

Journal of Official Statistics, Vol. 30, No. 3, 2014, pp. 547-562, http://dx.doi.org/10.2478/JOS-2014-0034

Are All Quality Dimensions of Equal Importance when Measuring the Perceived Quality of Official Statistics? Evidence from Spain

Alex Costa¹, Jaume García², and Josep Lluis Raymond³

Quality has become the key concept in official statistics. There is a general consensus that we have to consider several components when assessing the quality of statistical information. Relevance, accuracy, timeliness, punctuality, comparability, coherence, accessibility and clarity are the dimensions most frequently mentioned. In this article we use regression analysis to evaluate the contribution of these different dimensions when assessing the overall quality of statistical products. We do this using the information collected in the structured consultation with users and experts from both inside and outside the Spanish Central Administration carried out by the Working Group of the Spanish High Council on Statistics, responsible for the preliminary draft of the proposals and recommendations of this council for the Spanish Multiannual Statistical Programme 2013–2016. We find that the above-mentioned dimensions have different weights in the overall assessment of perceived quality (with accuracy and reliability having the highest weight, and relevance having the lowest) and that the structure differs between both types of users.

Key words: Quality; dimensions; consultation; weight; regression.

1. Introduction

When discussing the first of the general issues raised in the report by the Working Party of the Royal Statistical Society (1991), namely the importance of retaining public confidence in official statistics, Fellegi (1991) mentioned the existence of a virtuous circle in official statistics: Public confidence is a prerequisite for high-quality statistics and high-quality statistics must ultimately be the foundation for public confidence. Since then, we can find

¹ Institut d'Estadística de Catalunya, Via Laietana 58, 08003 Barcelona, Spain. Email: acosta@idescat.cat

² Departament d'Economia i Empresa, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain. Email: jaume.garcia@upf.edu

³ Department d'Economia i d'Història Econòmica, Universitat Autònoma de Barcelona, Edifici B, Campus Bellaterra, 08193 Bellaterra, Barcelona, Spain. Email: josep.raymond@uab.cat

Acknowledgments: We wish to thank the Spanish Statistics Institute (*Instituto Nacional de Estadística*, INE) for allowing us to access the data used in this paper and, in particular, Carlos Paulogorran, Paz Sánchez and Pilar Ochoa from INE, who helped us with the preparation of the data set. We are also grateful for the comments made by Ada Ferrer and Juan M. Rodríguez on a preliminary version of this paper. We also benefited from the comments by three anonymous referees, an associate editor and the Co-Editor-in-Chief. The usual disclaimer applies. Alex Costa was Director of Planning, Coordination and Statistical Dissemination of INE between June 2009 and February 2012. Jaume García was President of INE between April 2008 and December 2011. Josep Lluis Raymond was the Chairman of the Working Group of the Spanish High Council on Statistics (SHCS) which was responsible for the preliminary draft of the proposals and recommendations of the SHCS for the Spanish Multiannual Statistical Programme 2013-2016.

explicit references to quality as a key element in official statistics in most of the relevant documents of the statistical institutions. In that regard, the preamble of the European Statistics Code of Practice (European Statistical System 2011) includes an explicit reference to quality when defining the mission of the European Statistical System (ESS), taken from the Quality Declaration of the ESS (European Statistical System 2001), which considers the provision of high-quality information on the economy and society on European, national and regional levels. Moreover, the first of the principles governing international statistical activities (United Nations Statistical Commission 2006) refers to quality, stating that high-quality international statistics are a fundamental element of global information systems.

As mentioned by Vale (2010), the concept of quality in official statistics has moved from the traditional approach in which quality corresponds to how closely statistics reflect reality (e.g., the mean square error of an estimator), that is accuracy and reliability, to a situation in which the performance of statistical services is evaluated in terms of how they respond to users' needs (Castles 1991). This new concept fits perfectly with the ISO 9000 definition of quality – the ability of a set of characteristics to satisfy requirements – and is explicitly reflected in the European Statistics Code of Practice when talking about the statistical output, as well as in the first of the Fundamental Principles of Official Statistics (United Nations Economic Commission for Europe 1992). In fact, this actual concept of quality reinforces the consideration of statistics produced by public institutions as a public good (Giovannini et al. 2009).

As mentioned in the final report of the Leadership Expert Group on quality (Eurostat 2002), quality consists of a number of features reflecting users' needs and can be defined along a number of dimensions (i.e., quality is a multidimensional concept), which constitute the product quality. Nowadays, there is almost complete agreement among international statistical institutions about the components (or dimensions) of statistical quality, as is shown in Vale's (2010) mapping of quality components in international statistical organisations (p. 6). In particular, Article 12 of Regulation (EC) No 223/2009 of the European Parliament and of the Council on European Statistics refers to the criteria which shall apply to guarantee the quality of results. These are: relevance; accuracy; timeliness; punctuality; accessibility and clarity; comparability; and coherence. These are the dimensions referred to in Eurostat's definition of quality of statistics introduced in 2003 (Eurostat 2003). However, since statistical offices are confronted with a heterogeneous typology of users, who have different needs and also different perceptions of quality, the importance of the components (or dimensions) will not necessarily be uniform and may differ among users. In fact, Brackstone (2001) thinks of the dimensions of quality in a hierarchical fashion. Relevance is at the top, timeliness and accessibility are not important without relevance, and accuracy, interpretability and coherence only become important when the other three dimensions are satisfied.

