# PARTIAL AVERAGE CROSS-WEIGHT EVALUATION FOR ABC INVENTORY CLASSIFICATION

Giannis Karagiannis

Professor, Department of Economics, University of Macedonia, 156 Egnatia Str.
Thessaloniki, 54006 GREECE; karagian@uom.edu.gr

ABSTRACT: In this paper, we propose an alternative overall measure, inspired by the notion of average cross efficiency, that summarizes achievements across different descending ordering schemes regarding the relative importance of the considered indicators in the Ng model. The proposed overall measure is equal to the arithmetic average of the maximum partial averages across all possible descending ordering schemes. It can also be obtained by using the average (across ordering schemes) of the estimated multipliers. We apply the proposed measure to the ABC inventory classification problem and we compare our results with those by three information theory based methods that may be used for the same purpose, namely the Shannon entropy, the distance-based and the maximizing deviations methods.

KEYWORDS: Composite indicators, Descending ordering schemes; Overall measure; ABC inventory classification

ACKNOWLEDGEMENT: I would like to thank two anonymous referees and the associate editor for constructive comments and suggestions

July 2018

# Partial average cross-weight evaluation for ABC inventory classification

## 1. Introduction

In performance evaluation literature, the performance of individual decision-making units (DMUs) may be evaluated under three appraisal schemes: self-appraisal, peer-appraisal or preference-appraisal. Self-appraisal refers to the case that each evaluated DMU is allowed to choose its own "value system", by means of the weights attached to each performance indicator, in order to show in the best possible light relative to other DMUs included in the assessment.[1] Each evaluated DMU can exaggerate its own advantages and at the same time downplay its own weakness to obtain the maximal possible evaluation score. Data Envelopment Analysis (DEA), introduced by Charnes, Cooper and Rhodes (1978), is the main operation research tool for conducting self-appraisal performance evaluation. On the other hand, peer-appraisal gives the right to every DMU to have a "say" about the evaluation of the other DMUs. In particular, each DMU takes into account the "value system" of all evaluated units (included itself) in assessing its own performance. Since each "value system" results in a different evaluation score, performance is gauged by the average of the efficiency scores obtained by using all DMUs' self-appraisal weights (see Sexton, Silkman and Hogan, 1986; Doyle and Green, 1994), which is called average cross efficiency. Lastly, in the case of preference-appraisal, *a priori* information provided by experts, stakeholders, or policy makers is incorporated into the evaluation process by means of a predetermined "value system". Their preferences about the importance of the considered performance indicators are reflected in a set of restrictions that reduce the weight flexibility of self-appraisal DEA (Dyson *et al.*, 2001; Angulo-Meza and Lins, 2002). These restrictions may either take the form of numerical limits on the weights of the considered performance indicators or provide a ranking of their relative importance, with the latter having the advantage of being simple and intuitive (Joro and Viitala, 2004).

Sometimes, however, experts, stakeholders, or policy makers cannot reach a consensus about the relative importance of the considered performance indicators. This turns into an issue as long as it affects the estimated scores and/or the final ranking of DMUs. Then, the choice of a particular ordering scheme becomes difficult and debatable especially if there is no *a priori* reason to weight the opinion of an expert, stakeholder or policy maker more than another's. In such cases, one may either use all acceptable ordering schemes to examine the extent of changes (if any) in the final ranking of the evaluated DMUs or alternatively, may try to obtain an overall performance measure summarizing achievements across different ordering schemes. A similar situation may arise in a different occasion: e.g., when an analyst conducts a sensitivity analysis on the resulting efficiency scores by examining the entire set of possible ordering schemes regarding the importance of the considered performance indicators. Then, it may also be useful to end up with a single metric that reflects performance under all different norms or "value systems".

The problem of deriving such an overall or synthetic performance measure has been handled so far by information theory methods. In particular, Fu, Wang and Lai (2015) and Zheng *et al*. (2017) used Shannon entropy to aggregate a DMU's evaluation scores obtained under alternative ordering schemes regarding the importance of the considered performance indicators while for the same purpose Fu *et al*. (2016) and Cao *et al.* (2016) employed a distance–based method and Wu *et al*. (2018) the weighted least-square dissimilarity method of Wang and Wang (2013).[2] Both of these are purely data-driven methods trying to exploit the information of the data itself. For example, the entropy method rates higher for overall performance the preference-appraisal evaluation scores with the relatively larger variation across DMUs and consequently, assigns to them a higher aggregation weight. If an ordering scheme results in evaluation scores that have almost no variation across DMUs then its aggregation weight tends to zero. On the other hand, the distance-based method rates higher for overall performance the preference-appraisal evaluation scores with the smaller deviations from the mean and as a result, it assigns to them a higher aggregation weight.

The objective of this paper is to propose an alternative overall or synthetic performance measure rooted to performance evaluation rather than to information theory. In particular, the proposed measure is inspired by the notion of average cross efficiency. The difference is however that in preference-appraisal case, we consider all possible ordering schemes as reflecting the different “value systems” while the entire set of DMUs determines the different “value systems” in peer-appraisal, where the notion of average cross efficiency was firstly used. In our case, each possible ordering scheme, expressing the opinion of an expert, stakeholder or policy maker, provides weights for the relative importance of the considered performance indicators by means of the estimated DEA multipliers. We then evaluate all DMUs using each time a different “value system” reflecting a particular ordering scheme. Since in most of the cases there is no *a priori* reason to weigh the opinion of an expert, stakeholder or policy maker more than another’s we take the average of the resulting evaluation scores, as we do when computing the average cross efficiency in peer-appraisal. It also turns out that the proposed overall or synthetic measure can be computed by using the average (across ordering schemes) of the estimated multipliers for each considered performance indicator. The merits of the proposed measure are that it is easy to use and implement (even easier compared to those based on information theory) and intuitively appealing as it is based on weights that reflect the average relative importance of each considered performance indicator across all possible ordering schemes. These make the proposed measure very practical and attractive.

We also provide an empirical application of the propose measure along with comparative results from three information theory based methods (i.e., the Shannon entropy, the distance-based, the weighted least-square dissimilarity and the maximizing deviations) when applied to ABC inventory classification problem. The ABC analysis aims in classifying a DMU’s inventory items into three categories: A-class (very important), B-class (moderate important) and C-class (relatively unimportant). Our objective in this exercise is to examine the sensitivity of inventory items’ classification status across the resulting overall measures, paying particular attention on how the list of class-A items changes across different overall measures. This way we can examine robustness in the classification of inventory items across both different descending ordering schemes and aggregation methods.

The rest of this paper proceeds as follows: in the next section we provide a brief review of the ABC inventory classification literature related to construction of composite indicators and DEA. The proposed method is presented in the third section and in the fourth section we discuss the empirical results. In the fifth section, we provide a comparison of the proposed method with other three information theory based methods. Concluding remarks follow in the last section.

## 2. Literature Review

ABC is a well-known and practical inventory control method aiming to classify the stock of an organization's inventory items into three groups (i.e., A-class, B-class and C-class). The traditional ABC analysis is based on a single criterion such as annual dollar usage and Class A contains few items but constitutes the largest amount of annual dollar usage while class C contains a large amount of items with however a small amount of annual usage value. It has long been recognized however that using a single criterion may not necessarily provide a satisfactory classification of inventory items. For this reason, Flores, Olson and Dorai (1992) attempted to extent the ABC analysis into a multi-criteria decision-making tool where other criteria, such as inventory cost, lead time, etc., are also included in the analysis. For this purpose, Flores, Olson and Dorai (1992) and Partovi and Burtn (1993) applied an analytic hierarchy process-based approach that combines several criteria into a priority score for each inventory item. For more recent attempts in this literature stream see Lolli, Ishizaka and Gamberini (2014).

On the other hand, Ramanathan (2006) and Ng (2007) were the first to relate the multiple criteria ABC inventory classification problem to efficiency analysis by treating classification criteria as outputs or performance indicators. This turns the multiple criteria ABC inventory classification problem into a problem of constructing composite indicators. For this purpose, Ramanathan (2006) used what is now known as the Benefit-of-the-Doubt (BoD) model (Cherchye *et al.*, 2007)), namely an input-oriented DEA model with a single unitary input, while Ng (2007) relied essentially on Kao and Hung's (2003) (K&H hereafter) model. Even though both models assume that a helmsman attempts to aggregate a number of indicators to its best interest by assigning its own weights to each one of them, they have two main differences: *first*, in the K&H model the values of the performance indicators are normalized at the outset in the range of [0,1] while in the BoD model this is not necessary, and *second*, in the K&H model the resulting weights sum up to one while in the BoD model this is not necessarily true.[3] Nevertheless, the two models are related to each other (see Kao *et al.*, 2008) with the BoD model implying the K&H model but not the other way around, in the sense that once we solve the BoD model we can immediately obtain the solution of the K&H model while the opposite it is not always possible (Karagiannis and Paschalidou, 2017). However, the K&H model is computationally less demanding that the BoD model as it contains less constraints. In fact, the number of constraints in the K&H model is one plus the number of the considered performance indicators while in the BoD model is equal to the number of DMUs plus the number of the considered performance indicators.

Ng (2007) provided a refinement of the K&H model in the case of a particular form of weight restrictions, namely when the relative importance of the considered performance indicators can be ranked in a descending order. In such a case, where experts, stakeholders or policy makers can come *ex ante* with a descending ordering of the importance of the considered performance indicators, the resulting K&H model becomes even less computationally demanding. Ng (2007, 2008) showed that in this case there is no need to solve the model with an optimizer because the value of the composite performance indicator can be computed by simply using the maximum of indicators' partial averages. More importantly, the descending ordering of indicator's importance, as any other weight restriction form, tends to limit the problem that the conventional BoD and the K&H models are not compensatory in the sense that bad scores of the considered performance indicators may be completely ignored (Lolli, Ishizaka and Gamberini, 2014). For alternative attempts to deal with this problem in the case of the ABC inventory classification problem see Hadi-Vencheh (2010), who modified the normalization constraint of the K&H model, and Zhou and Fan (2007) and Chen (2011), who combined scores of the BoD and the anti-BoD model (namely, a BoD model searching for the least favorable weights of the considered performance indicators—see e.g. Zhou and Fan (2007)).

