Chapter 3
University League Tables

Methodological Options for Ranking Systems: Censis Approach and Alternatives

L. Bernardi, P. Bolzonello, and A. Tuzzi

3.1 Introduction

Since 2000, the Italian Censis research institute has compiled, on behalf of *La Repubblica* newspaper, the *Grande Guida all'Università*, a report which ranks Italian universities and faculties according to their quality. With the 2008 publication, devoted to students enrolled on degree courses in 2008–2009, the *Guida* has gone into its ninth edition.

For the administrators and practitioners of the Italian university system it may have been embarrassing to find themselves appraised and classified (even with unflattering rankings) in a competition in which, at least at the beginning, they did not know they were participating. All the more so if the league table was drawn up by a private organization assuming "civic responsibility" to inform the public about the work of the university system, and which was commissioned by a newspaper, which might therefore be more interested in sensationalism than in encouraging virtuous behaviour.

But how convincing and reliable are the general design, criteria, data sources, indicators and rules used to construct the league tables? And what are the possible reactions of the universities and faculties? Attack or defence, rejection or acknowledgement, acceptance or a decision to construct an alternative ranking system?

This study examined the contents and the methods of the Censis report and assessed possible alternatives. It explored the difficulties in achieving a reliable ranking system and sought ways to refine the Censis model. After a brief description of the Censis model, the discussion focuses on "evaluating" and "measuring", and

L. Bernardi (✉)
Dipartimento di Scienze Statistiche, Università di Padova, Padova, Italy
e-mail: bernardi@stat.unipd.it

In the Italian version the title "Arbitro, c'è rigore?" played upon words and with the meaning of the noun "arbitro" (referee and arbiter) and "rigore" (severity and penalty): "Arbiter, is there any penalty?"

M. Attanasio, V. Capursi (eds.), *Statistical Methods for the Evaluation of University Systems*, Contributions to Statistics, DOI 10.1007/978-3-7908-2375-2_3,

identified indicators as the means to produce objective, appropriate, and comparable measurement.

Data furnished by Censis on 27 Faculties of Political Science for 2006 were used to construct alternative ranking tables by employing a selection of current methods of normalization, aggregation and weighting. The Censis league table was then reconstructed on the basis of the *Note metodologiche* (methodological notes) attached to the *Guida* [10]. The results were compared and contrasted in terms of alternative rankings of the 27 Faculties. Although caution is obligatory when interpreting the results (either for the low number of statistical units or the context of the elaboration was different from that of Censis), this study finalized with comparative analyses of the rankings obtained using the various techniques, and with some proposals for alternative composite indicators.

3.2 The Censis Ranking System

Every year the *Grande Guida all'Università* proposes rankings of the Italian universities and faculties.[1] We decided to analyse only the Censis ranking system of the 27 Faculties of Political Science. The *Guida* can be evaluated from two viewpoints:

1. a vertical one, on which comparison is made among faculties as a whole and by "areas[2]" of indicators;
2. an horizontal one, on which the strengths and weaknesses of each faculty are assessed.

The *Guida* proposes a ranking of faculties to assist future freshmen and their families in making a more conscious choice. In order to translate this evaluative goal into quantities, Censis identified five areas:

- *productivity*, which measures a faculty's capacity to guarantee the regular fulfilment of examination requirements of degree courses;
- *educational sustainment*, which comprises a balanced student/academic staff ratio, the provision of adequate facilities, suitable course programmes, etc.;
- *research*, which evaluates the capacity of academic staff to plan their research, and the probability of a student to have lecturers with good research experience;
- *academic profile*, which identifies faculties that endeavour to rejuvenate their teaching staff and enhance international relations;
- *international relations*, which measures the openness of faculties to international study opportunities both for their students and their teaching staff [11];

[1] Censis evaluates universities along four dimensions: services, study grants, facilities, and website. Faculties are assessed by means of composite indicators of five areas. It should be borne in mind that the university league table does not depend on the results obtained by faculties, and *vice versa*.

[2] Censis calls "family" each "area" of the university system.

- *attractiveness* of each faculty, in regard to other universities and faculties of the same field.

Three methodological considerations seem to be necessary before the analysis:

1. The five areas have changed over time. The *attractiveness* was included in the *Guida* only in the first 2 years [8, 9], and the *academic profile* was introduced in subsequent years (but is no longer present in the [12] report). Moreover, the set of simple indicators has changed from year to year.
2. Data were available for the year 2006, i.e. when the reform of the Italian university system (according to D.M. 509/99) was just consolidated. Consequently, indicators would not be affected by institutional changes.
3. The Censis approach implies an underlying – and non-explicit – compensatory logic whereby good performances on a particular aspect off-set negative results on a different one (as often happens when attempts are made to synthesis the diverse features of a complex concept into a single measure).

3.3 Indicators for Evaluation and Measurement

The main task of the statistician is to translate the characteristic features of a phenomenon into numbers by means of a sensible definition of a pertinent concept. In social research, the quantitative and qualitative evaluation of a given phenomenon (hereafter designated by a concept) consists in a procedure whereby the particular feature possessed by a particular statistical unit is determined [11] by a number (quantitative information) or a category (qualitative information). Quantitative information lies at a higher level than that of qualitative information, even if the qualitative information is its basis. The question *how much*, in fact, often implies implicitly the other question *what*, whilst the reverse does not often occur. The process of measurement enables the feature measured to be represented and quantified by numbers, and it states the empirical relationships of interest in algebraic relationships among the numerical values assigned [4].

