A novel technique based on the *Law of Comparative Judgment* for quality-related problems

Fiorenzo Franceschini, Domenico Maisano

Politecnico di Torino (Dept. of Management and Production Engineering), Corso Duca degli Abruzzi 24, 10129, Turin, Italy

ABSTRACT
A relatively diffused problem in the quality engineering/management field is that in which a set of judges express their individual (*subjective*) judgments on a specific *attribute* of some *objects*, and these judgments have to be fused into a *collective* one.
The goal of this paper is to develop a new technique that combines the Thurstone's *Law of Comparative Judgment* with an *ad hoc* response mode based on preference orderings. This technique allows to express the collective judgment of the objects on a *ratio* scale and is applicable to a variety of practical contexts.

Corresponding author: Domenico Maisano, e-mail: domenico.maisano@polito.it

1. INTRODUCTION

A common problem in the field of quality engineering/management is that in which a set of judges express their individual (subjective) judgments on a specific attribute of some objects and they have to be fused into a collective one [1], [2], [15]. Possible examples can be: (i) the fusion of individual customer expectations on a set of product requirements, or (ii) the fusion of judgments by reliability and maintenance engineers on the severity of a set of possible process failures.

The scientific literature encompasses a plurality of fusion techniques, which differ from each other for (at least) three features [4]:

(1) The response mode for collecting the (input) judgments, e.g., expressed in the form of preference orderings, paired-comparison relationships, ratings, rankings, etc.;

(2) the underlying *principle* of the fusion technique, e.g., heuristic, mathematical/statistical, or fuzzy models;

(3) the type of (output) *collective judgment*, e.g., expressed in the form of rankings or ordinal/interval/ratio scale values.

The simplest fusion technique is probably that in which judges evaluate the objects using an ordinal response scale with a predetermined number of levels – e.g., five: *very low*, *low*, *intermediate*, *high*, and *very high* degree of the attribute. Then, for each object, the resulting scale levels are converted into conventional scores (e.g., 1, 2, 3, 4, 5 or -2, -1, 0, 1, 2, for five-level scales) and aggregated through central-tendency indicators.

In the field of *voting theory*, we recall the method by Borda [3], while in the field of multicriteria analysis decision-making, we recall the Analytic Hierarchy Process (AHP) [6]. For an exhaustive discussion of the existing techniques, we refer the reader to the vast literature and extensive reviews [1], [4].

Regardless of the peculiarities of the individual fusion techniques, a key element for their success is the simplicity of the response mode. Generally speaking, response modes that are relatively simple and understandable are more likely to be accepted, and the relevant data collection process is more likely to be accurate and reliable [3], [5]. For example, various studies show that comparative judgments of objects (e.g., "O_i is more preferred than O_j") are generally easier than judgments in absolute terms (e.g., "the degree of the attribute of O_i is low/high") [5], [8].

As to the typology of collective judgments, we note that they are often treated as if they were defined on a *ratio* scale, even when they actually are not; e.g., rankings or ordinal-scale values of the objects are often improperly "promoted", in the moment in which they are combined with other indicators through

weighted sums, geometric averages, or – more generally – statistics permissible to ratio-scale values [9], [10]. These promotions are potentially dangerous, as they can lead to significant distortions [3].

This article focuses on the Thurstone's *Law of Comparative Judgment* (LCJ) [11], i.e., a consolidated model that, starting from judgments expressed in the form of paired-comparison (ordinal) relationships of a set of objects, allows constructing an *interval* scaling of these objects. The LCJ is a milestone of psychometry and has stimulated the construction of numerous other models [7]. Despite its elegance and relative computational simplicity, the LCJ is not very popular in the quality engineering/management field, probably because of two major limitations:

(1) The response mode based on paired comparisons is inevitably tedious, especially when the number of objects tends to be high;

(2) The output is defined on an *interval* scale, which is not as powerful as a *ratio* scale.

The goal of this contribution is to develop a new fusion technique, which can be easily applied to typical problems in the quality engineering/management field. The proposed technique will be based on the combination of the canonical LCJ model with an *ad hoc* response mode based on preference orderings. Apart from the regular objects ($O_1, O_2, …, O_n$), these orderings will also include two *anchor* objects, to univocally identify the zero point and a conventional unit of the output scale. In this way, the output scale will (reasonably) be considered as a *ratio* one.

2. THURSTONE'S LCJ

Thurstone [11] postulated the existence of a *psychological continuum*, that is to say an abstract and unknown unidimensional scale, in which objects are positioned depending on the degree of a certain *attribute* – i.e., a specific feature of the objects, which evokes a subjective response in each judge. The position of one object is directly proportional to the degree of the attribute.

