Worthiness Based Interpretation of Equi-distanced Performance Scales.

Giulio D'Epifanio
Facoltà di Scienze Politiche, Univ. di Perugia, via A.Pascoli, 06100 Perugia, Italy

Abstract: Within the paradigm of value-focusing thinking (Keeney, 1996), a goal-driven approach is proposed to score levels of ordinal performance scale. This approach, which would satisfy the requirement of "relevance to decision-maker's purpose", is based on a principle of worthiness. Such a principle is grounded on an intrinsic substantive meaning. Furthermore, it could be implemented standardized also over a reference statistical setup. At light of a certain sequential evaluation process, over hierarchy of binary goals, we interpret why praxis of equi-distanced scaling is "naive" and potentially misleading in performance evaluating.

Keyords: equi-distanced interval scale, level scoring, performance scaling

1. Introduction

Over an ordinal performance scale with $(L + 1)$ categories of response, to score levels is crucial in constructing indexes which are *"relevant to purpose"* of the decision maker (DM), within his conceptual evaluation design. Of course, in formal measurement theory over ordinal scales (e.g. see Kampen and Swyngedouw, 2000), DM's evaluation should be invariant with respect to any monotone transformation of the *basic quantification*, which marks ordinal categories on conventional numerical labels $0, 1, \ldots, l, \ldots, L$. But despite of formal invariance principles, in substantive evaluations, "basic" quantification appears rather unnatural, perhaps "naive", in that it would violate typical situations, for instance of increasingly stronger/weaker resistance against further satisfying of goals, over list of performance requirements. As a consequence, the common praxis among practitioners of using some conventional "equidistant interval scale" (e.g. in Likert-like scaling) seems difficult to justify on the basis of relevance of a mere "criterion of simplifying". On the other hand, even sophisticated model-based approaches (e.g. see Agresti, 2001) suffer for "lack of relevance" (see Saaty, 1980, pp. 238), whenever their rationale (e.g. maximization of likelihood, correlation, etc.) in technical measurements is poorly interpreted and few meaningful over the DM's conceptual evaluation design.

Within a broader framework, the specific question which is addressed by this work is the following. What does it means, in evaluation performances over an ordinal scale, that some version of "equidistant interval scale" is assumed? What criterion and formal working assumptions could explain such a praxis?

We present a criterion, for eliciting level scores of an ordinal performance scale, which is based on a *principle of intrinsic worthiness*. Then, such a criterion is implemented, standardized over a statistical population of interest for DM. Designed "ad hoc" over hierarchy of goals, within the paradigm of "value-focusing thinking" (Keeney, 1996), an operative method is outlined, which should be (by construction) "relevant to DM's purpose" whenever the substantive meaning of the "principle of intrinsic worthiness" was accepted. It could guarantee some *conditional scientific objectivity* degree, whenever measurements were accurately estimated given reference statistical setup. We focus then attention on a very particular case, where the reference population may be virtually associated to a certain sequential random process of binary choices, on which versions of equidistant interval scales could be interpreted.

2. A worthiness based level scoring

2.1. Ordinal performance scale formatting

Instance of performance is intended as the realization of a service (act, process, etc.) over a certain domain of cases (subjects, occasions, etc.). In evaluating "intrinsic worthiness", an instance of performance is processed against a hierarchy of binary goals

$$O_0 \preceq O_1 \preceq \; O_2 \preceq \ldots \preceq O_l \preceq \ldots, \preceq O_L \tag{1}$$

Here, O_0 is the "base-line" goal which is a tautology, that is a logical clause which is always satisfied; "$\preceq$" denotes the *Guttman's ordering* ($O_l \preceq O_{l+1}$ holds if and only if "if goal O_{l+1} is satisfied then also goal O_l has been satisfied"). Sequentially processed over sequence of goals (1), any instance of performance may be classified on $(L+1)$ ordinal categories. This process of sequential discrimination may be sketched on a binary-decision tree. The l-th goal O_l, $l := 1, \ldots, L$, is associated to that level specific test-sentinel which decides, for any instance which temporarily has been located at level $(l-1)$, whether or not also goal O_l is satisfied. If this goal is satisfied, the processed instance is allowed to pass over, by adding value, toward higher classification levels. Otherwise, it is definitively confined at the l-th category. Thus, an ordinal performance scale remains formatted, which, by design, should be relevant to DM's purpose