If the concepts to be evaluated are not directly measurable, it is necessary to use *indicators*. Indicators must be simple and are specific tools which can be translated into terms tied to general concepts by a linkage of semantic representation [13].

Given the copious output of statistical information in social research, there is some enthusiasm for the construction and production of social indicators. As a result, a number of questions arise concerning the sensibleness of the choices taken and the methods used when constructing indicators.

One of the main issues is what indicators should be used, and for what purpose [7]. Indicators constitute the linkage between observations and the complex concept to be measured. On the assumption that the aim of the research determines the indicators, which assume the meaning of meta data: they help shed light on the concept to measure, and they perform the dual task of specifying and measuring the concept.

Indicators are items of information which synthesise the characteristics of a concept or highlight what is occurring within it. They often result from a compromise reached at reasonable cost between scientific accuracy and availability of information. A "composite" indicator[3] is not just the result of a thorough process of evaluation: it may also be the starting-point for political discussion of the phenomenon under study.

It is difficult to identify the best way to measure a complex, multidimensional, and abstract concept, both from the point of view of the sense of the measurement and the field of application. But what is the use of comparing specific components if the aim is to compare systems, and not a set of specific components of the system? If the objective is to receive warning signals, attention should focus on measuring the components, keeping the information disaggregated into "simple" indicators. When the purpose of the analysis is to compare systems or situations, synthesis with a "composite" indicator is necessary [3].

Indicators are classified according to various criteria. An important distinction is drawn between simple and composite indicators. *Simple* (or elementary) indicators refer to a simple unidimensional concept, or to one of the immediately quantifiable dimensions of a complex multidimensional concept. The aggregation and possible weighting of several simple indicators give rise to what is called a *composite* indicator. From the computational point of view, there may be three "key steps" which lead to the determination of a composite indicator: *normalization*, *aggregation*, and *weighting* of the simple indicators.

According to Land [18], an indicator is meaningful only when it possesses informational value within a theoretical model, however it may be defined – mathematically, operationally, logically, orally, etc. – for the analysis and interpretation of social phenomena. In recent decades, the history of indicators seems to have a further principle to this definition: an indicator is usually the outcome of the decomposition of a complex concept into its elementary components. It is a process of reassembling through procedures which normalize, aggregate and weight the simple indicators. This process obviously come from qualitative and quantitative information, subjective and objective observations, descriptions, analysis, and interpretation of existing sources or ad hoc surveys. The indicator therefore exists within a model and it is also produced by the model itself. According to this principle, the indicator often increases the content and meaning of the complex concept being examined within the model.

An indicator is a tool to convert the measurement of complex concepts into a systematic array of interpretative conjectures and relations incorporated into a functional model. But some distance persists between the heuristic intent and the operational feasibility. There exists, in other words, a gap between the (convinced, essential, sometimes normative/legislative) intention to assess a complex concept and its realistic measurement (broadly determined by the system of operational conditions actually adopted or adoptable, even when accompanied by careful and explicit reflection on the methodological rigour of the entire process).

[3] A definition of simple and composite indicators follows.

3.4 The Censis Data

Censis gave us the raw data used to compile the 2006 rankings, and made available the data for the 27 Faculties of Political Science of the Italian public universities.

The techniques of normalization and aggregation adopted by Censis, the list of simple indicators and the preliminary analyses are explained in the following paragraphs.

3.4.1 Normalization and Aggregation

The normalization technique used by Censis is a max-min standardization which converts the values into indicators, dividing by the range:

$$I = \frac{X - \min(X)}{\max(X) - \min(X)} \times 1{,}000 \qquad (1)$$

where X is the value of the raw indicator, whilst $\min(X)$ and $\max(X)$ are respectively the minimum and maximum value that the indicator assumes in the set of homogeneous faculties considered.[4] The transformed values will therefore vary from a minimum of 0 to a maximum of 1,000, and they will be comparable within each cluster of faculties: in fact, it is not possible to compare different faculties by means of the same indicator.

Censis rescales the values of the indicators in the interval 66–110, which represents the range of grades awarded for degrees in Italy. Because the formula for this transformation was not reported in the methodological notes, and since the results did not change because it is a linear transformation, this rescaling was not necessary and was not performed in our calculations.

The average final grade M attributed to each faculty was calculated as the arithmetic mean of the normalized scores of the five areas considered:

$$M_f = \frac{std(P_f) + std(D_f) + std(R_f) + std(PD_f) + std(RI_f)}{5} \qquad (2)$$

where f denotes each faculty (from 1 to 27) and P is the score for the productivity area, D the score for educational delivery, R for research, PD for the academic staff profile, and RI for international relations.

[4] An interesting alternative would be the use of the theoretical maximum and minimum with the simple indicators for which such values are determinable: this technique would make it possible to reduce the distances among units observed in terms of residuals among normalized values of the simple indicators.