According to Thurstone, the position of a generic *i*-th object (O_i) is distributed normally: $O_i \sim N(\mu_i, \sigma_i^2)$, where μ_i and σ_i^2 are the unknown mean value and variance of that object's attribute. This distribution has been postulated to reflect the intrinsic variation between judges in positioning the objects in the continuum. In addition, the objects' variances are all equal ($\sigma_i^2 = \sigma_j^2 = … = \sigma^2$), and the intercorrelations (in the form of Pearson coefficients ρ_{ij}) between pairs of objects (O_i, O_j) are all equal too ($\rho_{ij} = \rho, \forall i, j$).

The application of the LCJ is based on five steps:

(1) A set of (*m*) judges ($J_1, J_2, …, J_m$) formulate their preferences for each object (O_i) versus every other object (O_j), considering all possible $C_2^n = n(n–1)/2$ pairs, *n* being the total number of objects. Preferences are expressed through relations of *strict preference* (e.g., "$O_1 > O_2$" or "$O_1 < O_2$") or *indifference* (e.g., "$O_1 \sim O_2$") [10]. Results are then aggregated into a frequency matrix (**F**). Precisely, for each judge who prefers O_i to O_j, the general element $f_{ij} \in [0, m]$ is incremented by one unit; if two objects are considered indifferent (i.e., "$O_i \sim O_j$"), f_{ij} and f_{ji} are both conventionally incremented by 0.5. In mathematical terms:

$$f_{ij} = |A| + 0.5 \cdot |B| \qquad (1)$$

where "| |" is the *cardinality* operator that corresponds to the number of elements of a set, and the two sets $A \subseteq \{J_k: "O_i > O_j"\}$ and $B \subseteq \{J_k: "O_i \sim O_j"\}$. Of course, the complementarity relationship $f_{ij} = m - f_{ji}$ holds.

(2) Next, the f_{ij} values are transformed into p_{ij} values, through the relationship:

$$p_{ij} = f_{ij} / m \qquad (2)$$

where p_{ij} represents the observed proportion of times that O_i was chosen over O_j. The p_{ij} values are aggregated into a proportion matrix (**P**).

(3) Next, p_{ij} values are transformed into z_{ij} values, through the relationship:

$$z_{ij} = \Phi^{-1}(1 - p_{ij}), \qquad (3)$$

$\Phi()$ being the cumulative distribution function of the standard normal distribution. The element z_{ij} represents a unit normal deviate, which will be positive for all values of $(1 - p_{ij})$ over 0.50 and negative for all values of $(1 - p_{ij})$ under 0.50.

In general, objects are judged differently by judges. However, if all judges express the same preference for each outcome, the model is no more viable (p_{ij} values of 1.00 and 0.00 would correspond to z_{ij} values of $\pm\infty$). A simplified approach for tackling this problem is to associate values of $p_{ij} \le 0.023$ with $z_{ij} = \Phi^{-1}(1 - 0.023) = 1.995$ and values of $p_{ij} \ge 0.977$ with $z_{ij} = \Phi^{-1}(1 - 0.977) = -1.995$. More sophisticated solutions to deal with this issue have been proposed [14].

(4) Next, the z_{ij} values related to the possible paired comparisons are reported into a matrix **Z**. The element z_{ij} is reported in the *i*-th row and *j*-th column. The relationship $z_{ji} = -z_{ij}$ holds, being unit normal deviates related to complementary cumulative probabilities (i.e., $p_{ji} = (1 - p_{ij})$).

(5) The *scaling* can be performed by (i) summing the values into each column of the matrix **Z** and (ii) dividing these sums by *n*. It can be demonstrated that the result obtained for each *j*-th column (x_j) corresponds to the unknown average value (μ_j) of the object's attribute, up to a p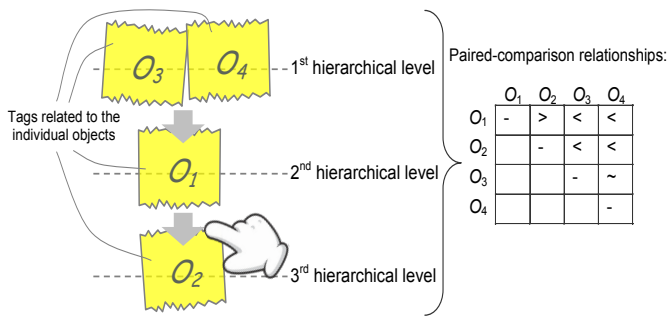ositive scale factor and an additive constant [12], [13]: $x_j = \sum_j (z_{ij}) / n = k_1 \cdot \mu_j + k_2$. In other words, the LCJ results into an *interval* scaling, i.e., objects are defined on a cardinal scale (*x*) with arbitrary zero point and unit [9], [10].