In this article we use the data collected in the structured consultation with users and experts from inside and outside the Spanish Central Administration carried out by the Working Group of the Spanish High Council on Statistics (SHCS) (*Consejo Superior de Estadística*) in 2010 when preparing the preliminary draft of the proposals and recommendations of the SHCS for the Spanish Multiannual Statistical Programme (SMSP) (*Plan Estadístico Nacional*) 2013–2016. The aim of these consultations was to provide

evidence on the importance of the different components when assessing data quality and, in particular, on whether different groups of users attach the same (or different) weights to the quality dimensions. Principles 11 to 15 of the European Statistics Code of Practice are used as references of the quality dimensions in the consultation questionnaire.

This study has some similarities with some recent exercises carried out in Greece (Nikolaidis 2012) and Portugal (Zilhao and Ribeiro 2012), but differs from them in terms of the way in which users' information is collected, the type of users considered and, most importantly, in terms of the main objective: the empirical analysis of the relationship between the general assessment of quality of official statistics and that of its different components.

The article is organised as follows. In the next section we describe the main features of the structured consultation. The basic results are presented in the third section. In Section 4 we report some evidence of how the dimensions of quality are related to its general assessment. The article ends with a summary of the main conclusions.

2. The Structured Consultation and the Institutional Framework

The SMSP is the main legal tool used by the Spanish Central Administration to define the statistical production to be developed by either its statistical services or any other entity dependent on them or in collaboration with the regional or local administration, covering a period of four years. The SMSP is approved by the government and afterwards is developed and carried out by means of the annual programmes, which are also approved by the government.

Additionally, the SHCS is an advisory body for the Central Administration statistical services in which informants, producers and users of official statistics are represented, for example trade unions and employers' organisations and other social, economic and academic groups, together with ministries and the Spanish Statistics Institute (*Instituto Nacional de Estadística*, INE). One important task of the SHCS is to contribute towards identifying the statistics to be included in the SMSP to improve the coverage of users' needs for information prior to the first draft of the SMSP. The SHCS also delivers opinions on the proposal of the SMSP before it is approved by the government.

In that regard, in March 2010 the SHSC set up a working group which was to be responsible for the preliminary draft of the proposals and recommendations for the SMSP 2013–2016. In its first meeting, the SHSC Working Group defined two main tools to evaluate the official statistics included in the actual SMSP (the Spanish official statistics from the Central Administration) as part of the first step before producing the draft: a Compliance Report by Eurostat, with an evaluation of the Spanish official statistics from the point of view of the European Statistical Regulations; and a Structured Consultation addressed to a wide range of users and experts from inside and outside the Administration. It was the first time in Spain that both the statistics produced by INE and those produced by the statistical services of the ministries were evaluated. In the past, similar exercises, but not as exhaustive, have been carried out only for the INE statistics.

The key concept of the consultation is quality, which the participants were asked to assess as such, in addition to different dimensions and aspects described below. Following the notion of data quality in the Handbook of Data Quality Assessment Methods and Tools by Eurostat (Ehling and Körner 2007), we must consider three elements: the characteristics of the statistical product; the perception of the statistical product by the user; and the characteristics of the statistical production process. The components (or dimensions) of the product's quality were used in the consultation as a framework to assess users' perceptions, although we know that the quality assessment by different users is not necessarily the same.

As explicitly stated in the questionnaires, the components (or dimensions) of quality included are those of the ESS, which are defined in Principles 11 to 15 of the European Statistics Code of Practice. They are the following, including the literal description contained in the questionnaire:

Relevance: Official statistics must meet users' needs. You must consider whether the objectives of each statistic are related to the users' expectations and to the potentialities of the data source.

Accuracy and Reliability: Official statistics must portray reality in an accurate and reliable way. You must consider the degree of closeness with reality based on the methodology used, paying attention to the sampling and nonsampling errors plus the biases associated with the different stages of the process.

Timeliness and Punctuality: Official statistics must be released in a timely and punctual way. Official statistics must be disseminated with the shortest time possible having passed between the date they become available and the event they describe, and according to a previously announced calendar.

Coherence and Comparability: Official statistics must be comparable over time, through space and between domains. Official statistics must allow related data from different sources to be used and combined in a reliable manner.

Accessibility and Clarity: Official statistics must be presented in a clear and understandable form, disseminated in a suitable and convenient manner, available and accessible on an impartial basis with supporting metadata and guidance.

The overall quality assessment follows the five aforementioned quality dimensions. Note that the dimensions are complex and highly abstract concepts and consequently that measurement could be an issue that could affect the interpretation of the results. However, as will be discussed below, since the participants in the consultation are users with a high level of expertise and experts, we assume that measurement problems are minimised.