However, there may be a problem with Ng (2007) refinement of the K&H model when experts, stakeholders, or policy makers cannot reach a consensus about the relative importance of the considered performance indicators because the derived optimal scores and rankings from the different descending ordering schemes are not the same. But as Fu *et al.* (2016, p. 970) put it "each of the … rankings and viewpoints has some valuable advantages that we could not ignore. While it is impossible to ignore any ranking completely, the best way to make decision is to accept all possible rankings first, and then aggregate the results of the different rankings and viewpoints". In most of the cases, combining the scores of different descending ordering schemes results in a more realistic classification of the inventory items than relying solely on the scores of any particular descending ordering scheme.[4] To provide an overall or synthetic score for every inventory item, Fu *et al.* (2016)

relied on a distance-based method and Wu *et al.* (2018) on the weighted least-square dissimilarity approach to obtain a set of common (across items) weights for aggregating the evaluation scores under the different descending ordering schemes while Zheng *et al.* (2017) used Shannon entropy for the same purpose.[5] However, the aggregation methods may not necessarily base on information theory and thus, be purely data-driven. As a matter of fact, in the next section, we provide an alternative overall or synthetic score rooted to performance evaluation literature.

## 3. Proposed Method

Consider the following model for constructing composite performance indicators used by Ng (2007, 2008):

$$
\begin{aligned}
S^k = \max_{w_j^k} & \sum_{j=1}^{J} w_j^k y_j^k \\
st & \sum_{j=1}^{J} w_j^k = 1 \\
& w_j^k \geq 0 \qquad\qquad j = 1, \dots, J \\
& w_j^k - w_{j+1}^k \geq 0 \qquad j = 1, \dots, J-1
\end{aligned}
\tag{1}
$$

where *S* refers to the estimated value of the composite performance indicator, *y* to the normalized value of performance indicators, *w* to their (DMU-specific) weights to be estimated, *k=1,…,K* is used to index DMUs and *j=1,…,J* to index indicators, and $y_j^k = \left(I_j^k - \min_k I_j^k\right)/(\max_k I_j^k - \min_k I_j^k)$.[6] This is essentially the model used by Kao and Hung (2003) augmented with the last inequality constraint that ranks the relative importance of the considered indicators in a descending order. This weak partial ordering of weights has the advantage of being simple and intuitive as it does not require information on how much more important one indicators is over another one but one has only to specify their relative importance (Joro and Viitala, 2004). By solving (1) we get a set of DMU-specific weights that lie within the bounds imposed by the three constraints and maximize the value of its composite performance indicator.

Ng (2007) showed however that there is no need to solve (1) as the values of the composite performance indicator can be computed by using the maximum of partial averages, namely $S^k = \max_j \left\{ \left(\frac{1}{J}\right) \sum_{j=1}^{J} y_j^k \, , j - 1, . . , J \right\}$. To verify this notice first that (1) can be written equivalently as:

$$S^k = \max_{u_j^k} \sum_{j=1}^{J} u_j^k x_j^k$$

$$st \sum_{j=1}^{J} j u_j^k = 1 \qquad (2)$$

$$u_j^k \geq 0 \qquad j = 1, \ldots, J$$

when $u_j^k = w_j^k - w_{j+1}^k, u_J^k = w_J^k$ and the $x_j^k$'s are the partial sum of the $y_j^k$'s, namely $x_1^k = y_1^k, x_2^k = \sum_{j=1}^{2} y_j^k$, …, $x_J^k = \sum_{j=1}^{J} y_j^k$. From (2) one can derive both the value of the composite performance indicator as well as the weights assigned to each indicator as follows: if $S^k = \max_j \left\{ \left(\frac{1}{J}\right) \sum_{j=1}^{J} y_j^k \, , j - 1, . . , J \right\} = y_1^k$ then $u_1^k = 1$ and $u_{j>1}^k = 0$, which in turn implies that $w_1^k = 1$ and $w_{j>1}^k = 0$ and thus, the resulting weights profile is $\{1,0, \ldots ,0\}$. If, on the other hand, $S^k = \max_j \left\{ \left(\frac{1}{J}\right) \sum_{j=1}^{J} y_j^k \, , j - 1, . . , J \right\} = \frac{1}{2}(y_1^k + y_2^k)$ then $u_2^k = 1/2$ and $u_1^k = u_{j>2}^k = 0$, which in turn implies that $w_1^k = w_2^k = 1/2$ and $w_{j>2}^k = 0$ and thus, the resulting weights profile is $\left\{\frac{1}{2}, \frac{1}{2}, 0, \ldots . ,0\right\}$. If instead $S^k = \max_j \left\{ \left(\frac{1}{J}\right) \sum_{j=1}^{J} y_j^k \, , j - 1, . . , J \right\} = \frac{1}{3}(y_1^k + y_2^k + y_3^k)$ then $u_3^k = 1/3$ and $u_1^k = u_2^k = u_{j>3}^k = 0$, which in turn implies that $w_1^k = w_2^k = w_3^k = 1/3$ and $w_{j>3}^k = 0$ and thus, the resulting weights profile is $\left\{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, \ldots . ,0\right\}$, and so on up to the case that $S^k = \max_j \left\{ \left(\frac{1}{J}\right) \sum_{j=1}^{J} y_j^k \, , j - 1, . . , J \right\} = \left(\frac{1}{J}\right) \sum_{j=1}^{J} y_j^k$. Then, $u_J^k = 1/J$ and $u_{j<J}^k = 0$, which in turn implies that $w_1^k = \cdots = w_J^k = 1/J$ and thus, the resulting weight profile corresponds to equal weights, namely $\left\{\frac{1}{J}, \ldots, \frac{1}{J}\right\}$.

If there is no consensus about the relative importance of the considered indicators we can obtain alternative estimates of the composite performance indicator based on the different descending ordering schemes. The number of possible descending ordering schemes depends on the number of the considered performance indicators. One can verify that the number of possible descending ordering schemes is equal to *J!*. That is, in the case of two performance indicators we have two possible descending ordering schemes, i.e., either the weight of the first indicators is greater or equal to the second one or that of the second indicators is greater or equal to the first one. In the case of three performance indicators we have six possible descending ordering schemes, in the case of four we have twenty four, and so on. In each of these cases we can obtain the weights of models (1) and (2) as explained above. For the cases of two and three performance indicators, we give the relevant weights as well as the partial averages for computing the relevant composite performance indicators in Tables 1 and 2, respectively.

Let then $S_R^k$ be the score of the k<sup>th</sup> DMU's composite performance indicator under the R<sup>th</sup> descending ordering scheme *(R=1,…,J!)* that is equal to $S_R^k = \sum_{j=1}^{J} w_{jR}^k y_j^k$ , where $w_{jR}^k$ are computed as described above and vary across the alternative descending ordering schemes. The performance of each DMU may be assessed as many times as the number of possible ordering schemes in a fashion similar to peer-appraisal evaluation where each DMU is evaluated by means of cross efficiencies using the input and output weights of all other DMUs, which in this case provide the different "value systems". For our purposes however each possible descending ordering scheme (and not the constituent DMUs) is considered as a different "value systems" on which performance evaluation is based on. In most of the cases, DMUs tend to perform better under some (but not all) ordering schemes as they do better in terms of some performance indicators and thus, are able to achieve a higher score whenever the relative importance of these indicators is valued relatively more.[7] Then, using the relevant scores we can find under which ordering scheme each DMU achieves its best and worst performance or alternatively, which ordering scheme favours the performance of each DMU.

It may however be more appropriate or fair to consider an overall measure for each DMU summarizing its achievements across all possible ordering schemes. The measure proposed here is inspired by the notion of average cross efficiency that provides a synthetic view for peer-appraisal evaluation by taking the arithmetic average of each DMU's cross efficiencies. In an analogous manner and as long as there is no *a priori* reason to weigh a particular descending ordering scheme more than another, we take the average of the resulting $S_R^k$ scores to obtain a simple metric providing a synthetic view of preference-appraisal evaluation. In the resulting metric, which is referred to as the *partial average cross-weight* score, each ordering scheme accounts equally for the overall performance of a DMU. One can verify that, for each DMU, the average (across ordering schemes) of the $S_R^k$ scores may also be obtained by using the average (across ordering schemes) of the estimated weights in the computation of the partial average cross-weight score $S_O^k$. That is,

$$S_O^k = \frac{1}{J!}\sum_{R=1}^{J!} S_R^k = \frac{1}{J!}\sum_{R=1}^{J!}\sum_{j=1}^{J} w_{jR}^k y_j^k = \sum_{j=1}^{J}\left(\frac{\sum_{R=1}^{J!} w_{jR}^k}{J!}\right) y_j^k = \sum_{j=1}^{J} \overline{w}_j^k y_j^k \qquad (3)$$

These weights reflect the average perspective of DMUs about the relative importance of each performance indicator across all possible descending ordering schemes and thus, the proposed preference-appraisal scheme may be seen as a cross-weight evaluation process in the sense used by Wang and Chin (2010), where the *J!* set of weights are not directly used for computing $S_R^k$'s but are used to generate an average set of weights for each DMU.