3. DESCRIPTION OF THE NEW TECHNIQUE

3.1. Introduction of preference orderings

A significant limitation of the canonical LCJ is that the direct formulation of paired-comparison (ordinal) judgments can be tedious and complex to manage, since much repetitious information is required from judges. Alternatively, judges can directly formulate preference orderings that are then turned into paired-comparison data. Precisely, each judge constructs a *linear* preference ordering, i.e., a chain of hierarchical levels containing the objects of interest, linked by arrows depicting the *strict preference* (">") relationship (see the example in Figure 1). Two or more objects in the same hierarchical levels are linked by the *indifference* ("~") relationship. We note that this

Figure 1. Practical technique to support the construction of preference orderings using tags and (indirect) determination of paired-comparison relationships.

response mode forces judges to be *transitive* (e.g., if "$O_1 > O_2$" and "$O_2 > O_3$", then "$O_1 > O_3$").

Even though the formulation of preference orderings may sometimes be less practical then the use of ordinal response scales (e.g., in the case of telephone or street interviews), the fact remains that it is less prone to the following problems: (i) ordinal scales tend to be used subjectively, as there is no absolute reference shared by all judges, and (ii) the number of categories in the ordinal scale (e.g., a five-level scale) may conflict with the real discriminatory power of judges.

3.2. *ZM* anchoring of the Thurstone's Scaling

Another limitation of the canonical LCJ is that the resulting (interval) scale is not "anchored" with respect to the (unknown) psychological continuum, in which objects are positioned, due to the arbitrary zero point and unit. This limitation makes the results of different scaling processes difficult to compare.

We have developed a new anchoring technique, denominated "*ZM*", based on the application of the LCJ, including two *anchor* objects among the regular ones (O_1, O_2, etc.): Z, i.e., an anchor object corresponding to the *absence* of the attribute of interest, and M, i.e., an anchor object corresponding to the *maximum-imaginable* degree of the attribute, consistently with the current technological and socio-economic development.

Likewise, the regular objects, Z and M are assumed to project a normal distribution on the psychological continuum, with unknown mean value and unknown variance, equal to that of the other objects (see Sect. 2). Each judge has to formulate a preference ordering of the objects, with two important requirements:

(1) Z should be positioned at the bottom of the preference ordering. In the case the attribute of another object is judged to be absent, that object will be considered indifferent to Z.

(2) M should be positioned at the top of the preference orderings. In the case the attribute of another object is judged to be the maximum-imaginable, that object will be considered indifferent to M.

Next, the Thurstone's scaling is performed and the resulting (interval) scale is transformed into a new one, defined in the conventional range [0, 100], through the following linear transformation:

$$\frac{y_j - 0}{100 - 0} = \frac{x_j - x_Z}{x_M - x_Z} \implies y_j = 100 \cdot \frac{x_j - x_Z}{x_M - x_Z} \quad (4)$$

where x_Z and x_M are the scale values of Z and M, resulting from the LCJ; x_j is the scale value of a generic j-th object, resulting from the LCJ; y_j is the relevant transformed scale value in the conventional range [0, 100].

The introduction of Z and M allows to anchor the LCJ scale (x) into a new scale (y) with a conventional unit and a zero point (which corresponds to the absence of the attribute); it is therefore reasonable to consider y as a *ratio* scale. We remark that setting the value of M to 100 is a normalization that makes the scale unit comparable with that ones obtained from other LCJ processes.

In the example in Figure 2, five judges (J_1 to J_5) have to compare four products (O_1 to O_4, i.e., the objects of interest), based on their *simplicity of use* (i.e., the attribute of interest). Precisely, each judge formulates a preference ordering of the four (regular) objects and two anchor objects (Z and M).

The introduction of Z and M increases the information content of preference orderings. For example, the information that the degree of an attribute is zero or the maximum-imaginable one is richer than the information that it is just lower or higher than the remaining ones. The price to pay for this information enrichment is the increased effort of judges, who should also consider the two dummy/anchor objects, envisaging their "absolute" meaning.

The preference orderings are then translated into paired-comparison relationships and the LCJ is applied (see Figure 3).