The evaluation of the quality of the product and its components in the consultation refer to single statistics. Results can be aggregated at a sectoral level by assigning each statistic to a sector. In fact, each participant in the consultation could evaluate all the statistics in a particular sector, that is, each questionnaire refers only to one sector. The 279 statistics included in the Structured Consultation came from the Spanish Annual Statistical Programme 2010, and were classified in 22 sectors.

There are also some questions in the questionnaire which correspond to assessments at the sectoral level. In particular, one question refers to the extent to which available statistics meet users' needs in a particular sector (coverage). This is proxying relevance at a sectoral level, but it is not necessarily equivalent to the aggregation of the relevance scores for each statistic in a sector, since it could happen that some relevant information is not covered by any statistic. There is also a question seeking the assessment of the quality of several dissemination channels (press releases, yearbooks, short-term newsletters, web pages and individual requests) at the sectoral level. All the assessments in the Structured Consultation are made according to an ordinal scale: very low (1); low (2); medium (3); high (4); and very high (5).

With respect to the selection of the sample, the difficulties with this type of users' satisfaction survey are well known. Considering the main objective in the Structured Consultation was to evaluate quality (and its components) and that in order to do this substantial knowledge of the official statistics is required, a purposive sample was used in which participants were selected from those proposed by the members of the SHSC Working Group as experts for a particular sector and with good prospects of collaboration. There are of course other alternative ways of selecting the sample, such as for instance a random sample of those making specific requests for information to the producers of official statistics but, given the data availability and the emphasis on having people with knowledge of the statistics, we ended up with a purposive sample. If possible, as a pending task for future research, it could be informative to compare the results of both approaches.

The sample of participants can be classified in three different groups: users with a high level of expertise, not coming from neither the Central nor the Regional Administration; users and experts from the Central Administration, who could also be producers (in particular the ministries and other public institutions) of some of the statistics; and users and experts from the Regional Administration, who are mainly users of the official statistics of the Central Administration, but are also producers of some regional information and collaborate in the production of some official statistics at national level. The questionnaires were sent as an Excel spreadsheet by e-mail to the participants, some of them receiving more than one questionnaire, each one including several single statistics. The data collection was carried out between June and July 2010. Table 1 presents the distribution of participants.

Table 2 summarises basic information about the questionnaires and the observations according to the three types of participants we considered. In total, 717 questionnaires (corresponding to sectors) were sent to the 236 participants with a response rate of 88.6%, and 599 of them were completed and included in the empirical analysis, generating 4,711 valid observations (complete assessments of single statistics), that is, on average each

	Number of Participants
Outside the Administration	130
Universities	79
Unions and employers' organisations	20
Other (media, research institutions etc.)	31
Central Administration	75
INE	32
Ministries	28
Other (agencies, institutes etc.)	15
Regional Administration	31
Regional Statistical Offices	17
Other (regional ministries, regional agencies etc.)	14
Total	236

Table 1. Distribution of participants (users and experts) in the consultation

	Sent question- naires	Received question- naires	%	Valid question- naires	%	Valid obser- vations
Outside the Administration	232	165	71.1	143	61.6	1,216
Central Administration	155	140	90.3	135	87.1	1,275
Regional Administration	330	330	100.0	321	97.2	2,220
Total	717	635	88.6	599	83.5	4,711

Table 2. Distribution of the number of questionnaires and observations

participant answered 2.5 valid questionnaires corresponding to 20 single statistics. It is important to note the significantly lower response rate among users outside both the Central and the Regional Administration.

3. Evidence from the Structured Consultation

In Tables 3 and 4 we report the basic results of the Structured Consultation: the distribution and the average of the scores for the different dimensions of quality and the dissemination channels (Table 3), and compared for the three types of participants in the consultation (Table 4). It is important to note that traces of straightlining in the responses, that is, not differentiating between the response categories for the dimensions and that of the overall assessment, have been observed for about 20 per cent of the sample. However, correction for this does not affect the empirical results significantly.

Based on the evidence from Tables 3 and 4, we can make the following comments:

- The frequency distributions of the different variables show that the mode for all the assessments is the score "high", except for the dimension Relevance which is "very high". All the average scores are between 3 ("medium") and 4 ("high"), again with

	Very low	Low	Medium	High	Very high	
	1	2	3	4	5	Average
Coverage	1.4	7.1	35.4	51.6	4.5	3.51
Quality of statistics	0.7	5.5	28.9	49.5	15.2	3.73
Quality dimensions						
Relevance	0.7	4.4	15.8	36.4	42.7	4.16
Accuracy and Reliability	1.5	6.6	27.5	47.2	17.2	3.72
Timeliness and Punctuality	1.8	8.2	25.7	40.8	23.5	3.76
Coherence and Consistency	1.9	8.0	28.6	43.1	18.4	3.68
Accessibility and Clarity	2.2	6.9	23.6	44.5	22.8	3.79
Dissemination channels						
Web page	3.8	5.7	21.6	43.2	25.7	3.81
Individual request	11.6	6.8	21.0	37.6	23.0	3.54
Press releases	7.8	15.8	23.9	37.0	15.5	3.37
Yearbooks	7.6	13.5	28.1	39.9	10.8	3.33
Short-term newsletters	7.4	13.5	34.7	35.5	8.9	3.25