3.4.2 The Simple Indicators Used by Censis

The simple indicators are reported in Tables 3.1, 3.2, 3.3, 3.4 and 3.5. To highlight that simple indicators are calculated for each faculty, we use the subscript f (from 1 to 27). Each simple indicator is normalized according to the formula 1. This transformation is indicated in Tables 3.1, 3.2, 3.3, 3.4 and 3.5 as $std(\cdot)$.

The thresholds for the k values were stated by Censis.

3.4.3 Preliminary Analysis

Since exploratory analysis of the data and study of relations among the variables is an important phase, we started with this operational step in order to verify the existence of relations among the simple indicators considered.

Table 3.1 Simple indicators for "productivity" area (Censis, 2006)

	Productivity
P_{1f}	Rate of persistence between 1st and 2nd year: *(students enrolled in the 2004–2005 academic year who were freshmen in the previous year)/(freshmen in 2003–2004)*
P_{2f}	Regularity index of students: *60 × (credits acquired in 2004 by students enrolled on the first level 3-year degree or on the "single-cycle" 5-year degree courses)/(students enrolled on the first level 3-year degree or on the "single-cycle" 5-year degree courses in the 2003–2004 academic year)*
P_{3f}	Rate of students enrolled "in corso[5]": *(total students enrolled – freshmen – students enrolled "fuori corso")/(total students enrolled – freshmen)*
P_{4f}	Rate of 3-year graduates: *(graduates in 2004 from 3-year degree courses who were enrolled in the 2001–2002 academic year)/(freshmen on 3-year degree courses in the 2001–2002 academic year)*
P_{5f}	Rate of graduates "in corso": *(graduates "in corso" in 2004 from 3-year single-cycle degree courses and from previous 4-year degree programmes)/(total graduates from the courses stated)*

Aggregation formula[6]

$$P_f = \frac{std(P_{1f}) + std(P_{2f}) + std(P_{3f}) + std\left(\frac{std(P_{4f})n_1 + std(P_{5f})n_2}{n_1+n_2}\right)}{4}$$

where
$k = 1$ if $D_{8f} < 75$
$k = 1.05$ if $D_{8f} \geq 75$

[5] Students "in corso" have fulfilled their examination requirements within the scheduled deadlines, whilst students "fuori corso" are still attending university beyond the duration of their courses because they have not yet completed their examination requirements.

[6] Because $P5$ was furnished by Censis as a rate, it was not possible to derive the value of n_2. The score for P was obtained as the simple arithmetic mean of the simple indicators.

Table 3.2 Simple indicators for "educational delivery" area (Censis, 2006)

	Educational delivery
D_{1f}	*Number of degree courses on the faculty programme in 2004–2005*
D_{2f}	*Number of subjects-courses on the faculty programme in 2003–2004*
D_{3f}	*(Tenured academic staff)/(number of subjects-courses in 2004 and 2004–2005)*
D_{4f}	*(Tenured academic staff on 31.12.2004)/students enrolled in 2004–2005)*
D_{5f}	*(Lecture room places NUCLEI 2004)/(students enrolled in 2002–2003)*
D_{6f}	*(Lecture room places NUCLEI 2005)/(students enrolled in 2003–2004)*
D_{7f}	*Student work experience placements (stage) in 2003–2004*
D_{8f}	*Monitoring and evaluation of courses in 2003–2004*

Aggregation formula[7]:

$$D_f = \frac{std\left(\frac{std(D_{1f})+std(D_{2f})+std(D_{3f})}{3}\right) + std(D_{4f}) + std\left(\frac{std(D_{5f})+std(P_{6f})}{2}\right) + 0.5std(D_{7f})}{4} \times k$$

where:
$k = 1$ if $D_{8f} < 75$
$k = 1.05$ if $D_{8f} \geq 75$

Table 3.3 Simple indicators for "research" area (Censis, 2006)

	Research
R_{1f}	*(Number of research units funded by the COFIN and FIRB programmes in 2003)/(tenured staff on 31.12.2002)*
R_{2f}	*(Number of research units funded by the COFIN and FIRB programmes in 2004)/(tenured staff on 31.12.2003)*
R_{3f}	*(Number of research units funded by the COFIN and FIRB programmes in 2005)/(tenured staff on 31.12.2004)*
R_{4f}	Average COFIN and FIRB funding:*(total funding obtained by research units from the COFIN and FIRB programmes in 2003)/(number of units funded)*
R_{5f}	Average COFIN funding:*(total funding obtained by research units from the COFIN programme in 2004)/(number of units funded)*
R_{6f}	Average COFIN and FIRB funding:*(total funding obtained by research units from the COFIN and FIRB programmes in 2005)/(number of units funded)*
R_{7f}	*Number of research projects funded by the EC V and VI Framework Programme and Tempus Programme*

Aggregation formula:

$$D_f = \frac{std\left(\frac{std(R_{1f})+std(R_{2f})+std(R_{3f})}{3}\right) + std\left(\frac{std(R_{4f})+std(R_{5f})+std(R_{6f})}{3}\right)}{2} \times k$$

where:
$k = 1$ if $R_{7f} = 0$
$k = 1.05$ if $R_{7f} > 0$

[7] In the case of the educational delivery, the weighting used by Censis raises obvious questions concerning the weights (it is not stated whether specific choices were made) because the denominator of the formula should be 3.5 instead of 4. For the sake of consistency, we decided to keep the formula applied by Censis.