We have verified that the new anchoring technique provides results in line with those obtained from other existing techniques, such as that suggested by Torgerson's [14].

4. CONCLUSIONS

This paper proposed a new fusion technique that combines the canonical LCJ model with an *ad hoc* response mode based on preference orderings. Apart from the regular objects, these orderings will also include the two anchor objects Z and M. This allows representing the objects on a conventional *ratio* scale, included between 0 to 100, without any conceptually prohibited promotion. In addition, the response mode based on preference orderings is relatively practical and user-friendly.

A limitation of the proposed technique is that the introduction of the two anchor objects (Z and M) requires an additional
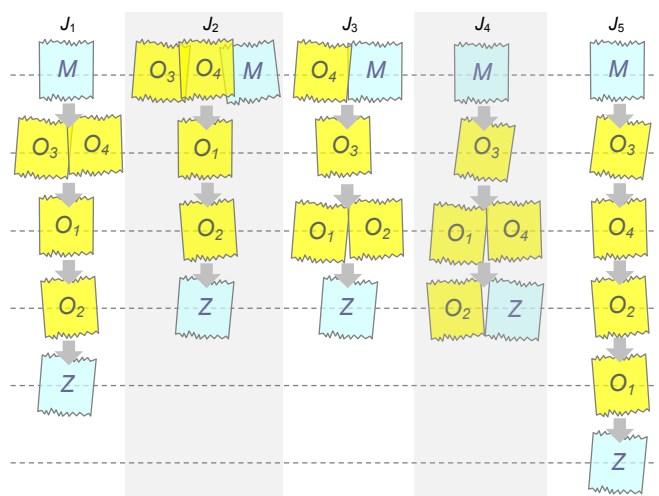

Figure 2. Example of preference orderings formulated by five judges (i.e., J_1 to J_5), including four regular objects (O_1 to O_4) and two anchor objects (Z and M).

(a) Judgements, paired-comparison relationships and f_{ij}, p_{ij}, z_{ij} indicators

Paired comparison	J_1	J_2	J_3	J_4	J_5	f_{ij}	p_{ij}	z_{ij}
(O_1, O_2)	>	>	~	>	<	3.5	0.70	-0.524
(O_1, O_3)	<	<	<	<	<	0	0.00*	1.995
(O_1, O_4)	<	<	<	~	<	0.5	0.10	1.282
(O_2, O_3)	<	<	<	<	<	0	0.00*	1.995
(O_2, O_4)	<	<	<	<	<	0	0.00*	1.995
(O_3, O_4)	~	~	<	>	>	3	0.60	-0.253
(O_1, Z)	>	>	>	>	>	5	1.00*	-1.995
(O_2, Z)	>	>	>	~	>	4.5	0.90	-1.282
(O_3, Z)	>	>	>	>	>	5	1.00*	-1.995
(O_4, Z)	>	>	>	>	>	5	1.00*	-1.995
(O_1, M)	<	<	<	<	<	0	0.00*	1.995
(O_2, M)	<	<	<	<	<	0	0.00*	1.995
(O_3, M)	<	~	<	<	<	0.5	0.10	1.282
(O_4, M)	<	~	~	<	<	1	1.00*	-1.995

(e) graphical representation of the scaling

decreasing degree of the attribute

increasing degree of the attribute

(b) matrix F

	O_1	O_2	O_3	O_4	Z	M
O_1	2.5	3.5	0.0	0.5	5.0	0.0
O_2	1.5	2.5	0.0	0.0	4.5	0.0
O_3	5.0	5.0	2.5	3.0	5.0	0.5
O_4	4.5	5.0	2.0	2.5	5.0	1.0
Z	0.0	0.5	0.0	0.0	2.5	0.0
M	5	5	4.5	4	5	2.5

(c) matrix P

	O_1	O_2	O_3	O_4	Z	M
O_1	0.50	0.70	0.00*	0.10	1.00*	0.00*
O_2	0.30	0.50	0.00*	0.00*	0.90	0.00*
O_3	1.00*	1.00*	0.50	0.60	1.00*	0.10
O_4	0.90	1.00*	0.40	0.50	1.00*	0.20
Z	0.00*	0.10	0.00*	0.00*	0.50	0.00*
M	1.00*	1.00*	0.90	0.80	1.00*	0.50
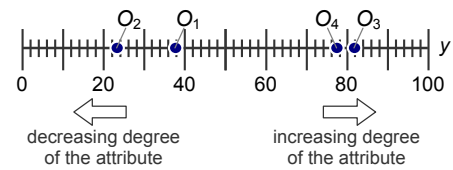
(d) matrix Z