Table 3. Frequencies (%) and averages of the scores of the different quality variables

Note: Mode in bold type

		Users not	Central	Regional
	Total	in Admin.	Admin.	Admin.
Coverage	3.51	3.40	3.76	3.44
Quality of statistics	3.73	3.63	3.90	3.69
Quality dimensions				
Relevance	4.16	4.19	4.36	4.03
Accuracy and Reliability	3.72	3.59	3.95	3.66
Timeliness and Punctuality	3.76	3.60	3.83	3.81
Coherence and Consistency	3.68	3.56	3.76	3.70
Accessibility and Clarity	3.79	3.69	3.93	3.76
Average of dimensions	3.82	3.72	3.97	3.79
Dissemination channels				
Web page	3.81	3.81	3.95	3.37
Individual request	3.54	3.36	3.92	3.38
Press releases	3.37	3.14	3.63	3.38
Yearbooks	3.33	3.30	3.71	3.16
Short-term newsletters	3.25	3.21	3.63	3.07
Average of channels	3.46	3.36	3.77	3.27

Table 4. Average scores by type of participants in the consultation

the exception of that of relevance (4.16). The perceived quality of the official statistics (3.73) is higher than the assessment of the coverage for the sectors (3.51) and also than that of dissemination (3.46 if measured as the average of the scores for the different channels). In fact, almost two thirds of the participants rate the quality of the statistical information as "high" or "very high".

- The average score of the assessment of the extent to which official statistics meet users' needs at a sectoral level (coverage) is significantly below the average score of relevance (3.51 vs 4.16), which also refers to meeting users' needs but for a single statistic. In fact, the score of coverage is also lower than that related to quality of the statistical product (3.51 vs 3.73). This significant difference between the scores of coverage and relevance shows that although available official statistics adequately meet the requirements of the users in those fields mentioned, there is still a lack of statistical information at the sectoral level. Coverage may be poor even if each of the statistics is highly relevant.
- The average score of the five dimensions of the quality of statistical information (3.82) is above the overall average (3.73). According to a t-test for the equality of the means of two variables, the differences of these scores with respect to the overall average are statistically significant, except for Accuracy and Reliability. In any case, except for Relevance, the averages of these scores and that of the overall assessment appear to be quite homogenous given that they do not differ more than two decimal points. Three dimensions (Relevance; Timeliness and Punctuality; Accessibility and Clarity) have average scores of above 3.73.
- The assessment of quality of dissemination by channel (contrary to what was observed in the assessment of the quality of statistical information) is clearly heterogeneous. The average value of the scores of the different channels (3.46), see Table 4, is derived from a positive evaluation of web pages (3.81), a medium-high evaluation of

individual requests (3.54) and less positive ratings for the other channels: press releases (3.37), yearbooks (3.33) and newsletters (3.25). It is interesting to note that the accessibility and clarity in the dissemination of information, as a dimension of the quality of statistical products, has a similar average to the web page channel (both are valued at around 3.8). Therefore, the quality of access does not seem to be significantly influenced by the less positive ratings of the other channels, probably as a consequence of the fact that, given the profile of the participants in the consultation, they rely more on those channels that have a higher evaluation (web pages and individual requests).

- As mentioned in Ehling and Körner (2007), evidence from Table 4 shows that different types of users and experts of official statistics perceive product quality differently. The average scores for all the variables are higher for the group of participants from the Central Administration, who also have a profile of producers. On the other hand, except for the assessment of relevance, the average scores of quality dimensions for the group of users not in the Administration are the lowest. Almost all the differences between the averages among the different groups of participants are statistically significant (t-test).
- The differences between the averages of the score of Relevance for the Central Administration and the Regional Administration are more important than for the other dimensions. This could be explained by the fact that those participants belonging to the Regional Administration are mainly users of the official statistics produced by the Central Administration, and they have a genuine interest in the territorial information at a more disaggregated level than that of the official figures. This is also corroborated at the sectoral level. The average of the coverage variable is quite similar between users and participants from the Regional Administration and is substantially different from the participants of the Central Administration.
- The genuine interest of the Regional Administration in the territorial information could also explain the substantial difference between the average scores for accuracy and reliability. The rating of these characteristics of the quality of the statistical product worsen when we consider more disaggregated geographical areas. However, since users not in the Administration also have a low score (3.59), part of this observed difference could be explained by some overrating of this dimension by the Central Administration, given that they are also responsible for the statistical production.
- It must be pointed out that the results for the total sample in Table 4 do not change significantly if we weight the observations of each group of participants differently. For instance, if we calculate the total values as the average of the estimates for each group, rather than the average of all responses, since regional administrators reported for many more statistics than other groups, then the average total scores are: 3.74 (overall assessment of quality), and 4.19, 3.73, 3.75, 3.67 and 3.69 (scores for the quality dimensions respectively). This pattern of different weights having no relevant effects applies to all the estimates in the article.
- When considering participants' assessments of the dissemination channels, it can be observed that the contents of the web page are scored highly by all three groups. Something similar occurs for the individual requests, although the score of users not in the Administration is significantly lower than that of the web page. This fact may be reflecting the difficulties of accessing microdata in some cases, in particular

administrative data. Finally, the evidence from the score of press releases seems to indicate that they are aimed more at users of the Regional Administration, who produce some regional data.