Table 3.4 Simple indicators for "academic staff profile" area (Censis, 2006)

	Academic staff profile
PD_{1f}	*Average age of tenured academic staff in 2005*
PD_{2f}	Ageing: *(average age of tenured academic staff in 2005) – (average age of tenured academic staff in 2001)*
PD_{3f}	Outgoing Erasmus students per member of academic staff: *(students with Erasmus grants in 2004–2005)/(tenured academic staff on 31.12.2004)*
PD_{4f}	*(Courses taught by untenured "extra-academic" lecturers)/(total courses taught in 2003–2004)*
PD_{5f}	"Rientro dei cervelli" programme: *number of lecturers participating in the international mobility programme for Italian and foreign scholars in thet three-year period 2004–2006*

Aggregation formula[8]:

$$PD_f = \frac{std\left(\frac{std(PD_{1f})+std(PD_{2f})}{2}\right) + 0,5std(PD_{3f}) + 0,5std(PD_{4f})}{3} \times k$$

where:
$k = 1$ if $PD_{5f} = 0$
$k = 1.05$ if $PD_{5f} > 0$

Table 3.5 Simple indicators for "international relations" area (Censis, 2006)

	International relations
RI_{1f}	Outgoing Erasmus grant-holders per student: *(outgoing students with Erasmus grants in 2004–2005)/(students enrolled net of matriculants in 2004–2005)*
RI_{2f}	Incoming Erasmus grant-holders per student: *(average number of foreign students who obtained an Erasmus grant at the faculty in 2003–2004 and 2004–2005)/(students enrolled in 2004–2005)*
RI_{3f}	Host universities per lecturer: *(number of foreign universities which hosted Erasmus students in 2004–2005)/(tenured lecturers on 31.12.2004)*
RI_{4f}	International opportunities: *(number of contributions obtained by the faculty for international cooperation schemes in 2003–2006: lecturer exchanges financed by Miur in 2004; Programma Vigoni 2003–2004; Programma Italia-Germania 2003–2004; Azioni Italia-Spagna 2004–2005; Programma Italia-Germania 2004–2005; Programma Galileo Italia-Francia 2004–2005; Cooperazione Internazionale finanziata dal Ministero degli Esteri – Accordi Bilaterali 2002–2006)*

Aggregation formula:

$$RI_f = \frac{std(RI_{1f}) + std(RI_{2f}) + std(RI_{3f})}{3} \times k$$

where:
$k = 1$ if $RI_{4f} = 0$
$k = 1.05$ if $RI_{4f} > 0$

[8] Likewise the case of the educational delivery, in the academic staff profile the denominator should be 2 and not 3. For the sake of consistency, we decided to keep the formula applied by Censis.

The independence between pairs of variables was measured by means of the Bravais-Pearson coefficient of correlation. The statistical significance was tested with a null hypothesis equal to zero.

We first examined the correlation among the simple indicators belonging to the same area: evidence of correlations among the various indicators would indicate that some aspects had been measured – and therefore considered – several times within the same area. This would not have complied with the parsimony criterion which should guide the construction of composite indicators.

Of course, the results were affected by the small number of faculties available: 27 units, in fact, did not represent a number of observations sufficient to produce stable and convincing results. Moreover results were not extendable to the universe of the Italian faculties. Table 3.6 lists the correlations higher than ±0.4 within the areas (we consider only the values of the correlations because the analysis refers to all the faculties).

We analyzed the correlations among all the indicators. The resulting matrixes showed correlations among indicators belonging to different areas, which suggested the existence of a hypothetical – and not unrealistic – effect of the same measures on different dimensions by means of indicators belonging to different areas.

Table 3.6 Pairwise correlations among simple indicators (values higher than ±0.4)

Area	Value	Indicators
Productivity	0.648	P2 vs. P4
Educational delivery	0.701	D_1 vs. D_2
	0.617	D_6 vs. D_7
	0.587	D_5 vs. D_6
	0.546	D_5 vs. D_7
	–0.435[9]	D_2 vs. D_3
Research	0.773	R_1 vs. R_3
	0.527	R_1 vs. R_6
	0.491	R_4 vs. R_6
Academic staff profile	0.519	PD_1 vs. PD_2
	0.486	PD_3 vs. PD_5
	–0.445[10]	PD_1 vs. PD_4
International relations	0.894	RI_1 vs. RI_2
	0.645	RI_1 vs. RI_3
	0.528	RI_2 vs. RI_3

[9] The correlation between D_2 and D_3 is negative because in the Italian university system who offer a higher number of courses usually has a minor tenured academic staff.

[10] The correlation between PD_1 and PD_4 is negative because a higher average age of tenured academic staff implies that the same staff taught the majority of the courses (untenured "extra-academic" staff usually taught a minor number of courses).

Table 3.7 Matrix of correlations among areas (data obtained using the Censis procedure)

	P	D	R	PD	RI
P	1	0.454	0.423	0.443	0.297
D		1	0.419	0.093	0.111
R			1	0.145	0.087
PD				1	0.200
RI					1

We finally calculated the matrix of correlations among the overall scores of the areas using the scores of the areas constructed by means of the Censis methodology (in some cases obtaining values slightly different from those published) as shown in Table 3.7.