	O_1	O_2	O_3	O_4	Z	M
O_1	0.00	-0.52	2.00	1.28	-2.00	2.00
O_2	0.52	0.00	2.00	2.00	-1.28	2.00
O_3	-2.00	-2.00	0.00	-0.25	-2.00	1.28
O_4	-1.28	-2.00	0.25	0.00	-2.00	0.84
Z	2.00	1.28	2.00	2.00	0.00	2.00
M	-2.00	-2.00	-1.28	-0.84	-2.00	0.00
Σ_j	-2.75	-5.23	4.96	4.18	-9.26	8.11
$x_j = \Sigma_j / n$	-0.46	-0.87	0.83	0.70	-1.54	1.35
y_j [0,100]	**37.5**	**23.2**	**81.9**	**77.4**	**0**	**100**

Notes:
Z is an anchor object denoting the zero preference level;
M is an anchor object denoting the maximum-imaginable preference level;
$n=6$ is the total number of objects, including Z and M;
(*)values of $p_{ij} \leq 0.023$ and ≥ 0.977 have been conventionally associated with $z_{ij} = 1.995$ and -1.995 respectively;
f_{ij} denotes the number of times that O_i is preferred to O_j;
p_{ij} denotes the proportion of times that O_i is preferred to O_j;
$z_{ij} = \Phi^{-1}(1 - p_{ij})$;
x_j is the (interval) scale value of the j-th object, resulting from the LCJ;
y_j is the x_j value transformed in the conventional range [0, 100] (transformation in Eq. 4).

Figure 3. Example of LCJ application to the preference orderings in Figure 2: (a) paired-comparison relationships, (b) matrix F, (c) matrix P, (d) matrix Z and resulting scaling, and (e) graphical representation of the scaling.

effort for judges.

Regarding the future, we plan to simplify the response mode, e.g., assuming that judges formulate preference orderings that include only a few most/least preferred objects [5], and to apply the proposed technique to real-life problems in the quality engineering/management field.

REFERENCES

[1] K. Coaley, An Introduction to Psychological Assessment and Psychometrics, 2nd edn., Sage, London, 2014, ISBN 978-1847874795.

[2] J.C. Krynicki, Introduction to soft metrology, Proc. of XVIII IMEKO World Congress, Sep. 17-22, 2006, Rio de Janeiro, Brazil.

[3] F. Franceschini, M. Galetto, M., D. Maisano, Management by Measurement: Designing Key Indicators and Performance Measurement Systems, Springer, Berlin, 2007, ISBN 978-3-540-73212-9.

[4] R.F. DeVellis, Scale Development: Theory and Applications, 4th edn., Sage, London, 2016, ISBN 978-1412980449.

[5] A.W. Harzing, J. Baldueza, ... and Y.K. Liang, Rating versus ranking: What is the best way to reduce response and language bias in cross-national research?. International Business Review, 18 (2009) pp. 417-432.

[6] T.L. Saaty, Decision making with the analytic hierarchy process. International journal of services sciences, 1(2008): 83-98.

[7] D. Andrich, Relationships between the Thurstone and Rasch approaches to item scaling. Applied Psychological Measurement, 2 (1978) pp. 451-462.

[8] G. Moors, I. Vriens, J.P. Gelissen, J.K. Vermunt, Two of a kind. Similarities between ranking and rating data in measuring values. Survey Research Methods, 10 (2016) pp. 15-33.

[9] F.S. Roberts, Measurement Theory: with Applications to Decision Making, Utility, and the Social Sciences, in: Encyclopedia of Mathematics and its Applications, vol. 7, Addison-Wesley, Reading, MA, 1979, ISBN 978-0521102438.

[10] S.S. Stevens, On the theory of scales of measurement. Science, 103 (1946) p. 2684.

[11] L.L. Thurstone, A law of comparative judgments. Psychological Review, 34 (1927) p. 273.

[12] F. Franceschini, D. Maisano, Prioritization of QFD customer requirements based on the law of comparative judgments. Quality Engineering, 27 (2015) pp. 437-449.

[13] L.L. Thurstone, L.V. Jones, The Rational Origin for Measuring Subjective Values. Journal of the American Statistical Association, 52 (1957) pp. 458-471.

[14] W.S. Torgerson, Theory and Methods of Scaling, Wiley, Oxford, England, 1958, ISBN 978-0471879459.

[15] J.J. Dahlgaard, G.K. Khanji, K. Kristensen, Fundamentals of total quality management, Routledge, London, 2008, ISBN 978-0748772933.