivity analysis***),
4. assessing the discriminant capacity of the composite indicator (***ascertainment of selectivity*** and ***identification of cut-point or cut-off values***).

C. Beyond condensation: the creation of composite indicators

1. **weighting criteria**

assign weights to each elementary indicator

Equal Weighting ← → Different Weighting

Both approaches have pros and cons.

The choice depends on theoretical and methodological concerns.

C. Beyond condensation: the creation of composite indicators

1. **weighting criteria**

assign weights to each elementary indicator

Different Weighting → defined by

1. **statistical methods**:
   a. *Correlation*,
   b. *Principal Component Analysis*,
   c. *Data Envelopment Analysis*,
   d. *Unobserved Components Models*.
2. **multi-attribute models**:
   a. *Multi-Attribute Decision Making* (in particular, *Analytic Hierarchy Processes*),
   b. *Multi-Attribute Compositional Model* (in particular, *Conjoint Analysis*).
3. **expertise methods** (in particular, *Budget Allocation* (BAL).

C. Beyond condensation: the creation of composite indicators

2. aggregating-over-items techniques

aggregate the elementary indicators in order to define the new synthetic indicator

The different techniques present different technical characteristics concerning the admissibility of:

a. compensability among the elementary indicators to be aggregated,
b. comparability among elementary indicators (in terms of directionality),
c. homogeneity of the levels of measurement of the elementary indicators.

		Aggregating approaches					
		Linear aggregation			Geometrical aggregation		Non-compensatory aggregation
		Classical additive		Cumulative			
Assumptions	Dimensionality (relationships between items)	Uni	Independency	Uni	Uni	Independency	Multi
	Relationship "elementary indicators – latent variable"	Monotonic		Differential relationship	Monotonic		
	Compensation among items	Admitted		Not admitted (graduality, scalability)	Admitted		Not admitted
	Homogeneity of scaling techniques	Requested		Requested	Requested		Not requested

C. Beyond condensation: the creation of composite indicators

3. **verify the robustness** of the obtained synthetic indicator

Assessing the robustness allows to evaluate the role and the consequences of the subjectivity of the choices made as regards:

- the model to estimate the measurement error;
- the procedure for selecting the elementary indicators;
- the procedure of data management (missing data imputation, data standardization and normalization, etc.);
- the criterion for weight assignment;
- the used aggregation technique.

Nardo M., M. Saisana, A. Saltelli and S. Tarantola (EC/JRC), A. Hoffman and E. Giovannini (OECD) (2005) *Handbook on Constructing Composite Indicators: Methodology and Userguide*, OECD, Statistics Working Paper.

C. Beyond condensation: the creation of composite indicators

3. verify the robustness of the obtained synthetic indicator

This procedure, which can be included in the wider field of the *what-if analysis*, is conducted through two stages; each stage corresponds to a different methodology of analysis:

- ***uncertainty analysis*** → analyzes how much the synthetic indicator depends on the information that constitutes it.
  - → identification of different scenarios for each individual case;
  - → each scenario corresponds to a certain combination of choices that produces a certain synthetic value;
- ***sensitivity analysis***: → evaluates the contribution of each identified source of uncertainty by decomposing the total variance of the obtained synthetic score.

C. Beyond condensation: the creation of composite indicators

4. **verify the capacity to discriminate** of the obtained synthetic indicator

Assessing the discriminant capacity of the synthetic indicator requires exploring its capacity in:

- discriminating between cases and/or groups (traditional approaches of statistical hypothesis testing),
- distributing all the cases without any concentration of individual scores in few segments of the continuum (some coefficients were defined),
- showing values that are interpretable in terms of selectivity through the identification of particular values or reference scores
  - → **cut-point** (continuous data)
  - → **cut-off** (discrete data).