## 4. Empirical results for the ABC inventory classification problem

In this section we apply the proposed method to the ABC inventory classification problem in order to obtain an overall classification of items by combining their scores from all possible descending ordering schemes. For these purposes we use Flores, Olson and Doral (1992) data set for 47 items and following Ng (2007), we base our analysis into three classification criteria, namely annual dollar usage, average unit cost, and lead time (in weeks), with the relevant data given in Table 3.[8] All criteria are positively related to the priority score of each item and its classification status. That is, the higher the demand (as measured by its annual dollar use), the value (as measured by its average unit cost) or the lead time for an item is, the higher the required service level will be and thus, it should be classified into a higher class. For the latter criterion, notice that a stock-out of an item with high lead time will be refilled after a longer time period than an item having a lower lead time and thus its importance as a stock item is greater (Lolli, Ishizaka and Gamberini, 2014). In the case of three classification criteria considered here, there are six possible descending ordering schemes under which to evaluate the 47 items. Lastly, following previous literature, e.g., Ramanathan (2006), Ng (2007), Zhou and Fan (2007), Hadi-Vencheh

(2010), and Chen (2011), we classify the first 10 ranked items into class A, the next 14 items into class B and the last 23 items into class C.

The scores of items and their classification under the six descending ordering schemes are given in Table 4. Several interesting results emerge from these figures: *first,* when the third criterion (i.e., lead time) is considered as the most valuable (see the last two columns in Table 4) the estimated scores have a higher mean but also a larger dispersion, as measured by standard deviation. On the other hand, the lowest standard deviation is obtained when the second criterion (i.e., average unit cost) is considered as the most valuable (see the two middle columns in Table 4). *Second*, the classification status of 21 items (44.7%) remains unchanged across all descending ordering schemes. These include four items (#2, #9, #13 and #29) in class A, three items (#19, #31 and #39) in class B and fourteen items (#11, #16, #24, #25, #26, #30, #32, #35, #36, #38, #41, #42, #44 and #46) in class C. That is, 40% of A-class items, 29% of B-class items and 61% of C-class items do not change classification status under the six ordering schemes. On the other hand, the classification of only three items (#1, #5 and #6) alternate across all classes across the six descending ordering schemes while that of the remaining items between two classes. *Third*, only one item came up with the maximum score of one in all ordering schemes except the two considering the third criterion as the most valuable. This is item #1 when the first criterion (i.e., annual dollar usage) is considered as the most valuable (see the first two columns in Table 4) and item #2 when the second criterion (i.e., average unit cost) is considered as the most valuable (see the two middle columns in Table 4). In contrast, when the third criterion (i.e., lead time) is considered as the most valuable we found four items with the maximum score of one, namely items #13, #29, #34 and #45.[9] On the other hand, at the bottom of items' ranking, we found items #25 and #11 having the lowest score under respectively three and two ordering schemes.

The resulting weight profiles for the six descending ordering schemes are summarized in Table 5. According to our results, when either the first (i.e., annual dollar usage) or the second (i.e., average unit cost) criterion is considered as the most valuable the resulting weight profiles contain two or three of the considered criteria while when the third criterion (i.e., lead time) is considered as the most valuable the resulting weight profiles contain only one criterion. In particular, 72% (34 out of 47) of the items have a weight profile involving all criteria, i.e., (1/3, 1/3, 1/3), when $w_1 \geq w_2 \geq w_3$ while 51% (24 out of 47) of items have a weight profile involving two criteria, i.e., (1/2, 0, 1/2) and 32% (15 out of 47) of items have a weight profile involving all criteria considered, i.e., (1/3, 1/3, 1/3), when $w_1 \geq w_3 \geq w_2$. On the other hand, almost 62% (29 out of 47) of the items have a weight profile involving all criteria considered, i.e., (1/3, 1/3, 1/3), when $w_2 \geq w_1 \geq w_3$ while 70% (33 out of 47) of items have a weight profile involving two criteria, i.e., (0, 1/2, 1/2), when $w_2 \geq w_3 \geq w_1$. In contrast, 74% (35 out of 47) and 81% (38 out of 47) of items have a weight profile involving only one criterion, i.e., (0, 0, 1), when respectively $w_3 \geq w_1 \geq w_2$ and $w_3 \geq w_2 \geq w_1$.

The results in Table 5 also show that a single weights profile, common to the vast majority of items, came up in four descending ordering schemes. In particular, for the $w_1 \geq w_2 \geq w_3$ scheme this profile regarding the 72% of items is (1/3, 1/3, 1/3), for the $w_2 \geq w_3 \geq w1_1$ scheme the profile regarding the 70% of items is (0, 1/2, 1/2), and for the $w_3 \geq w_1 \geq w_2$ and the $w_3 \geq w_2 \geq w_1$ schemes the profile regarding respectively the 74% and the 81% of the items is (0, 0, 1). On the other hand, in the other two descending ordering schemes, two weights profiles are shared by the vast majority of items. For the $w_1 \geq w_3 \geq w_2$ scheme, these profiles are (1/2, 0, 1/2) and (1/3, 1/3, 1/3) regarding respectively the 51% and the 32% of items while for the $w_2 \geq w_1 \geq w_3$ scheme the corresponding profiles are (0, 1, 0) and (1/3, 1/3, 1/3) regarding respectively the 32% and the 62% of the items.

In Figure 6 we report the A-class items according to the six ordering schemes. We see that there is consensus only on four items, namely #2, #9, #13 and #29. This accounts for 40% of the A-class items. Apart from this, it seems that the first three descending ordering schemes, namely $w_1 \geq w_2 \geq w_3$ , $w_1 \geq w_3 \geq w_2$ and $w_2 \geq w_1 \geq w_3$, tend to favor items #1, #2, #3, #4, and #5 to be classified as A-class items while the last three ordering schemes, namely $w_2 \geq w_3 \geq w_1$, $w_3 \geq w_1 \geq w_2$ and $w_3 \geq w_2 \geq w_1$, tend to favor items #18, #28, #34, #40 and #45. Thus, it turns to a difficult task to obtain the list of A-class items by simply looking separately at the results from the different descending ordering schemes. One can confirm that this is also the case for the other two classes. Instead, the use of an overall measure that takes into account the results of all six descending ordering schemes may resolve this issue when there is no consensus on about the relative importance of the considered classification criteria.

The results of such an overall measure based on partial average cross-weight scores are given in Table 7. In this Table we report for each item the average criteria weights and the resulting score and classification status. From there we can see that for almost half of the items (23 out of 47) the average criteria weights are (0.193, 0.193, 0.610). These refer to five items in class A, nine items in class B and nine items in class C. To these, we have to add nine other items with average criteria weights of (0.110, 0.310, 0.527), from which one item is in class A, two are in class B, and six are in class C. Taking together, for the 68% of items, the third criterion (i.e., lead time) is considered as the most valuable and the second (i.e. average unit cost) as being either equally or more important than the first one (i.e., annual dollar usage). In other words, for the vast majority of items, the $w_3 \geq w_2 \geq w_1$ scheme is applied which is completely different than the ordering scheme used by Ng (2007), namely $w_1 \geq w_2 \geq w_3$. Nevertheless, the classification of items reported in the last column of Table 7 is only by 19% (9 out of 47) different than that in the third column of Table 4, which correspond to Ng (2007) results, and their differences in terms of A-class items is just on three items (#4, #5 and #6 versus #28, #34 and #45) as can be seen from Table 6.

## 5. Comparison with other methods

In this section we compare the results of the partial average cross-weight method presented in the previous section with those of three information theory based methods used previously for the same purpose, i.e., Shannon entropy (Zheng *et al.*, 2017), the distance-based method (Fu *et al.*, 2016), and the weighted least-square dissimilarity approach (Wu *et al.*, 2018). In addition, we include in the comparison the results of another information theory based method, namely the maximizing deviation method, used previously by Chen (2011) for aggregating the most and the least favorable items' scores but which can as well be employed for our purpose. Before that however we show how overall scores obtained from these information theory based methods may be related to a kind of average weights as in the case of the partial average cross-weight method.

Notice first that all the aforementioned information theory based methods result in a set of common (across items) but unequal (across ordering schemes) aggregation weights. Let these aggregation weights be $\alpha_R^h$ with *R=1,..,J!*, where $h$ is used to index aggregation weights' computation methods, i.e., h={Shannon entropy, distance-based, maximizing deviations}. Then one can verify that

$$\tilde{S}_h^k = \sum_{R=1}^{J!} \alpha_R^h S_R^k = \sum_{R=1}^{J!} \alpha_R^h \sum_{j=1}^{J} w_{jR}^k y_j^k = \sum_{j=1}^{J} \left( \sum_{R=1}^{J!} \alpha_R^h w_{jR}^k \right) y_j^k = \sum_{j=1}^{J} \widehat{w}_{jh}^k y_j^k \qquad (4)$$

where $\widehat{w}_{jh}^k$ are the sum of the products of the aggregation and the criteria weights for each ordering scheme. They differ across the computation methods as each one results in different aggregation weights that are endogenously determined by a different procedure.

In the Shannon entropy method, this procedure involves three steps: *first*, set $\bar{S}_R^K = S_R^k / \sum_{k=1}^K S_R^k$ for *k=1,…,K* and compute the value of entropy as $e_R = -e_0 \sum_{k=1}^K \bar{S}_R^K \, ln\bar{S}_R^K$ for *R=1,…,J!* where the entropy constant is $e_0 = 1/lnK$. *Second*, set $b_R = 1 - e_R$ for *R=1,…,J!* and *third*, compute the degree of importance of each descending ordering scheme as $\alpha_R = b_R / \sum_{R=1}^{J!} b_R$ for *R=1,..J!* which implies that the smaller is the value of the entropy then the larger is the degree of importance and, *vice versa*. As the value of the entropy is inversely related to the variation of items' scores, a descending ordering scheme with a larger variation in items' scores receives a larger aggregation weight, and *vice versa*. According to our empirical results reported in Table 8, the largest aggregation weight is assigned to the $w_1 \geq w_2 \geq w_3$ ordering scheme and the smallest to the $w_2 \geq w_3 \geq w_1$. Notice also that the aggregation weights related to the two ordering schemes in which the third criterion (i.e., lead time) is considered to be the most valuable are the same.