High correlations among simple indicators belonging to the same area were highlighted: this may indicate that two indicators cover areas that overlap each other. This affects the validity property of the measurement process. The Censis aggregation method was used to synthesise the simple indicators of the same area (without changing the weights).

The aim of this correlation analysis was to point out the redundancy among the indicators used by Censis and to notice that this redundancy did not exist among areas. This analysis would be done by Censis considering the complete dataset relative to all the Italian faculty: in this way it could have stable results.

In order to complete a preliminary analysis of the data, we wanted to devote a specific section to multivariate analysis [20] intended to evaluate the number of latent statistical dimensions derivable from the simple indicators. Given the small size of the dataset available, it was not possible to obtain information useful to help us in constructing a different configuration of the areas.

3.5 Alternative Ways to Analyse the Data

Before adopting our strategy of analysis, we considered a list of techniques of normalization, aggregation and weighting [2, 5, 14, 16, 20–22].

- Normalization comprises all the operations performed to transform the simple indicators so that they are comparable with each other in terms of direction, unit of measurement, and order of magnitude. It can be performed by means of:
  - *linear transformations* ($Y = \alpha + \beta X$ where the response variable Y is a linear function of the explanatory variable X [1]) as dividing by the range, as transformation into index numbers, standardization, comparison with the unit leader or a control group, distance from the median;
  - *non-linear transformations* (where the relationship f between Y and X, $Y = f(X)$, is nonlinear); the most used non linear function essentially to convert the data into ordinal values (ranks).

- Aggregation is the choice of merging through an appropriate function which combines different dimensions of the concept under study. It can be performed by means of:
  - *ordinal approach*, which synthesises the indicators transformed into ranks with their mean or sum;
  - *additive cardinal approaches*, which involve calculation of the mean of the transformed values;
  - *non-compensatory multi-criteria approach*, which solves the compensation problem via comparisons among couples of units;
  - *geometric aggregation*, an intermediate solution in terms of compensation between additive aggregations and the multi-criteria approach;
  - *multivariate aggregation techniques*, based on principal components analysis or factor analysis, which draw the latent dimensions that the data describe.
- Weighting is the phase of the process when weights are assigned to the indicators and/or to the dimensions of the concept. The weights may be:
  - *equal* for every variable: this is not a "non-choice" but it grants equal status to all the indicators;
  - based on multivariate models (the most common are *regression* and *factor analysis*);
  - derived from the application of *participatory methods*;
  - calculated by applying the *hierarchical analytical process* which breaks a problem down into a hierarchy and systematically collects opinions on the indicators through pairwise comparisons;
  - derived by the distance from a defined *efficiency frontier*;
  - estimated using an *unobserved components model*.

Among all the normalization, aggregation and weighting techniques listed above, we decided to use those that the literature indicates as the most robust and convincing. Some methods of analysis were discarded due to the small amount of data available. We wanted to adopt techniques which were mutually compatible but based on different approaches and selected two normalization methods: linear and non-linear. We consequently decided to use two different aggregation methods applicable to any normalization. Finally we also adopted two systems of weights: equal for every area (as in the Censis procedure), and the other one based on the participatory method.

It is worth to mention that we first applied the Censis aggregation and weighting techniques to our data, in order to obtain the same results published in the *Guida*. The starting point for our procedure was the set of simple indicators that we had constructed from the variables furnished by Censis.

- The simple indicators were therefore normalized by means of three different techniques:
  1. dividing by the range (as in the Censis procedure);
  2. standardization with z scores;
  3. rank transformation.

- Indicators were aggregated in two distinct steps: first the simple indicators of the same area were aggregated; then the scores of the five areas were aggregated to produce the final league table. We performed only the second aggregation, keeping unchanged the one made by Censis to calculate the final values of each area. The two methods selected for the aggregation were the following:

  1. arithmetic mean;
  2. geometric aggregation.

  Rather than the non-compensatory multi-criteria approach, we opted for geometric aggregation for several reasons: because it is a simply-to-use technique; it is easy to understood; it is better suited to a small dataset; it is a good compromise in terms of compensation between the multi-criteria approach (which excludes compensation) and linear approaches (which do not concern compensation). Moreover the geometric aggregation enabled us to compare our results with those published keeping our assumption close to those adopted by Censis.

- Two methods were selected for the weighting:

  1. equal weights for each area;
  2. participatory method with the "allocation of a budget" by experts.

  The weighting based on the expert judgments was done by us: we "arrogated" this role to ourselves by assigning a weight equal to 0.25 to *educational delivery*, *research* and *academic staff profile* areas, and a weight equal to 0.125 to *productivity* and *international relations* areas. A lower weight was assigned to *productivity* because it was too closely tied to the composition of the student component, and because of the ambivalence of the indicator's information content (good rates of graduates and students "in corso" do not necessary mean a good performance in terms of *productivity*). A lower weight was given also to *international relations* because these substantially only concerned the Erasmus Programme, whilst other activities were omitted. It would be interesting to use the participatory approach with experts on the university system to obtain a shared system of weights. This could also be done by Censis using the results of the surveys conducted with the faculty deans.