4. The Relationship Between the Quality of the Statistical Product and Its Dimensions

The main objective of this article is to try to evaluate how the different dimensions of the quality of statistical products influence the overall assessment, that is, to quantify what weight each dimension has when comprising the overall assessment of quality. In that regard, the simplest and most intuitive way of relating a variable (overall assessment) and its components (the different dimensions), given that they have the same scale (1-5), is by interpreting the overall assessment as an average of the scores of the different dimensions. This simplest interpretation of how the overall assessment is formed implies an equal weight for each of the dimensions.

We can compare the unweighted average scores of the five dimensions for all the participants in the consultation to the reported overall assessment. These averages are the result of calculating the average of the scores for the dimensions and recoding them on a scale of 1 to 5 by rounding the value of this average: being less than 1.5 (1); between 1.5 and 2.5 (2); between 2.5 and 3.5 (3); between 3.5 and 4.5 (4); and more than 4.5 (5). We obtain the result that in 82.6% of the cases the average prediction coincides with the observed assessment; in 13.5% of the cases it is higher, whereas in only 3.9% of the cases it is lower. This asymmetry can be interpreted as a downwards "correction" of the prediction, which can be thought of as subtracting a constant term from the average.

To provide further evidence about the type of relationship we expect to observe between overall quality and its components, we examine the correlation coefficients between these variables, reported in Table 5. It can be observed that the correlation coefficients between the different dimensions are positive and that they are above 0.5 except for the cases involving the relevance dimension (around 0.35). Furthermore, when looking at correlations with the overall assessment a similar pattern emerges, that is, correlations above 0.7 except for the relevance dimension (0.52). This shows that the influence of the dimensions will probably not be uniform and that of relevance will be lower than those

Tuble 5. Correlation matrix of the assessment	variabic.	5					
	С	Q	R	A-R	T-P	C-C	A-C
Coverage (C)	1.00						
Quality (Q)	0.44	1.00					
Relevance (R)	0.16	0.52	1.00				
Accuracy and Reliability (A-R)	0.28	0.75	0.35	1.00			
Timeliness and Punctuality (T-P)	0.34	0.71	0.34	0.51	1.00		
Comparability and Coherence (C-C) Accessibility and Clarity (A-C)	0.35 0.34	0.79 0.72	0.38 0.36	0.64 0.54	0.60 0.61	1.00 0.68	1.00

Table 5. Correlation matrix of the assessment variables

Note: Correlations are calculated for the average values of the assessments of the quality dimensions by each participant in the consultation (not for each single statistic), given that the assessment of coverage is unique for each participant and sector.

of the others. On the other hand, when looking at correlation with the coverage assessment, correlations are much lower, reflecting the fact that the coverage assessment implicitly takes into account those needs in a sector which are not met by the actual statistics (those for which we evaluate the quality components).

By using multidimensional scaling (Kruskal and Wish 1978), we can transform the correlations (proximities or similarities) between the seven assessment variables (objects) in the matrix of Table 5 into seven vectors in \mathbb{R}^N , usually N equal to 2 or 3, preserving the "hidden structure" of the data in terms of similarities and where the output is a spatial map, in our case a two-dimensional spatial map.

The seven points in Figure 1 correspond to the seven vectors in \mathbb{R}^2 , the elements of which correspond to *Dimension 1* and *Dimension 2*. Moreover, the smaller the correlation (proximity) between two assessment variables, the further apart the corresponding points (vectors) should be on the map, facilitating the visualisation of the structure of similarities of the data. In particular, in our case we identify five assessment variables that are really close (overall quality, accuracy and reliability, timeliness and punctuality, comparability and coherence, and accessibility and clarity). Since we are interested in measuring how the different components of quality contribute to the perceived overall quality, the structure of the spatial map leads us to expect that the four components included in this group will have a more substantial influence on the overall assessment than relevance.

All the previous evidence seems to indicate that a simple and intuitive mechanism such as a weighted average (to capture the potentially different influence of each component) corrected by a constant term (to take into account the overprediction which was suggested by the use of the simple average) can approximate the relationship between the overall assessment of quality and its components, that is, a mechanism based on a regression model with a constant term where the dependent variable is the score of the quality of the statistical product, the regressors are the scores of the five dimensions of quality defined in the European Code of Practice, and its coefficients are restricted to add one (weighted average).

We must point out that nonlinear specifications, such as Cobb-Douglas (a weighted geometric average with a constant term) or Translog functions, which allow for an



Euclidean distance model

Fig. 1. Multidimensional scaling for the assessment variables

interpretation as a production function of the relationship between the overall assessment and its components, were tried and none of these alternatives could offer a relevant better fit than the simple linear relationship we propose from a practical point of view. If a simple model can offer a similar fit capacity to more complex models, while at the same time facilitating the interpretation of the results, the Occam's razor principle suggests that the selection of the more parsimonious formulation is appropriate.