The results for the overall measure obtained by the Shannon entropy method are given in Table 9 where we also report the average criteria weights and the resulting classification status of items.[10] From there we can see that for almost half of the items (23 out of 47) the average criteria weights are (0.218, 0.177, 0.601).[11] These refer to five items in class A, nine items in class B and nine items in class C. Thus, for almost half of the items, the third criterion (i.e., lead time) is considered as the most valuable and the second (i.e., average unit cost) as being less important than first one (i.e., annual dollar usage). These results differ with respect to the second and the third most valuable criterion compared to those obtained by the partial average cross-weight method, where for the great majority of items the second criterion (i.e., average unit cost) is either equally or more important than first one (i.e., annual dollar usage). On the other hand, the mean and the standard deviation of items' overall scores obtained by the two methods are very similar. The same is essential true for the resulting classification: the only difference between the partial average cross-weigh and the Shannon entropy aggregation methods is on whether item #6 or #23 is a B- or a C-class item. Otherwise, they provide exactly the same classification despite their differences in the average weights profiles.

In the distance-based method, the objective is to find a set of common (across items) aggregation weights $\alpha_R$ that minimizes the weighted square of the difference $d_R^k = S_R^k - S^k$ over all descending ordering schemes (Fu *et al.*, 2016; Cao *et al.*, 2016). The solution of this multi-objective programming problem results in $\alpha_R = 1/\left[\left(\sum_{k=1}^{K}(d_R^k)^2\right)\left(\sum_{R=1}^{J!}\left(\sum_{k=1}^{K}(d_R^k)^2\right)^{-1}\right)\right]$ for *R=1,…,J!*. In the distance-based method, as opposed to the Shannon entropy method, a descending ordering scheme with larger deviations from the mean receives a smaller aggregation weight, and *vice versa*. According to Fu *et al.* (2016) empirical results, reproduced in Table 8, the largest aggregation weight is assigned to the $w_2 \geq w_3 \geq w_1$ ordering scheme and the smallest to the $w_3 \geq w_2 \geq w_1$. Notice that the distance-based method assigns the largest aggregation weight to the descending ordering scheme that the Shannon entropy method assigns the smallest aggregation weight.

The results for the overall measure obtained by the distance-based method are given in Table 10 where we also report the average criteria weights and the resulting classification status of items.[12] From there we can see that for almost half of the items (23 out of 47) the average criteria weights are (0.173, 0.288, 0.534). Thus, for almost half of the items, the third criterion (i.e., lead time) is considered as the most valuable and the second (i.e., average unit cost) as being more important than the first one (i.e., annual dollar usage). These results differ with respect to the second and the third most valuable criterion compared to those obtained by the Shannon entropy method but they are similar to those obtained by the partial average cross-weight method. On the other hand, the mean and the standard deviation of items' overall scores are only slightly different from those obtained by both the Shannon entropy and the distance-based methods. Regarding item's classification status, the only difference between the distance-based and the partial average cross-weigh aggregation methods is on whether item #14 or #34 is an A- or B-class item. Otherwise, they provide exactly the same classification despite their differences in the average weights profiles. There are however differences between the Shannon entropy and the distance-based methods as to whether items #6 and #23 belongs to class B or C and items #14 and #34 belongs to class A or B.

In the weighted least-square dissimilarity method, the objective is to find a set of common (across items) aggregation weights $\alpha_R$ that minimizes the weighted least-square of the total dissimilarity measure $g_R = K - \sum_{k=1}^{K} W_{RR'}$ over all descending ordering schemes, where $W_{RR'} = \sum_{k=1}^{K} S_R^k S_{R'}^k \,/\, \left(\sum_{k=1}^{K}\left(S_R^k\right)^2 \sum_{k=1}^{K}\left(S_{R'}^k\right)^2\right)^{1/2}$ and *R* and *R'* are two different descending ordering schemes (Wang and Wang, 2013; Wu *et al.*, 2018). The solution of this quadratic programming problem results in $\alpha_R = (1/g_R^2)/\sum_{k=1}^{K}(1/g_R^2)$ for *R=1,…,J!*. A descending ordering scheme with large total dissimilarity will be given a small aggregation weight, and *vice versa*. According to Wu *et al.* (2018) empirical results, reproduced in Table 8, the largest aggregation weight is assigned to the $w_2 \geq w_3 \geq w_1$ ordering scheme and the smallest to the $w_1 \geq w_2 \geq w_3$. Notice that the weighted least-square dissimilarity and the distance-based method assign the largest aggregation weight to the same descending ordering scheme, namely, $w_2 \geq w_3 \geq w_1$.

The results for the overall measure obtained by the weighted least-square dissimilarity method are given in Table 11 where we also report the average criteria weights and the resulting classification status of items.[13] From there we can see that for almost half of the items (23 out of 47) the average criteria weights are (0.182, 0.238, 0.577). Thus, for almost half of the items, the third criterion (i.e., lead time) is considered as the most valuable and the second (i.e., average unit cost) as being more important than the first one (i.e., annual dollar usage). These results are similar to those obtained by the partial average cross-weight and the distance-based methods but differ from those obtained by the Shannon entropy method. The weighted least-square dissimilarity method results in the third largest mean overall score after the partial average cross-weight and the Shannon entropy methods. On the other hand, the weighted least-square dissimilarity and the partial average cross-weight methods result in the same classification status while the only difference between the weighted least-square dissimilarity and the Shannon entropy aggregation methods is on whether item #6 or #23 is a B- or a C-class item and between the weighted least-square dissimilarity and distance-based methods on whether item #14 or #34 is an A- or B-class item.

In the maximizing deviations method, the objective is to find a set of common (across items) aggregation weights $\alpha_R$ that maximize $h_R = \sum_{k=1}^{K} \sum_{i=1}^{K} \left| S_R^k - S_R^i \right|$ over all descending ordering schemes (Chen, 2011). The solution of this problem results in $\alpha_R = h_R / \sum_{R=1}^{J!} h_R$. In the maximizing deviations method, a descending ordering scheme in which the items have smaller deviations in performance receives a smaller aggregation weight, and *vice versa*. According to our empirical results reported in Table 8, the largest aggregation weight is assigned to the $w_3 \geq w_1 \geq w_2$ ordering scheme and the smallest to the $w_3 \geq w_2 \geq w_1$. Notice that both the distance-based and the maximizing deviations methods assign the smallest aggregation weight to the same descending ordering scheme.

The results for the overall measure obtained by the maximizing deviations method are given in Table 12 where we also report the average criteria weights and the resulting classification status of items. From there we can see that for almost half of the items (23 out of 47) the average criteria weights are (0.195, 0.178, 0.623). Thus, for almost half of the items, the third criterion (i.e., lead time) is considered as the most valuable and the first (i.e., annual dollar usage) as being more important than the second one (i.e., average unit cost). These results differ with respect to the second and the third most valuable criteria compared to those obtained by the partial average cross-weight, the distance-based and the weighted least-square dissimilarity methods but they are similar to those obtained by Shannon entropy. On the other hand, the maximizing deviations method results in the lowest mean overall score compared to the other three methods. Regarding the classification status of items, the differences between the maximizing deviations and the partial average cross-weigh and the weighted least-square dissimilarity aggregation methods is on whether items #6 and #7 belong to class C or B and items #23 and #43 to class. B or C. There are however differences between the maximizing deviations and the distance-based methods as to whether items #6 and #7 belong to class B or C, items #23 and #43 to class C or B, and items #14 and #34 to class A or B. In contrast, the only difference between the maximizing deviations and the Shannon entropy methods is on whether item #7 or #43 is an A- or B-class item.

We may now summarize the differences and the similarities among the four alternative methods: *first*, aggregation weights differ across methods. The Shannon entropy method puts more weight on the scores from a descending ordering scheme placing more importance on the first criterion (i.e., annual dollar usage), the distance–based and the weighted least-square dissimilarity method on those from a descending ordering scheme placing more importance on the second criterion (i.e., average unit cost), the maximizing deviations method on those from an descending ordering scheme that place more importance on the third criterion (i.e., lead time) while the partial average cross-weight method weights them equally. *Second*, in all methods, more importance is on average placed on the third criterion (i.e., lead time) while the second (i.e., average unit cost) criterion is on average valued more than the first (i.e., annual dollar usage) in the partial average cross-weight, the distance-based and the weighted least-square dissimilarity methods while the opposite is true in the Shannon entropy and the maximizing deviations methods. *Third*, despite these differences, the mean and the standard deviation of the overall scores are similar across methods. *Fourth*, all methods result in similar classifications with only minor differences: in particular, there are differences in the classification of only two items between the partial average cross-weight and the Shannon entropy methods as well as between the partial average cross-weight and the distance-based methods and between the Shannon entropy and the maximizing deviations methods while there are no differences between the partial average cross-weight and the weighted least-square dissimilarity methods. *Fifth*, there is consensus on the A-class items in all but the distance-based method, which classifies item #14 instead of item #34 in class A.

## 6. Concluding Remarks

In this paper we propose an alternative overall measure, inspired by the notion of average cross efficiency, that summarizes achievements across different descending ordering schemes regarding the relative importance of the considered indicators in the Ng (2007) model. The proposed measure is equal to the arithmetic average of the maximum partial averages across all possible descending ordering schemes. One can verify that it may also be obtained by using the average (across descending ordering schemes) of the estimated multipliers and for this reason it is referred to as partial average cross-weight measure. Compared to other information theory based measures used previously in the literature for the same purpose, it is computationally less demanding and has a more intuitively appealing interpretation.