Hence 12 ranking tables were obtained by applying the three different normalization methods, the two aggregation techniques, and the two systems of weights. They are summarized in Table 3.8.

In the following analysis we did not considered two methods out of 12. There were marked differences for the C2 and D2 methods due to the computational problems in the geometric aggregation of the standardized z scores.

Censis prefers simple mathematical processes instead of complex statistical models because the readers of the *Guida* are future freshmen and their families which could not appreciate complex statistical methods. For this reason we decided to work in the same perspective.

Table 3.8 Combination of the normalization, aggregation, and weighting techniques in constructing the 12 ranking tables

	Range	Z scores	Ranks
Mean	A1 equal weights B1 weights by experts	A2 equal weights B2 weights by experts	A3 equal weights B3 weights by experts
Geometric Aggregation	C1 equal weights D1 weights by experts	C2 equal weights D2 weights by experts	C3 equal weights D3 weights by experts

3.6 Results

For each method we obtained a list of 27 values and a position for each faculty in a ordered list (ranking). We compared and contrasted the 10 ranking tables of the combination of the normalization, aggregation, and weighting techniques and the league table published by Censis. Finally, we synthesized them into a combined ranking table (the best estimation of the "true" league table of the faculties).

The results of the 10 rankings are reported in Table 3.9, where the cells show the position of each faculty according to each method. The last column of the table reports the position of the faculties in the league table published by Censis.

Table 3.9 Rankings of the faculties (10 ranking methods and the league table of Censis)

Faculty	A1	A2	A3	B1	B2	B3	C1	C3	D1	D3	Censis
Bari	26	26	27	27	27	27	21	27	21	27	26
Bologna	2	3	1	2	2	3	1	3	1	4	1
Cagliari	15	16	14	14	16	13	14	15	13	13	12
Calabria – Cosenza	7	6	11	3	4	9	8	8	6	5	8
Catania	25	24	25	25	24	25	20	25	20	25	25
Firenze	3	2	4	4	3	5	3	2	3	3	5
Genova	10	13	16	15	15	18	12	14	16	16	9
Macerata	23	23	23	24	25	23	23	23	23	23	21
Messina	21	21	21	23	22	22	22	22	22	22	23
Milano 1	8	8	3	7	7	1	6	4	4	2	7
Napoli Orientale	17	19	16	19	19	17	23	17	23	18	17
Napoli 1 – Federico II	27	27	26	26	26	26	23	26	23	26	27
Padova	12	12	11	13	14	14	11	12	12	15	15
Palermo	24	25	20	22	23	20	23	21	23	21	24
Pavia	5	5	4	6	5	5	4	6	5	6	2
Perugia	9	9	7	11	12	7	7	7	10	11	10
Piemonte Orientale	1	1	4	1	1	2	17	1	9	1	6
Pisa	22	20	21	20	20	21	23	20	23	20	20
Roma 1	20	22	24	21	21	24	19	24	19	24	22
Roma 3	6	7	9	8	8	9	5	11	7	12	4
Salerno	18	17	19	17	18	19	16	19	17	19	18
Sassari	16	15	15	16	13	14	15	16	15	14	16
Siena	11	10	8	10	10	8	9	10	11	8	11
Teramo	13	14	13	9	9	12	10	13	8	10	13
Torino	14	11	9	12	11	9	13	9	14	9	14
Trieste	4	4	1	5	6	4	2	5	2	7	3
Urbino – Pesaro	19	18	18	18	17	16	18	18	18	17	19

In order to compare and contrast the positions of the faculties in the ten ranking tables with respect to the position in the league table of Censis we reported the frequencies of the absolute differences in Table 3.10. We noted a general concordance of the results, with the exception of some faculties for which the distances from the Censis values seemed rather wide: Genova, Piemonte Orientale, Roma 3, and to a lesser extent, Padova and Torino.

We found a good concordance between our results and the league table published by Censis. A brief inspection of the tables immediately showed that the positions of the faculties were not particularly variable among ranking methods and with respect to the ranking table published by Censis. The highest cograduation was between A1 and Censis. The same normalization, aggregation and weighting techniques were used to construct the two ranking tables; the only difference consisted in the initial set of simple indicators, for which, however, there was no evidence of a close correspondence between our indicators and those elaborated by Censis. Results showed a generally high consistency between the league table of Censis and our methods that use data normalized in ranks (in order, A3, C3, B3 and D3). The attribution of different weights to the areas became important (D-type methods): despite the presence of few data, the weighting had an important role in defining the positions in the ranking tables.

The 10 ranking tables obtained with different normalization, aggregation and weighting methods showed a high level of concordance. To obtain a measure of this concordance we used Kendall's coefficient W([17]: 95) and its chance-corrected version W_1[6, 15]. For our ten ranking tables we obtained $W = 0.91$ and $W_1 = 0.92$. Since the coefficients are close to 1 we could estimate a combined "best" ranking table of the faculties. According to Kendall [17] the best ranking table could be obtained by means of the sum of the ranks, i.e. the position of a faculty is determined by the sum of its positions in the ten ranking tables. Although this approach would increase the computational complexity, it would ensure greater robustness and reliability of the final results [6, 15].