Using the acronyms of the variables in Table 5, the estimated model has the following specification:

$$Q_{ij} = \beta_0 + \beta_1 R_{ij} + \beta_2 (A - R)_{ij} + \beta_3 (T - P)_{ij} + \beta_4 (C - C)_{ij} + \beta_5 (A - C)_{ij} + u_{ij}$$
(1)

where indexes *i* and *j* correspond to the participant and the single statistics, respectively, *u* is the error term and the βs are the coefficients, which can be interpreted as the expected value of the weights in a random coefficient model (Swami and Tavlas 1995), where $\beta_{k,ij} = \beta_k + \eta_{k,ij}$, k = 0, 1, ..., 5, $\beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 = 1$, and η is a random variable with a zero mean, that is, the weights are different for each participant and each statistic, although the $\beta_{k,ij}$ coefficients cannot be estimated.

The model in Equation (1) is estimated by OLS with 4,711 observations. The estimation results are presented in Table 6. Those corresponding to the whole sample (Total) are in the first column, whereas from the second to the fourth column we report those corresponding to the groups of participants we are considering: users not in the Administration (U), Central Administration (C.A.) and Regional Administration (R.A.), respectively. By doing this we are allowing for some variability in the weights depending on the type of user we consider.

The results in the first column show that this simple and intuitive model (a corrected weighted average) has a substantial explanatory power, almost 80% of the variability of the dependent variable. Moreover, the significance of the coefficient of the constant term corroborates the correction we mentioned when looking at the descriptive analysis with the expected (negative) sign, that is, the weighted average is corrected downwards when configuring the overall assessment.

At the same time, the coefficients of the components (the weights) are statistically different from 0.2, which would be the weight for each component in the case of a simple average, meaning that not all the components have the same influence in assessing overall quality. Note that the sum of the estimated coefficients (weights) without imposing the constraint of the sum being equal to one would be 1.0188, 1.0278, 1.0010, and 1.0110 for the four models in Table 6, respectively, that is, the sums are practically equal to one.

In particular, accuracy and reliability is the dimension with the highest weight (23.2%), whereas relevance has the lowest (15.4%), as we expected from the descriptive analysis of the previous section. There are almost eight percentage points of difference between both weights, which implies accuracy and reliability receive more than 50 per cent more weight than relevance, that is, the differences are not only statistically significant but also practically relevant. This feature of different weights for each dimension is particularly relevant when considering and analysing the trade-offs between the dimensions, an issue that is becoming increasingly important, as mentioned in Ehling and Körner (2007). But this point cannot be interpreted, for instance, as a recommendation to prioritise accuracy

	Total Coef. (t-stat.)	Users not in Admin. <i>Coef.</i> (<i>t-stat.</i>)	Central Admin. Coef. (t-stat.)	Regional Admin. Coef. (t-stat.)
Relevance (β_1) Accuracy and Reliability (β_2) Timeliness and Punctuality (β_3) Coherence and Comparability (β_4) Accessibility and Clarity (β_5) Constant (β_0) Adjusted R ² Sample size H ₀ : $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0.2$ F statistic, p-values in brackets H ₀ : $\beta_{k,U} = \beta_{k,CA} = \beta_{k,RA}$ F statistic, p-values in brackets <i>Relevance</i> <i>Accuracy and Reliability</i> <i>Timeliness and Punctuality</i> <i>Coherence and Comparability</i> <i>Accuracy and Comparability</i> <i>Accuracy and Comparability</i> <i>Accuracy and Comparability</i> <i>Accessibility and Clarity</i> H ₀ : $\beta_{k,U} = \beta_{k,CA} = \beta_{k,RA}$ F statistic, p-values in brackets	$\begin{array}{c} 0.154 \ (19.37) \\ 0.232 \ (23.00) \\ 0.182 \ (21.17) \\ 0.224 \ (21.30) \\ 0.208 \ (22.35) \\ 0.208 \ (22.35) \\ 0.787 \\ 4711 \\ 12.11 \ (0.000) \end{array}$	$\begin{array}{c} 0.129 \ (8.59) \\ 0.243 \ (12.61) \\ 0.162 \ (10.13) \\ 0.240 \ (11.50) \\ 0.226 \ (13.18) \\ 0.226 \ (13.18) \\ 0.226 \ (13.18) \\ 0.811 \\ 1216 \\ 9.22 \ (0.000) \end{array}$	$\begin{array}{c} 0.124 \ (8.15) \\ 0.201 \ (10.94) \\ 0.228 \ (14.76) \\ 0.251 \ (12.99) \\ 0.196 \ (9.59) \\ 0.196 \ (9.59) \\ 0.811 \\ 1275 \\ 8.09 \ (0.000) \\ 8.09 \ (0.000) \\ 1.77 \ (0.183) \\ 5.31 \ (0.003) \\ 1.97 \ (0.140) \\ 0.73 \ (0.480) \\ 0.73 \ (0.480) \\ 0.73 \ (0.001) \\ 3.35 \ (0.001) \end{array}$	$\begin{array}{c} 0.179 \ (15.29) \\ 0.241 \ (15.85) \\ 0.171 \ (12.30) \\ 0.205 \ (13.02) \\ 0.204 \ (14.61) \\ 0.204 \ (14.61) \\ 0.749 \\ 0.749 \\ 2220 \\ 3.36 \ (0.009) \end{array}$
Note. The standard errors used in the calculation of the	t-statistics are calculated with	a robust estimator to allow for a no	mscalar variance-covariance m	atrix of the errors