We apply the proposed measure to the ABC inventory classification problem and compare our results with those obtained by four information theory based methods, namely, the Shannon entropy, the distance-based, weighted least-square dissimilarity and the maximizing deviations methods. The empirical results indicate that besides differences in aggregation and criteria weights among the alternative methods, there is a general consensus in the final classification of items as well as on the items that belong to class A, with only minor differences that are relatively more pronounced among the information theory based methods. On these grounds, it seems that in the particular study case the partial average cross-weight method has an advantage over the aforementioned information theory based methods due to its computation simplicity but it remains to be proved whether the same would be true in other applications of the Ng (2007) model, such as supplier selection, evaluation of faculty members' publication record, performance in Olympic games, and other cases where there is no consensus about the descending ordering scheme regarding the importance of the considered performance indicators.

## References

Angulo-Meza, L. and M.P.E. Lins. Review of methods for increasing discrimination in Data Envelopment Analysis, *Journal of Productivity Analysis*, 2002, 116, 225-42.

Cao, X., Fu, Y., Du, J., Sun, J. and M. Wang. Measuring Olympics performance based on a distance-based approach, *International Transactions in Operational Research*, 2016, 23, 979-90.

Charnes, A., Cooper, W.W. and E. Rhodes. Measuring the efficiency of decision making units, *European Journal of Operational Research*, 1978, 2, 429-444.

Chen, J.X. Peer-estimation for multiple criteria ABC inventory classification, *Computers and Operations Research*, 2011, 38, 1784-91.

Cherchye, L., Moesen, W., Rogge, N. and T. van Puyenbroeck. An introduction to "Benefit of the Doubt" composite indicators, *Social Indicators Research*, 2007, 82, 111-45.

Doyle, J.R. and R. Green. Efficiency and cross efficiency in DEA: Derivation, meanings and uses, *Journal of Operational Research Society*, 1994, 45, 567-78.

Dyson, R.G., Allen, R., Camanho. A.S., Podinovski, V.V., Sarrico, C.S. and E.A. Shale. Pitfalls and protocols in DEA, *European Journal of Operational Research*, 2001, 132, 245-59.

Flores, B.E., Olson, D.I. and V.K. Doral. Management of multicriteria inventory classification, *Mathematical and Computer Modeling*, 1992, 16, 71-82.

Fu, Y., Wang, M. and K.K. Lai. A modified nature publishing index via Shannon entropy, *Discrete Dynamics in Nature and Society*, 2015.

Fu, Y., Lai, K.K., Miao, Y. and J.W.K. Leung. A distance-based decision-making method to improve multiple criteria ABC inventory classification, *International Transactions in Operational Research*, 2016, 23, 969-78.

Hadi-Vencheh, A. An improvement to the multiple criteria ABC inventory classification, *European Journal of Operational Research*, 2010, 201, 962-65.

Joro, T. and E.J. Viitala. Weight-restricted DEA in action: From expert opinions to mathematical models, *Journal of Operational Research Society*, 2004, 55, 814-21.

Kao, C. and H.T. Hung. Ranking university libraries with a posteriori weights, *Libri*, 2003, 53, 282-89.

Kao, C., Wu, W.Y., Hsieh, W.J., Wang, T.Y., Lin, C. and L.H. Chen. Measuring the national competitiveness of Southeast Asian countries, *European Journal of Operational Research*, 2008, 187, 613-28.

Karagiannis, G. and G. Paschalidou. Assessing research effectiveness: A comparison of alternative non-parametric models, *Journal of Operational Research Society*, 2017, 68, 456-68.

Ladhari, T., Babai, M.Z. and I. Lajili. Multi-criteria inventory classification: New consensual procedures, *IMA Journal of Management Mathematics*, 2016, 26, 335-51.

Li, Z., Wu, X., Liu, F., Fu, Y. and K. Chen. Multicriteria ABC inventory classification using acceptability analysis, *International Transactions in Operational Research*, 2018 (forthcoming).

Lolli, F., Ishizaka, A. and R. Gamberini. New AHP-based approaches for multi-criteria inventory classification, *International Journal of Production Economics*, 2014, 156, 62-74.

Ng, W.L. A simple classifier for multiple criteria ABC analysis, *European Journal of Operational Research*, 2007, 177, 344-53.

Ng, W.L. An efficient and simple model for multiple criteria supplier selection problem, *European Journal of Operational Research*, 2008, 186, 1059-67.

Partovi, F.Y. and J. Burton. Using the analytic hierarchy process for ABC analysis, *International Journal of Production and Operations Management*, 1993, 13, 29-44.

Ramanathan, R. ABC inventory classification with multiple-criteria using weighted linear optimization, *Computers and Operations Research*, 2006, 33, 695-700.

Sexton, T.R., Silkman, R.H. and A.J. Hogan. Data envelopment analysis: Critique and extensions, in Silkman, R.H. (ed), *Measuring Efficiency: An Assessment of Data Envelopment Analysis*, 1986, Jossey-Bass.

Shannon, C.E. A mathematical theory of communication, *Bell System Technical Journal*, 1948, 27, 379-423.

Wang, Y.M. and K.S. Chin. A neutral DEA model for cross-efficiency evaluation and its extension. *Expert Systems with Applications*, 2010, 37, 3666-75.

Wang, Y.M. and S. Wang. Approaches to determining the relative importance weights for cross-efficiency aggregation in data envelopment analysis, *Journal of Operational Research Society*, 2013, 64, 60-69.

Wu, J., Sun, J. and l. Liang. DEA cross-efficiency aggregation method based upon Shannon entropy, *International Journal of Production Research*, 2012, 50, 6726-36.

Wu, S., Fu, Y., Lai, K.K. and W.K.J. Leung. A weighted least-square dissimilarity approach for multiple criteria ABC inventory classification, *Asia-Pacific Journal of Operational Research*, 2018 (forthcoming).

Zheng, S., Fu, Y., Keung, L.K. and L. Liang. An improvement to multiple criteria ABC inventory classification using Shannon entropy, *Journal of Systems Science and Complexity*, 2017, 30, 857-65.

Zhou, P. and L. Fan. A note on multi-criteria ABC inventory classification using weighted linear optimization, *European Journal of Operational Research*, 2007, 182, 1488-91.

Table 1: Weights and composite indicator's value in the Ng model with two indicators

| | $u_1$ | $u_2$ | $w_1$ | $w_2$ | $S^k$ |
|---|---|---|---|---|---|
| $w_1 \geq w_2$ | 1 | 0 | 1 | 0 | $y_1$ |
| | 0 | ½ | 1/2 | ½ | $(y_1 + y_2)/2$ |
| $w_2 \geq w_1$ | 0 | 1 | 0 | 1 | $y_2$ |
| | ½ | 0 | 1/2 | ½ | $(y_1 + y_2)/2$ |

Table 2: Weights and composite indicator's value in the Ng model with three indicators

| | $u_1$ | $u_2$ | $u_3$ | $w_1$ | $w_2$ | $w_3$ | $S^k$ |
|---|---|---|---|---|---|---|---|
| $w_1 \geq w_2 \geq w_3$ | 1 | 0 | 0 | 1 | 0 | 0 | $y_1$ |
| | 0 | ½ | 0 | ½ | 1/2 | 0 | $(y_1 + y_2)/2$ |
| | 0 | 0 | 1/3 | 1/3 | 1/3 | 1/3 | $(y_1 + y_2 + y_3)/3$ |
| $w_1 \geq w_3 \geq w_2$ | 1 | 0 | 0 | 1 | 0 | 0 | $y_1$ |
| | 0 | 0 | ½ | ½ | 0 | ½ | $(y_1 + y_3)/2$ |
| | 0 | 1/3 | 0 | 1/3 | 1/3 | 1/3 | $(y_1 + y_2 + y_3)/3$ |
| $w_2 \geq w_1 \geq w_3$ | 0 | 1 | 0 | 0 | 1 | 0 | $y_2$ |
| | ½ | 0 | 0 | ½ | 1/2 | 0 | $(y_1 + y_2)/2$ |
| | 0 | 0 | 1/3 | 1/3 | 1/3 | 1/3 | $(y_1 + y_2 + y_3)/3$ |
| $w_2 \geq w_3 \geq w_1$ | 0 | 1 | 0 | 0 | 1 | 0 | $y_2$ |
| | 0 | 0 | ½ | 0 | 1/2 | ½ | $(y_2 + y_3)/2$ |
| | 1/3 | 0 | 0 | 1/3 | 1/3 | 1/3 | $(y_1 + y_2 + y_3)/3$ |
| $w_3 \geq w_1 \geq w_2$ | 0 | 0 | 1 | 0 | 0 | 1 | $y_3$ |
| | ½ | 0 | 0 | ½ | 0 | ½ | $(y_1 + y_3)/2$ |
| | 0 | 1/3 | 0 | 1/3 | 1/3 | 1/3 | $(y_1 + y_2 + y_3)/3$ |
| $w_3 \geq w_2 \geq w_1$ | 0 | 0 | 1 | 0 | 0 | 1 | $y_3$ |
| | 0 | ½ | 0 | 0 | 1/2 | ½ | $(y_2 + y_3)/3$ |
| | 1/3 | 0 | 0 | 1/3 | 1/3 | 1/3 | $(y_1 + y_2 + y_3)/3$ |