2.2. Principle of worthiness and level scoring

To numerically graduate ordinal levels, over hierarchy of goals (1) a principle of "worthiness" is proposed as follows: *the more the resistance to satisfy a goal, and so to pass over associated level, the more the worthiness credit due to the capability of that instance which is allowed to pass over.* Thus, at each stage whenever it is allowed to pass over, an instance of performance (subject) would gain the value which is intrinsic to the merit of winning against some designed resistance (difficulty). Such a principle may be statistically interpreted and then implemented over some reference (standard) population $\mathcal{P}^*$, of interest within the DM's conceptual evaluation framework. Given hierarchy of goals (1) over reference population $\mathcal{P}^*$, for transition from $(l-1)$-th to l-th ($l := 1, \ldots, L$) level, *the larger the statistical risk, for an instance which has been randomly drawn by population $\mathcal{P}^*$, of failing the l-level-specific objective O_l, the higher the "increment of worthiness credit", which is associated to an instance-type in $\mathcal{P}^*$, of passing over".*

Let Y denote numerical response under *basic quantification*, which used numerical labels $0, 1, \ldots, l, \ldots, L$ over ordinal categories of performance; let ${}_{O_l}\chi$ denote the binary indicator of satisfaction, on $\{0, 1\}$, for $l-$th goal O_l. Then, criterion above yields level-score increments:

$$\omega_l^* := \omega_l[\mathcal{P}^*] := Pr\{_{O_l}\chi = 0|_{O_{l-1}}\chi = 1; \mathcal{P}^*\} = \frac{Pr\{Y = l-1; \mathcal{P}^*\}}{Pr\{Y \geq l-1; \mathcal{P}^*\}} \geq 0, \quad (2)$$
$$l := 1, \ldots, L$$
$$\omega_0^* := Pr\{_{O_0}\chi = 0; \mathcal{P}^*\} = Pr\{Y \leq 0; \mathcal{P}^*\} = 0$$

A meaningful interpretation is that, *the larger the probability of failing goal O_l, given that O_{l-1} has been realized, the larger the (uni-directional) worthiness based distance of l-th category from previous $l-1$-th category*.

2.3. Worthiness based scoring

Sequentially processed over an ordinal-scale-formatting hierarchy of goals (1), any instance of performance i would receive final value $\omega_1^* + \omega_2^* \ldots + \omega_l^*$, whenever it remains confined at l-th category of Y. Thus, level-scores (2) will enter the following performance score:

$$Score(i; \mathcal{P}^*) = \sum_{l:=1}^{L} c(\mathcal{P}^*) \cdot \frac{Pr\{Y = l-1; \mathcal{P}^*\}}{Pr\{Y \geq l-1; \mathcal{P}^*\}} \cdot_{Y \geq l} \chi(i) \quad (3)$$

Here, $_{Y\geq l}\chi(i) := \begin{cases} 1 & \text{if } Y(i) \geq l, \\ 0 & \text{otherwise} \end{cases}$; c is a positive scale constant, which might depends, but non necessarily, on $\mathcal{P}^*$. Here, specific re-scaling of (2) may be crucial in meaningful interpreting of worthiness values. Notice that, for example in evaluating students, comparative ranking is invariant with respect to the choice of $\mathcal{P}^*$. Of course, rates of difference, among performance of subjects, due to worthiness would depend on choice of $\mathcal{P}^*$. But, given $\mathcal{P}^*$, they are invariant with respect to any re-scaling of (2). However, DM would expect that intrinsic value of worthiness is low/high whether/or not it is referenced against best/worst practices. Therefore, normalizing scores (3), over interval $[0, 1]$, so that $c(\mathcal{P}^*) = \{\sum_{l:=1}^{L} \frac{Pr\{Y=l-1;\mathcal{P}^*\}}{Pr\{Y\geq l-1;\mathcal{P}^*\}} \cdot_{Y\geq l} \chi(i)\}^{-1}$ might imply some type of self-referencing. Alternatively, "intrinsic worthiness scores" might be relativized, over interval $[0, 1]$, as follows:

$$RScore(i; \mathcal{P}^*) = \frac{Score(i; \mathcal{P}^*) - Score(m; \mathcal{P}_{BEST})}{Score(M; \mathcal{P}_{WORST}) - Score(m; \mathcal{P}_{BEST})} = \quad (4)$$
$$= \sum_{l:=1}^{L} \frac{1}{L} \frac{Pr\{Y = l-1; \mathcal{P}^*\}}{Pr\{Y \geq l-1; \mathcal{P}^*\}} \cdot_{Y \geq l} \chi(i)$$