Table 6. Estimation results for the regression model

variance matrix of the errors. ŝ vallalle IOT & nonscalar esumator to allow calculated with a rodust t-statistics are nne calculation of used in the Note: The standard errors and reliability at the expense of relevance. In fact, relevance is the basis for other quality dimensions and to some extent is a precondition for being a user at all.

The results for the three groups of participants confirm what has been pointed out by several authors (e.g., Ehling and Körner 2007), namely that different users have a different perception of quality and assign a different level of importance to the dimensions. As shown in Table 6, we can reject the null hypothesis of a unique structure of weights (coefficients) for the three groups of participants, as well as the hypothesis that for each group the weights for the different dimensions are the same (0.2).

In particular, the most significant differences between the weights for the different groups correspond to Relevance and Timeliness and Punctuality. Relevance has the lowest coefficient out of all three groups, but it is given significantly higher weight by the Regional Administration than by the Central Administration and users not in the Administration. Having the lowest weight could be a consequence of the fact that, as mentioned by Brackstone (1999), relevance is a dimension which should be considered across the whole output of a statistical office rather than per statistic, as we do in the exercise. The higher weight in the Regional Administration group can be explained by its special interest, as a user, in acquiring more detailed territorial information.

The weight of Timeliness and Punctuality is significantly higher for Central Administration. In fact, this dimension as well as Coherence and Comparability have the higher weights in this group, whereas Accuracy and Reliability occupies the third place in terms of the magnitude of the weights; by contrast, this dimension has the highest weight both for users not in the Administration and Regional Administration. This can be explained by the fact that Central Administration participants also have the role of producers, and this dimension has traditionally been associated with quality and essentially been under the control of producers. Additionally, in recent years and because of (among other things) the economic crisis and the need for data for international comparisons, increasing attention has been paid to issues related to timeliness and punctuality as well as to comparability. This has been translated into a closer association of these two dimensions with quality from the producers' side. The same conclusion is reached if we look at the proportion of coincidences between the score of one dimension and that of the overall assessment of quality. In the case of Timeliness and Punctuality and Comparability and Coherence, the Central Administration presents the highest proportions, whereas in the case of Accuracy and Reliability it has the lowest.

Finally, there are no significant differences between the participant groups in the case of Accessibility and Clarity. In principle, this is also a dimension that should be evaluated for the whole set of statistics produced by a statistical office rather than for one individual statistic, but in our case, since not only the statistical office (INE) is producing official statistics (ministries and other public institutions also do so), the importance of the weight could be capturing some heterogeneity in the way the different producers are making the information accessible and clear.

5. Conclusions

Quality has become a key concept in official statistics. At the same time however, this is a concept which can refer to different aspects (product, process) and which can be evaluated

from different viewpoints (users, producers) and with different tools (reports, users' satisfaction surveys, peer reviews).

In this article we have approached the evaluation of the quality of the statistical products from the users' side and considered the characteristics (dimensions or components) of the statistical product as they are described in the European Code of Practice. The evaluation has been performed using the information generated by the structured consultation carried out in 2010 by the SHCS Working Group in Spain when preparing the draft of the proposals and recommendations for the Spanish Multiannual Statistical Programme 2013–16. In this consultation both users not in the Administration and users in the Administration, who are also producers, were included in a purposive sample with the objective of evaluating the official statistics included in the Annual Statistical Programme 2010 by answering a questionnaire for the assessment of the different dimensions of quality.

Additionally, we have included evidence of the importance of the different dimensions of quality in order to assess the overall perceived quality of the statistical products. In order to do so, we have followed a simple and intuitive approach based on regression analysis to estimate the weights of a weighted average of the different components plus a constant term.

The main conclusions of this study referring to the evaluation of the quality of the Spanish official statistics can be summarised as follows:

- The overall assessment of the quality of individual official statistics in Spain is "high" or "very high" for 64.7% of the participants in the structured consultation, and only in 6.3% of the cases the score was "very low" or "low". In fact, for all the different variables analysed the mode is "high" with the exception of relevance ("very high"). Relevance is the dimension with the highest score (above 4), whereas the other four dimensions of quality have very similar scores.
- The assessment of coverage of official statistics at the sectoral level is lower than that of the relevance dimension for individual statistics and that of the overall quality. This indicates that some users' needs are still not fully satisfied in some fields (sectors), showing that there is still a lack of information at the sectoral level.
- Traditional dissemination channels have relatively low scores compared to the web page channel.
- There are significant differences between the scores for the three groups of participants we consider (users not in the Administration, Central Administration and Regional Administration). In general, the highest scores are found for Central Administration and the lowest for users not in the Administration.