Table 3: Measures of inventory items

| Item | Annual dollar usage | Average unit cost | Lead time |
|---|---|---|---|
| 1 | 5840.64 | 49.92 | 2 |
| 2 | 5670 | 210 | 5 |
| 3 | 5037.12 | 23.76 | 4 |
| 4 | 4769.56 | 27.73 | 1 |
| 5 | 3478.8 | 57.98 | 3 |
| 6 | 2936.67 | 31.24 | 3 |
| 7 | 2820 | 28.2 | 3 |
| 8 | 2640 | 55 | 4 |
| 9 | 2423.52 | 73.44 | 6 |
| 10 | 2407.5 | 160.5 | 4 |
| 11 | 1075.2 | 5.12 | 2 |
| 12 | 1043.5 | 20.87 | 5 |
| 13 | 1038 | 86.5 | 7 |
| 14 | 883.2 | 110.4 | 5 |
| 15 | 854.4 | 71.2 | 3 |
| 16 | 810 | 45 | 3 |
| 17 | 703.68 | 14.66 | 4 |
| 18 | 594 | 49.5 | 6 |
| 19 | 570 | 47.5 | 5 |
| 20 | 467.6 | 58.45 | 4 |
| 21 | 463.6 | 24.4 | 4 |
| 22 | 455 | 65 | 4 |
| 23 | 432.5 | 86.5 | 4 |
| 24 | 398.4 | 33.2 | 3 |
| 25 | 370.5 | 37.05 | 1 |
| 26 | 338.4 | 33.84 | 3 |
| 27 | 336.12 | 84.03 | 1 |
| 28 | 313.6 | 78.4 | 6 |
| 29 | 268.68 | 134.34 | 7 |
| 30 | 224 | 56 | 1 |
| 31 | 216 | 72 | 5 |
| 32 | 212.08 | 53.02 | 2 |
| 33 | 197.92 | 49.48 | 5 |
| 34 | 190.89 | 7.07 | 7 |
| 35 | 181.8 | 60.6 | 3 |
| 36 | 163.28 | 40.82 | 3 |
| 37 | 150 | 30 | 5 |
| 38 | 134.8 | 67.4 | 3 |
| 39 | 119.2 | 59.6 | 5 |
| 40 | 103.36 | 51.68 | 6 |
| 41 | 79.2 | 19.8 | 2 |
| 42 | 75.4 | 37.7 | 2 |
| 43 | 59.78 | 29.89 | 5 |
| 44 | 48.3 | 48.3 | 3 |
| 45 | 34.4 | 34.4 | 7 |
| 46 | 28.8 | 28.8 | 3 |
| 47 | 25.38 | 8.46 | 5 |

Source: Flores, Olson and Doral (1992)

Table 4: Items' score and classification status under different descending ordering schemes

| | $w_1 \geq w_2 \geq w_3$ | | $w_1 \geq w_3 \geq w_2$ | | $w_2 \geq w_1 \geq w_3$ | | $w_2 \geq w_3 \geq w_1$ | | $w_3 \geq w_1 \geq w_2$ | | $w_3 \geq w_2 \geq w_1$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.000 | A | 1.000 | A | 0.609 | A | 0.462 | B | 0.583 | B | 0.462 | C |
| 2 | 0.985 | A | 0.971 | A | 1.000 | A | 1.000 | A | 0.879 | A | 0.879 | A |
| 3 | 0.862 | A | 0.862 | A | 0.484 | A | 0.484 | B | 0.681 | A | 0.500 | B |
| 4 | 0.816 | A | 0.816 | A | 0.463 | A | 0.309 | C | 0.408 | C | 0.309 | C |
| 5 | 0.594 | A | 0.594 | A | 0.426 | A | 0.395 | B | 0.464 | C | 0.395 | C |
| 6 | 0.501 | A | 0.501 | B | 0.320 | B | 0.320 | C | 0.417 | C | 0.333 | C |
| 7 | 0.481 | B | 0.481 | B | 0.309 | B | 0.309 | C | 0.407 | C | 0.333 | C |
| 8 | 0.450 | B | 0.475 | B | 0.398 | B | 0.398 | C | 0.500 | B | 0.500 | B |
| 9 | 0.526 | A | 0.623 | A | 0.526 | A | 0.583 | A | 0.833 | A | 0.833 | A |
| 10 | 0.584 | A | 0.556 | A | 0.758 | A | 0.758 | A | 0.556 | B | 0.629 | B |
| 11 | 0.181 | C | 0.181 | C | 0.116 | C | 0.116 | C | 0.174 | C | 0.167 | C |
| 12 | 0.306 | B | 0.421 | B | 0.306 | C | 0.372 | C | 0.667 | B | 0.667 | B |
| 13 | 0.524 | A | 0.587 | A | 0.524 | A | 0.699 | A | 1.000 | A | 1.000 | A |
| 14 | 0.443 | B | 0.443 | B | 0.514 | A | 0.590 | A | 0.667 | B | 0.667 | B |
| 15 | 0.266 | C | 0.266 | C | 0.323 | B | 0.328 | C | 0.333 | C | 0.333 | C |
| 16 | 0.221 | C | 0.234 | C | 0.221 | C | 0.264 | C | 0.333 | C | 0.333 | C |
| 17 | 0.221 | C | 0.308 | C | 0.221 | C | 0.273 | C | 0.500 | B | 0.500 | B |
| 18 | 0.383 | B | 0.466 | B | 0.383 | B | 0.525 | A | 0.833 | A | 0.833 | A |
| 19 | 0.322 | B | 0.380 | B | 0.322 | B | 0.437 | B | 0.667 | B | 0.667 | B |
| 20 | 0.279 | C | 0.288 | C | 0.279 | C | 0.380 | B | 0.500 | B | 0.500 | B |
| 21 | 0.223 | C | 0.288 | C | 0.223 | C | 0.297 | C | 0.500 | B | 0.500 | B |
| 22 | 0.289 | C | 0.289 | C | 0.292 | C | 0.396 | B | 0.500 | B | 0.500 | B |
| 23 | 0.322 | B | 0.322 | C | 0.397 | B | 0.449 | B | 0.500 | B | 0.500 | B |
| 24 | 0.178 | C | 0.199 | C | 0.178 | C | 0.235 | C | 0.333 | C | 0.333 | C |
| 25 | 0.108 | C | 0.072 | C | 0.156 | C | 0.156 | C | 0.072 | C | 0.078 | C |
| 26 | 0.176 | C | 0.194 | C | 0.176 | C | 0.237 | C | 0.333 | C | 0.333 | C |
| 27 | 0.219 | C | 0.146 | C | 0.385 | B | 0.385 | B | 0.146 | C | 0.193 | C |
| 28 | 0.414 | B | 0.441 | B | 0.414 | B | 0.596 | A | 0.833 | A | 0.833 | A |
| 29 | 0.558 | A | 0.558 | A | 0.631 | A | 0.815 | A | 1.000 | A | 1.000 | A |
| 30 | 0.141 | C | 0.094 | C | 0.248 | C | 0.248 | C | 0.094 | C | 0.124 | C |
| 31 | 0.342 | B | 0.350 | B | 0.342 | B | 0.497 | B | 0.667 | B | 0.667 | B |
| 32 | 0.144 | C | 0.144 | C | 0.234 | C | 0.234 | C | 0.167 | C | 0.200 | C |
| 33 | 0.304 | B | 0.348 | B | 0.304 | C | 0.442 | B | 0.667 | B | 0.667 | B |
| 34 | 0.346 | B | 0.514 | A | 0.346 | B | 0.505 | B | 1.000 | A | 1.000 | A |
| 35 | 0.210 | C | 0.210 | C | 0.271 | C | 0.302 | C | 0.333 | C | 0.333 | C |
| 36 | 0.177 | C | 0.179 | C | 0.177 | C | 0.254 | C | 0.333 | C | 0.333 | C |
| 37 | 0.270 | C | 0.344 | B | 0.270 | C | 0.394 | B | 0.667 | B | 0.667 | B |
| 38 | 0.219 | C | 0.219 | C | 0.304 | C | 0.319 | C | 0.333 | C | 0.333 | C |
| 39 | 0.316 | B | 0.341 | B | 0.316 | B | 0.466 | B | 0.667 | B | 0.667 | B |
| 40 | 0.358 | B | 0.423 | B | 0.358 | B | 0.530 | A | 0.833 | A | 0.833 | A |
| 41 | 0.083 | C | 0.088 | C | 0.083 | C | 0.119 | C | 0.167 | C | 0.167 | C |
| 42 | 0.111 | C | 0.111 | C | 0.159 | C | 0.163 | C | 0.167 | C | 0.167 | C |
| 43 | 0.264 | C | 0.336 | C | 0.264 | C | 0.394 | B | 0.667 | B | 0.667 | B |
| 44 | 0.183 | C | 0.183 | C | 0.211 | C | 0.272 | C | 0.333 | C | 0.333 | C |
| 45 | 0.381 | B | 0.501 | B | 0.381 | B | 0.571 | A | 1.000 | A | 1.000 | A |
| 46 | 0.150 | C | 0.167 | C | 0.150 | C | 0.224 | C | 0.333 | C | 0.333 | C |
| 47 | 0.228 | C | 0.333 | C | 0.228 | C | 0.341 | C | 0.667 | B | 0.667 | B |
| avrg | 0.354 | | 0.380 | | 0.340 | | 0.401 | | 0.526 | | 0.516 | |
| stdev | 0.219 | | 0.222 | | 0.171 | | 0.181 | | 0.257 | | 0.257 | |

Table 5: Items' weight profiles under different descending ordering schemes

| | Weight profiles including | | |
|---|---|---|---|
| | one criterion | two criteria | three criteria |
| $w_1 \geq w_2 \geq w_3$ | 8 | 5 | 34 |
| $w_1 \geq w_3 \geq w_2$ | 8 | 24 | 15 |
| $w_2 \geq w_1 \geq w_3$ | 15 | 3 | 29 |
| $w_2 \geq w_3 \geq w_1$ | 6 | 33 | 8 |
| $w_3 \geq w_1 \geq w_2$ | 35 | 7 | 5 |
| $w_3 \geq w_2 \geq w_1$ | 38 | 5 | 4 |