The result of this combined ranking table is reported in Table 3.11. The second and third columns report the position of the faculty in the league table published by Censis and the absolute difference between the positions. We noted again a general concordance of the results, with the exception of some faculties for which the distances between the position in the combined ranking table and the position proposed by Censis were wider: Genova, Pavia, Roma 3, and to a lesser extent, Piemonte Orientale, Firenze and Macerata.

3.7 Conclusions

The ranking of university institutions always causes controversies, expectations, and criticisms in the actors (areas, actual and potential university students, and academic "actors"). In this study we have addressed the core of the problem by focusing on the ranking method proposed by Censis in its *Grande Guida all'Università* and analysing its structure, our purpose being to understand what measurement

Table 3.10 Distribution of absolute differences (10 ranking methods vs. the league table of Censis)

Differences	A1	%	A2	%	A3	%	B1	%	B2	%	B3	%	C1	%	C3	%	D1	%	D3	%
0	11	40.7	4	14.8	3	11.1	5	18.5	3	11.1	3	11.1	2	7.4	6	22.2	4	14.8	3	11.1
1	6	22.2	10	37.0	8	29.6	9	33.3	9	33.3	10	37.0	8	29.6	6	22.2	7	25.9	7	25.9
2	6	22.2	5	18.5	7	25.9	6	22.2	4	14.8	4	14.8	6	22.2	4	14.8	3	11.1	5	18.5
3	3	11.1	5	18.5	3	11.1	1	3.7	4	14.8	4	14.8	5	18.5	6	22.2	7	25.9	5	18.5
4	0	0.0	2	7.4	3	11.1	3	11.1	5	18.5	2	7.4	2	7.4	1	3.7	1	3.7	2	7.4
5	1	3.7	1	3.7	2	7.4	2	7.4	1	3.7	2	7.4	2	7.4	3	11.1	3	11.1	3	11.1
6+	0	0.0	0	0.0	1	3.7	1	3.7	1	3.7	2	7.4	2	7.4	1	3.7	2	7.4	2	7.4
Tot.	27	100	27	100	27	100	27	100	27	100	27	100	27	100	27	100	27	100	27	100

Table 3.11 Rankings of the faculties in the combined ranking table

Faculty	Comb. Class.	Censis	diff.
Bari	26	26	0
Bologna	1	1	0
Cagliari	14	12	2
Calabria – Cosenza	7	8	1
Catania	25	25	0
Firenze	2	5	3
Genova	15	9	6
Macerata	24	21	3
Messina	21	23	2
Milano 1	5	7	2
Napoli 1 – Federico II	27	27	0
Napoli Orientale	19	17	2
Padova	13	15	2
Palermo	23	24	1
Pavia	6	2	4
Perugia	9	10	1
Piemonte Orientale	3	6	3
Pisa	20	20	0
Roma 1	22	22	0
Roma 3	8	4	4
Salerno	18	18	0
Sassari	16	16	0
Siena	10	11	1
Teramo	11	13	2
Torino	12	14	2
Trieste	4	3	1
Urbino – Pesaro	17	19	2

instruments can be used, how to combine them, and how to obtain robust final results. Our intention has not been to criticise the Censis ranking system a priori, but rather to analyse how it can be adjusted and/or improved, as well as to suggest possible alternatives to it. However, we wish to make a proposal: we regard it as both necessary and desirable for Censis to clearly state how it has selected and/or determined the "areas" used to evaluate the university system when its university league table is published.

We have considered indicators as basic tools to operate, and we have argued that the synthesis of indicators is crucial for evaluation processes. When discussing the complex process of constructing a composite indicator, we highlighted the normalization, aggregation and weighting phases, and we illustrated a set of techniques based on different theories and suited to different purposes. With a view to comparison among several situations, as well to give warning signals on individual aspects, analysis must synthesise the information. This, therefore, is what we have sought to do: apply different operational techniques to the data in order to obtain results that enable comparison among university faculties on the basis of a synthesis of a wide range of alternative applications. Geometric aggregation becomes preferable to the simple linear aggregation which calculates the average of the items; weighting assumes significant importance in synthesis of the information; simple normalization techniques (e.g. the transformation of the simple indicators into ranks) and more complex standardizations conduct to similar results. In general, we would suggest the use of normalization, aggregation and weighting techniques that are not overly complex with respect to the assumptions and objectives of the analysis. This will foster better understanding of the methodology employed by the *Guida* for the readers, and especially its target audience of future university students and their families. Moreover, standardization by the range proved not to be a good normalization technique, because large distances between maximum and minimum values were amplified. The use of ranks was a good alternative method of data normalization instead of the method based on the range. Geometric aggregation (except in the case based on z scores) was a good aggregation technique based on a logic of non-complete compensability among the indicators (for the role assumed by weights in different aggregation methods see [19]).

The data used in our calculations have been collected and furnished by Censis. It was, therefore, essential to regard them as "quality" data and attribute to them – a posteriori for obvious reasons – the properties of accuracy, validity and consistency. This, however, prompts a necessary consideration: it is essential to verify the quality of data also by making careful selection of the information deemed useful and necessary, without giving in to the temptation to "cherry-pick" information from the sources available.