Here, $\mathcal{P}_{BEST}$ and $\mathcal{P}_{WORST}$ denote, respectively, virtual reference set of "uniformly best practices" (where distribution of Y is $\delta_Y := (0, 0, \ldots, 1)$, so that $\omega_l = 0$, $l := 1, \ldots, L$) and the set of "uniformly worst practices" (where $\delta_Y := (1, 0, \ldots, 0)$, so that $\omega_1 = 1$ and, conventionally, $\omega_l = \frac{0}{0} = 1$, $l := 2, \ldots, L$). Thus, letting "m" and "M" to denote the instances which have, respectively, performance level $Y = 0$ (the lower) and $Y = L$ (the higher), $Score(M; \mathcal{P}_{WORST}) = c \cdot L$ is maximal and $Score(m; \mathcal{P}_{BEST}) = 0$ is minimal, so that their difference is maximized.

3. Worthiness based interpretation of equi-distance interval scales

Processed over hierarchy of goals (1), over statistical population $\mathcal{P}^*$, assume that transition probabilities remain constant, that is that: $q_l := Pr\{_{O_l}\chi = 1|_{O_{l-1}}\chi = 1\} = p$, $l := 1, \ldots, L$. Notice that probability distribution, over categories of Y, was provided by $p_0 := (1-p), p_1 := p(1-p), p_3 := p^2(1-p), \cdots, p_{L-1} := p^{L-1}(1-p), p_L := p^L$. Then, using (2), $\omega_l := \frac{p^{l-1}(1-p)}{p^{l-1}(1-p)+p^l(1-p)+\cdots+p^{L-1}(1-p)+p^L} = \frac{p^{l-1}}{p^{l-1}} \frac{1-p}{(1-p)+(1-p)p+\cdots+(1-p)p^{L-1}+p^L} = \frac{1-p}{(1-p)+(1-p)p+\cdots+(1-p)p^{L-1}+p^L} = (1-p)$, $l := 1, \ldots, L$ (conventionally, we would set $\frac{0}{0} = 1$). By normalizing now sum of level score increments to unity we would have $\omega_l = 1/L$, $l := 1, \ldots, L$, irrespective of value p. Therefore, level scores $\{0, \frac{1}{L}, \frac{2}{L}, \ldots, 1\}$ would be equivalent to those of ”basic quantification”, up to normalizing full attainment of performance on unity. Of course, any ”equi-distant interval” scale is compatible with some translation and/or re-sizing of ”basic quantification”. Therefore, any ”equi-distant interval” scale might be re-interpreted by thinking on some virtual ”random device” (reference population), which realizes a stationary ”without memory” sequential process of binary decisions, over hierarchy (1) of goal-tests, whenever probability (propensity of subject-type) p of passing over would remain constant. At light of this interpretation, since such situations rarely occur in practice, the common praxis among practitioners of using basic quantification should be considered as a *”naive quantification”*, which is potentially misleading in performance evaluating. But, it is also crucial in meaningful evaluation of performance, to avoid that even very different versions of ”equi-distant interval” scale, which are associated to differently p-sized process, may be confused. Thus, (4) would be recommended to even distinguish among different versions of equi-distanced scales.

References

Agresti A. (2001), *Categorical Data Analysis*, second edition, Wiley, New York

D'Epifanio G. (2007), Level Scoring of Ordinal Verbal Scales in Construction of Model-based Performance Indexes, in: *Atti della Riunione Scientifica ”Valutazione e Customer Satisfaction per la Qualità dei Servizi”, Roma 12-13 Aprile, Università di Tor Vergata*

Kampen J., Swyngedouw M. (2000), The Ordinal Controversy Revisited, *Quality and Quantity*, Volume 34, Number 1 February, Springer Netherlands

Keeney, R.L (1996), Value focused thinking: Identifying decision opportunities and creating alternatives, European Journal of Operational Research, 92, 537-549

Saaty L. Thomas (1980), *The Analytic Hierarchy Process. Planning, Priority Setting, Resource Allocation.*, McGraw-Hill, New York