Table 6: A-class items under different descending ordering schemes and overall measures

| Item | Descending ordering scheme | | | | | | Overall measure | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (A) | (B) | (C) | (D) | (E) |
| 1 | ▯ | ▯ | ▯ | | | | ▯ | ▯ | ▯ | ▯ | ▯ |
| 2 | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ |
| 3 | ▯ | ▯ | ▯ | | ▯ | | ▯ | ▯ | ▯ | ▯ | ▯ |
| 4 | ▯ | ▯ | ▯ | | | | | | | | |
| 5 | ▯ | ▯ | ▯ | | | | | | | | |
| 6 | ▯ | | | | | | | | | | |
| 9 | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ |
| 10 | ▯ | ▯ | ▯ | ▯ | | | ▯ | ▯ | ▯ | ▯ | ▯ |
| 13 | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ |
| 14 | | | ▯ | ▯ | | | | | ▯ | | |
| 18 | | | | ▯ | ▯ | ▯ | | | | | |
| 28 | | | | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ |
| 29 | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ |
| 34 | | ▯ | | | ▯ | ▯ | ▯ | ▯ | | ▯ | ▯ |
| 40 | | | | ▯ | ▯ | ▯ | | | | | |
| 45 | | | | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ | ▯ |

Note: (1) refers to $w_1 \geq w_2 \geq w_3$, (2) to $w_1 \geq w_3 \geq w_2$, (3) to $w_2 \geq w_1 \geq w_3$, (4) to $w_2 \geq w_3 \geq w_1$, (5) to $w_3 \geq w_1 \geq w_2$, (6) to $w_3 \geq w_2 \geq w_1$, (A) to partial average cross-weight, (B) to Shannon entropy, (C) to distance-based (D) to weighted least-square dissimilarity and (E) to maximizing deviations.

Table 7: Partial average cross-weight scores and classification status of items

| Item | Average criteria weights | | | Item Score | Classification Status |
|---|---|---|---|---|---|
| | $w_1$ | $w_2$ | $w_3$ | | |
| 1 | 0.610 | 0.193 | 0.193 | 0.686 | A |
| 2 | 0.360 | 0.527 | 0.110 | 0.952 | A |
| 3 | 0.527 | 0.110 | 0.360 | 0.646 | A |
| 4 | 0.610 | 0.193 | 0.193 | 0.520 | B |
| 5 | 0.610 | 0.193 | 0.193 | 0.478 | B |
| 6 | 0.527 | 0.110 | 0.360 | 0.399 | C |
| 7 | 0.527 | 0.110 | 0.360 | 0.387 | C |
| 8 | 0.360 | 0.110 | 0.527 | 0.453 | B |
| 9 | 0.193 | 0.193 | 0.610 | 0.654 | A |
| 10 | 0.193 | 0.610 | 0.193 | 0.640 | A |
| 11 | 0.527 | 0.110 | 0.360 | 0.155 | C |
| 12 | 0.193 | 0.193 | 0.610 | 0.456 | B |
| 13 | 0.193 | 0.193 | 0.610 | 0.722 | A |
| 14 | 0.110 | 0.360 | 0.527 | 0.554 | B |
| 15 | 0.110 | 0.360 | 0.527 | 0.308 | C |
| 16 | 0.193 | 0.193 | 0.610 | 0.268 | C |
| 17 | 0.193 | 0.193 | 0.610 | 0.337 | C |
| 18 | 0.193 | 0.193 | 0.610 | 0.570 | B |
| 19 | 0.193 | 0.193 | 0.610 | 0.466 | B |
| 20 | 0.193 | 0.193 | 0.610 | 0.371 | C |
| 21 | 0.193 | 0.193 | 0.610 | 0.339 | C |
| 22 | 0.110 | 0.360 | 0.527 | 0.378 | C |
| 23 | 0.110 | 0.360 | 0.527 | 0.415 | B |
| 24 | 0.193 | 0.193 | 0.610 | 0.243 | C |
| 25 | 0.193 | 0.610 | 0.193 | 0.107 | C |
| 26 | 0.193 | 0.193 | 0.610 | 0.241 | C |
| 27 | 0.193 | 0.610 | 0.193 | 0.246 | C |
| 28 | 0.193 | 0.193 | 0.610 | 0.588 | A |
| 29 | 0.110 | 0.360 | 0.527 | 0.760 | A |
| 30 | 0.193 | 0.610 | 0.193 | 0.158 | C |
| 31 | 0.193 | 0.193 | 0.610 | 0.477 | B |
| 32 | 0.110 | 0.527 | 0.360 | 0.187 | C |
| 33 | 0.193 | 0.193 | 0.610 | 0.455 | B |
| 34 | 0.193 | 0.193 | 0.610 | 0.618 | A |
| 35 | 0.110 | 0.360 | 0.527 | 0.277 | C |
| 36 | 0.193 | 0.193 | 0.610 | 0.242 | C |
| 37 | 0.193 | 0.193 | 0.610 | 0.435 | B |
| 38 | 0.110 | 0.360 | 0.527 | 0.288 | C |
| 39 | 0.193 | 0.193 | 0.610 | 0.462 | B |
| 40 | 0.193 | 0.193 | 0.610 | 0.556 | B |
| 41 | 0.165 | 0.248 | 0.582 | 0.118 | C |
| 42 | 0.110 | 0.360 | 0.527 | 0.146 | C |
| 43 | 0.193 | 0.193 | 0.610 | 0.432 | B |
| 44 | 0.110 | 0.360 | 0.527 | 0.252 | C |
| 45 | 0.193 | 0.193 | 0.610 | 0.639 | A |
| 46 | 0.193 | 0.193 | 0.610 | 0.226 | C |
| 47 | 0.193 | 0.193 | 0.610 | 0.411 | C |
| Average | | | | 0.420 | |
| Stdev | | | | 0.190 | |

Table 8: Aggregation weights

| | Partial-average cross-weight method | Shannon entropy | Distance-based method | Maximizing deviations method | Weighted least-square dissimilarity method |
|---|---|---|---|---|---|
| $w_1 \geq w_2 \geq w_3$ | 0.167 | 0.210 | 0.111 | 0.175 | 0.102 |
| $w_1 \geq w_3 \geq w_2$ | 0.167 | 0.203 | 0.176 | 0.185 | 0.161 |
| $w_2 \geq w_1 \geq w_3$ | 0.167 | 0.142 | 0.148 | 0.136 | 0.206 |
| $w_2 \geq w_3 \geq w_1$ | 0.167 | 0.121 | 0.406 | 0.151 | 0.273 |
| $w_3 \geq w_1 \geq w_2$ | 0.167 | 0.161 | 0.082 | 0.225 | 0.148 |
| $w_3 \geq w_2 \geq w_1$ | 0.167 | 0.162 | 0.076 | 0.127 | 0.110 |

Table 9: Shannon entropy scores and classification status of items

| Item | Average criteria weights | | | Item Score | Classification |
|---|---|---|---|---|---|
| | $w_1$ | $w_2$ | $w_3$ | | Status |
| 1 | 0.658 | 0.164 | 0.174 | 0.725 | A |
| 2 | 0.415 | 0.475 | 0.107 | 0.952 | A |
| 3 | 0.580 | 0.087 | 0.329 | 0.675 | A |
| 4 | 0.658 | 0.164 | 0.174 | 0.556 | B |
| 5 | 0.658 | 0.164 | 0.174 | 0.493 | B |
| 6 | 0.580 | 0.087 | 0.329 | 0.413 | B |
| 7 | 0.580 | 0.087 | 0.329 | 0.400 | C |
| 8 | 0.398 | 0.087 | 0.511 | 0.457 | B |
| 9 | 0.218 | 0.177 | 0.601 | 0.652 | A |
| 10 | 0.225 | 0.569 | 0.201 | 0.627 | A |
| 11 | 0.580 | 0.087 | 0.329 | 0.160 | C |
| 12 | 0.218 | 0.177 | 0.601 | 0.454 | B |
| 13 | 0.218 | 0.177 | 0.601 | 0.712 | A |
| 14 | 0.136 | 0.339 | 0.520 | 0.543 | B |
| 15 | 0.136 | 0.339 | 0.520 | 0.303 | C |
| 16 | 0.218 | 0.177 | 0.601 | 0.265 | C |
| 17 | 0.218 | 0.177 | 0.601 | 0.335 | C |
| 18 | 0.218 | 0.177 | 0.601 | 0.562 | B |
| 19 | 0.218 | 0.177 | 0.601 | 0.459 | B |
| 20 | 0.218 | 0.177 | 0.601 | 0.364 | C |
| 21 | 0.218 | 0.177 | 0.601 | 0.335 | C |
| 22 | 0.136 | 0.339 | 0.520 | 0.371 | C |
| 23 | 0.136 | 0.339 | 0.520 | 0.406 | C |
| 24 | 0.218 | 0.177 | 0.601 | 0.239 | C |
| 25 | 0.225 | 0.569 | 0.201 | 0.102 | C |
| 26 | 0.218 | 0.177 | 0.601 | 0.238 | C |
| 27 | 0.225 | 0.569 | 0.201 | 0.232 | C |
| 28 | 0.218 | 0.177 | 0.601 | 0.577 | A |
| 29 | 0.136 | 0.339 | 0.520 | 0.742 | A |
| 30 | 0.225 | 0.569 | 0.201 | 0.150 | C |
| 31 | 0.218 | 0.177 | 0.601 | 0.467 | B |
| 32 | 0.136 | 0.480 | 0.378 | 0.180 | C |
| 33 | 0.218 | 0.177 | 0.601 | 0.447 | B |
| 34 | 0.218 | 0.177 | 0.601 | 0.611 | A |
| 35 | 0.136 | 0.339 | 0.520 | 0.270 | C |
| 36 | 0.218 | 0.177 | 0.601 | 0.237 | C |
| 37 | 0.218 | 0.177 | 0.601 | 0.428 | B |
| 38 | 0.136 | 0.339 | 0.520 | 0.280 | C |
| 39 | 0.218 | 0.177 | 0.601 | 0.453 | B |
| 40 | 0.218 | 0.177 | 0.601 | 0.546 | B |
| 41 | 0.183 | 0.244 | 0.567 | 0.115 | C |
| 42 | 0.136 | 0.339 | 0.520 | 0.142 | C |
| 43 | 0.218 | 0.177 | 0.601 | 0.425 | B |
| 44 | 0.136 | 0.339 | 0.520 | 0.246 | C |
| 45 | 0.218 | 0.177 | 0.601 | 0.629 | A |
| 46 | 0.218 | 0.177 | 0.601 | 0.222 | C |
| 47 | 0.218 | 0.177 | 0.601 | 0.405 | C |
| Average | | | | 0.417 | |
| Stdev | | | | 0.192 | |