The main conclusions regarding the weight of the quality dimensions when assessing overall quality can be summarised as follows:

- A weighted average of the five quality dimensions we have considered (those in the European Statistics Code of Practice) plus a constant term is a fairly good representation of how dimensions are taken into account when making an overall assessment of perceived quality of official statistics. All the dimensions make a significant contribution for all the different groups of participants we have considered in the consultation.

- The contribution of the different dimensions to the overall assessment of the quality of the statistical products is not uniform, that is, the weights are significantly different between dimensions. Accuracy and Reliability is the dimension with the highest weight in the model estimated for the whole sample, although this ranking and the value of the weights differ depending on the type of participants considered. In future work we plan to analyse whether there are significant differences in the weights across groups of statistics (e.g., by sector, short term vs. structural).
- Relevance is the dimension seen to have the highest score but is given the lowest weight in judging overall quality as perceived by users. This may reflect the fact that the participants in the structured consultation assume that the statistics they evaluate are relevant, so they do not focus too much on relevance in determining overall quality.

6. References

- Brackstone, G. 1999. "Managing Data Quality in a Statistical Agency." Survey Methodology 25:139-149.
- Brackstone, G. 2001. "How Important is Accuracy?" In Proceedings of Statistics Canada Symposium 2001, "Achieving Data Quality in a Statistical Agency: A Methodological Perspective". Available at: http://www.statcan.gc.ca/pub/11-522-x/2001001/session24/ 6311-eng.pdf (accessed February 4, 2013).
- Castles, I. 1991. "Responding to Users' Needs." *Journal of the Royal Statistical Society* (*Series A*) 154:6–10.
- Ehling, M. and T. Körner. 2007. *Handbook on Data Quality Assessment Methods and Tools*. European Commission, Eurostat.
- European Statistical System 2001. "Quality Declaration of the European Statistical System." Statistical Programme Committee, September 20, 2001. Available at: http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/DECLARATIONS.pdf (accessed May 20, 2014).
- European Statistical System 2011. "European Statistics Code of Practice." European Statistical System Committee, September 28, 2011. Available at: http://epp.eurostat.ec. europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-32-11-955 (accessed May 20, 2014).
- Eurostat 2002. "Quality in the European Statistical System. The Way Forward". Luxembourg: Office for Official Publications of the European Communities.
- Eurostat 2003. "Methodological Documents Definition of Quality in Statistics." Sixth meeting of the Working Group "Assessment of Quality in Statistics", Luxembourg, Available at: http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/ ess%20quality%20definition.pdf (accessed May 20, 2014).
- Fellegi, I.P. 1991. "Maintaining Public Confidence in Official Statistics." *Journal of the Royal Statistical Society (Series A)* 154:1–6.
- Giovannini, E., J. Oliveira Martins, and M. Gamba. 2009. "Statistics, Knowledge and Governance." *Statistika* 6:471–490.

- Kruskal, J.B. and M. Wish. 1978. "Multidimensional Scaling." Sage University Paper series on Quantitative Applications in the Social Sciences 07-011. Beverly Hills and London: Sage Publications.
- Nikolaidis, I. 2012. "Users' Perceptions of Statistics for Improvement Actions." Paper presented at the European Conference on Quality in Official Statistics, Athens, May 2012.
- Statistics Canada 2009. Corporate Business Plan. 2009/10 to 2011/12. Statistics Canada.
- Swami, P.A.V.B. and G.S. Tavlas. 1995. "Random Coefficient Models: Theory and Applications." *Journal of Economic Surveys* 9:165–196.
- United Nations Economic Commission for Europe 1992. *Fundamental Principles of Official Statistics*. Available at: http://www.unece.org/stats/archive/docs.fp.e.html (accessed May 20, 2014).
- United Nations Statistical Commission 2006. Report of the Secretary-General on Principles Governing International Statistical Activities. Report on the Thirty-Seventh Session (March 6–10, 2006). Available at: http://unstats.un.org/unsd/statcom/doc06/2006-13e-Principles.pdf (accessed May 20, 2014).
- Vale, S. 2010. "Statistical Data Quality in the UNECE. 2010 Version." Available at: http://unstats.un.org/unsd/dnss/docs-nqaf/UNECE-Quality%20Improvement%20 Programme%202010.pdf (accessed May 20, 2014).
- Working Party of the Royal Statistical Society 1991. "Official Statistics: Counting with Confidence." *Journal of the Royal Statistical Society (Series A)* 154:23–44.
- Zilhao, M.J. and M. Ribeiro. 2012. "Measuring Client's Satisfaction the Integrated Management System of the Post-Service Satisfaction Survey at Statistics Portugal." Paper presented at the European Conference on Quality in Official Statistics, Athens, May 2012.

Received October 2012 Revised February 2014 Accepted March 2014