The small amount of data available for our calculations has restricted the range of possible applications. In particular, it has precluded analysis of the structure underlying the data using multivariate analysis methods. Factor analysis of the entire set of simple indicators might yield areas different – in number and significance – from those (pre-)determined by Censis, considering the complexity and delicacy of establishing them a priori. Ex post cluster analysis might instead be useful for verifying the existence of geographical areas or types of faculties which are problematic or virtuous according to the aspects analyzed. Also preliminary analyses based on correlations, if performed on a larger dataset, could highlight redundancies among the indicators belonging to the same area or overlaps among areas.

Given the overall structure of the league table of Censis, an important observation concerns its lack of measures of variation in position within the ranking table and of year-by-year changes in the scores for the areas for each faculty. Of course, any evaluation in this sense must consider the changes that take place every year in the structure of the indicators and areas, changes which entail that annual rankings are not entirely comparable.

In this study we have asked whether Censis is a reliable "referee". A series of choices made by Censis produce results quite similar to those yielded by the alternative strategies used here. From this point of view, we may say that the technical-methodological aspects of constructing composite indicators seemingly do not give rise to significant differences in the results. There appear to be two main discriminating factors: the nature, articulation and quality of the database used to represent the sub-dimensions of the concept considered; and the strategy used for their weighting within the areas (or "family" as Censis calls it), and among areas. In its *Guida* and on its website, Censis allows examination and evaluation of the general procedure that it adopts. We might call this an ex-post "search for transparency": we believe that there should be a joint effort by the actors involved, and more generally by the stakeholders. All that in order to establish the objects, the rules of the game could be a way, laborious but necessary, to improve the process and to achieve an outcome which creates less wrangling, less discontent, less indifference, less ill-feeling. A participatory process involving all the stakeholders is less agile and efficient than appointment of an actor external to the system. Nevertheless, the issue is a highly sensitive one, and it warrants higher-level discussion if the results of a ranking method will be more believable and have a real effect on the university system. The literature on evaluation devotes ample space to the issue of the quality of the interaction among actors, especially in complex, dynamic and turbulent contexts. The correct management of relations among actors when a ranking system is adopted is necessary so that there is a co-responsibility (collective assumption of responsibility) for processes, greater recognition of the value of the results achieved and, therefore, also greater future use of the indications obtained.

References

1. Agresti A, Finlay B (1997) Statistical methods for the social sciences. Prentice-Hall Inc, Englewood Cliffs, NJ
2. Aiello F, Attanasio M (2004) How to transform a batch of simple indicators to make up a unique one? Atti del Convegno SIS giugno 2004, Bari – Sessioni Specializzate
3. Aiello F, Attanasio M (2006) Some issues in constructing composite indicators. VIII international meeting on quantitative methods for applied sciences, Certosa di Pontignano, 11–13 settembre 2006
4. Aiello F, Librizzi L (2006) Gli indicatori nelle scienze sociali: dal qualitative al quantitativo. In: Diamond I, Jefferies J (eds) Introduzione alla statistica per le scienze sociali. Mc Graw-Hill, Milano
5. Allegra FS, La Rocca A (2004) Sintetizzare misure elementari: una sperimentazione di alcuni criteri per la defnizione di un indice composto. Istat, Roma
6. Attanasio M, Capursi V (1997) Graduatorie sulla qualità della vita: prime analisi di sensibilità delle tecniche adottate. Atti XXXV Riunione Scientifica SIEDS, Alghero
7. Bernardi L, Capursi V, Librizzi L (2004) Measurement awareness: the use of indicators between expectations and opportunities, SIS: Sezione Specializzata. Atti della XLIII Riunione Scientifica, Bari
8. Censis (2000) Grande guida all'università, La Repubblica, Milano
9. Censis (2001) Grande guida all'università, La Repubblica, Milano
10. Censis (2006) Grande guida all'università, La Repubblica, Milano
11. Censis (2007) Grande guida all'università, La Repubblica, Milano
12. Censis (2008) Grande guida all'università, La Repubblica, Milano
13. Corbetta PG (1999) Metodologia e tecniche della ricerca sociale. Il Mulino, Bologna
14. Delvecchio F (1995) Scale di misura e indicatori sociali. Cacucci Editore, Bari
15. Fagot RF (1994) An ordinal coefficient of relational agreement for multiple judges. Psychometrika 59(2):241–251
16. Freudenberg M (2003) Composite Indicators of Country Performance: A Critical Assessment, OECD Science, Technology and Industry Working Papers 2003/16, OECD, Directorate for Science, Technology and Industry

17. Kendall MG (1962) Rank correlation methods. C. Griffin & Co. Ltd, London
18. Land KC (1971) On the definition of social indicators. Am Sociol 6:322
19. Munda G, Nardo M (2005) Constructing consistent composite indicators: the issue of weights. Institute for the Protection and Security of the Citizen, Luxemburg
20. Nardo M, Saisana M, Saltelli A, Tarantola S (2005a) Handbook on constructing composite indicators: methodology and user guide. OECD statistics working papers, 2005/3. OECD Publishing, Paris
21. Nardo M, Saisana M, Saltelli A, Tarantola S (2005b) Tools for composite indicators building. European Commission-Joint Research Centre, Ispra, Italy
22. Saisana M, Tarantola S (2002) State-of-the-art report on current methodologies and practices for composite indicator development. European Commission-Joint Research Centre, Ispra, Italy