Table 10: Distance-based method scores and classifications status of items

| Item | Average criteria weights | | | Item Score | Classification |
|---|---|---|---|---|---|
| | $w_1$ | $w_2$ | $w_3$ | | Status |
| 1 | 0.561 | 0.233 | 0.200 | 0.648 | A |
| 2 | 0.284 | 0.662 | 0.052 | 0.973 | A |
| 3 | 0.511 | 0.183 | 0.300 | 0.609 | A |
| 4 | 0.561 | 0.233 | 0.200 | 0.485 | B |
| 5 | 0.561 | 0.233 | 0.200 | 0.462 | B |
| 6 | 0.511 | 0.183 | 0.300 | 0.381 | C |
| 7 | 0.511 | 0.183 | 0.300 | 0.368 | C |
| 8 | 0.382 | 0.183 | 0.429 | 0.433 | B |
| 9 | 0.173 | 0.288 | 0.534 | 0.614 | A |
| 10 | 0.141 | 0.733 | 0.123 | 0.676 | A |
| 11 | 0.511 | 0.183 | 0.300 | 0.143 | C |
| 12 | 0.173 | 0.288 | 0.534 | 0.410 | B |
| 13 | 0.173 | 0.288 | 0.534 | 0.681 | A |
| 14 | 0.095 | 0.446 | 0.456 | 0.548 | A |
| 15 | 0.095 | 0.446 | 0.456 | 0.310 | C |
| 16 | 0.173 | 0.288 | 0.534 | 0.258 | C |
| 17 | 0.173 | 0.288 | 0.534 | 0.301 | C |
| 18 | 0.173 | 0.288 | 0.534 | 0.526 | B |
| 19 | 0.173 | 0.288 | 0.534 | 0.433 | B |
| 20 | 0.173 | 0.288 | 0.534 | 0.356 | C |
| 21 | 0.173 | 0.288 | 0.534 | 0.308 | C |
| 22 | 0.095 | 0.446 | 0.456 | 0.366 | C |
| 23 | 0.095 | 0.446 | 0.456 | 0.412 | B |
| 24 | 0.173 | 0.288 | 0.534 | 0.229 | C |
| 25 | 0.141 | 0.733 | 0.123 | 0.123 | C |
| 26 | 0.173 | 0.288 | 0.534 | 0.228 | C |
| 27 | 0.141 | 0.733 | 0.123 | 0.290 | C |
| 28 | 0.173 | 0.288 | 0.534 | 0.558 | A |
| 29 | 0.095 | 0.446 | 0.456 | 0.742 | A |
| 30 | 0.141 | 0.733 | 0.123 | 0.187 | C |
| 31 | 0.173 | 0.288 | 0.534 | 0.457 | B |
| 32 | 0.095 | 0.687 | 0.215 | 0.200 | C |
| 33 | 0.173 | 0.288 | 0.534 | 0.425 | B |
| 34 | 0.173 | 0.288 | 0.534 | 0.543 | B |
| 35 | 0.095 | 0.446 | 0.456 | 0.276 | C |
| 36 | 0.173 | 0.288 | 0.534 | 0.233 | C |
| 37 | 0.173 | 0.288 | 0.534 | 0.396 | B |
| 38 | 0.095 | 0.446 | 0.456 | 0.290 | C |
| 39 | 0.173 | 0.288 | 0.534 | 0.437 | B |
| 40 | 0.173 | 0.288 | 0.534 | 0.514 | B |
| 41 | 0.144 | 0.347 | 0.505 | 0.112 | C |
| 42 | 0.095 | 0.446 | 0.456 | 0.148 | C |
| 43 | 0.173 | 0.288 | 0.534 | 0.393 | B |
| 44 | 0.095 | 0.446 | 0.456 | 0.247 | C |
| 45 | 0.173 | 0.288 | 0.534 | 0.577 | A |
| 46 | 0.173 | 0.288 | 0.534 | 0.212 | C |
| 47 | 0.173 | 0.288 | 0.534 | 0.362 | C |
| Average | | | | 0.402 | |
| Stdev | | | | 0.181 | |

Table 11: Weighted least-square dissimilarity method scores and classifications status of items

| Item | Average criteria weights | | | Item Score | Classification Status |
|---|---|---|---|---|---|
| | $w_1$ | $w_2$ | $w_3$ | | |
| 1 | 0.566 | 0.229 | 0.200 | 0.622 | A |
| 2 | 0.297 | 0.615 | 0.085 | 0.878 | A |
| 3 | 0.495 | 0.158 | 0.342 | 0.599 | A |
| 4 | 0.566 | 0.229 | 0.200 | 0.476 | B |
| 5 | 0.566 | 0.229 | 0.200 | 0.442 | B |
| 6 | 0.495 | 0.158 | 0.342 | 0.387 | C |
| 7 | 0.495 | 0.158 | 0.342 | 0.374 | C |
| 8 | 0.341 | 0.158 | 0.497 | 0.445 | B |
| 9 | 0.182 | 0.238 | 0.577 | 0.612 | A |
| 10 | 0.153 | 0.687 | 0.157 | 0.616 | A |
| 11 | 0.495 | 0.158 | 0.342 | 0.168 | C |
| 12 | 0.182 | 0.238 | 0.577 | 0.429 | B |
| 13 | 0.182 | 0.238 | 0.577 | 0.662 | A |
| 14 | 0.087 | 0.429 | 0.481 | 0.540 | B |
| 15 | 0.087 | 0.429 | 0.481 | 0.333 | C |
| 16 | 0.182 | 0.238 | 0.577 | 0.294 | C |
| 17 | 0.182 | 0.238 | 0.577 | 0.333 | C |
| 18 | 0.182 | 0.238 | 0.577 | 0.527 | B |
| 19 | 0.182 | 0.238 | 0.577 | 0.445 | B |
| 20 | 0.182 | 0.238 | 0.577 | 0.377 | C |
| 21 | 0.182 | 0.238 | 0.577 | 0.337 | C |
| 22 | 0.087 | 0.429 | 0.481 | 0.376 | C |
| 23 | 0.087 | 0.429 | 0.481 | 0.421 | B |
| 24 | 0.182 | 0.238 | 0.577 | 0.268 | C |
| 25 | 0.153 | 0.687 | 0.157 | 0.129 | C |
| 26 | 0.182 | 0.238 | 0.577 | 0.266 | C |
| 27 | 0.153 | 0.687 | 0.157 | 0.273 | C |
| 28 | 0.182 | 0.238 | 0.577 | 0.551 | A |
| 29 | 0.087 | 0.429 | 0.481 | 0.705 | A |
| 30 | 0.153 | 0.687 | 0.157 | 0.184 | C |
| 31 | 0.182 | 0.238 | 0.577 | 0.463 | B |
| 32 | 0.087 | 0.621 | 0.290 | 0.214 | C |
| 33 | 0.182 | 0.238 | 0.577 | 0.436 | B |
| 34 | 0.182 | 0.238 | 0.577 | 0.542 | A |
| 35 | 0.087 | 0.429 | 0.481 | 0.301 | C |
| 36 | 0.182 | 0.238 | 0.577 | 0.269 | C |
| 37 | 0.182 | 0.238 | 0.577 | 0.412 | B |
| 38 | 0.087 | 0.429 | 0.481 | 0.314 | C |
| 39 | 0.182 | 0.238 | 0.577 | 0.445 | B |
| 40 | 0.182 | 0.238 | 0.577 | 0.514 | B |
| 41 | 0.155 | 0.291 | 0.549 | 0.161 | C |
| 42 | 0.087 | 0.429 | 0.481 | 0.189 | C |
| 43 | 0.182 | 0.238 | 0.577 | 0.409 | B |
| 44 | 0.087 | 0.429 | 0.481 | 0.273 | C |
| 45 | 0.182 | 0.238 | 0.577 | 0.569 | A |
| 46 | 0.182 | 0.238 | 0.577 | 0.251 | C |
| 47 | 0.182 | 0.238 | 0.577 | 0.383 | C |
| Average | | | | 0.409 | |
| Stdev | | | | 0.159 | |

Table 12: Maximizing deviations method scores and classification status of items

| Item | Average criteria weights | | | Item Score | Classification Status |
|---|---|---|---|---|---|
| | $w_1$ | $w_2$ | $w_3$ | | |
| 1 | 0.632 | 0.160 | 0.204 | 0.702 | A |
| 2 | 0.389 | 0.491 | 0.116 | 0.948 | A |
| 3 | 0.567 | 0.095 | 0.334 | 0.664 | A |
| 4 | 0.632 | 0.160 | 0.204 | 0.534 | B |
| 5 | 0.632 | 0.160 | 0.204 | 0.486 | B |
| 6 | 0.567 | 0.095 | 0.334 | 0.407 | B |
| 7 | 0.567 | 0.095 | 0.334 | 0.392 | B |
| 8 | 0.362 | 0.095 | 0.539 | 0.444 | B |
| 9 | 0.195 | 0.178 | 0.623 | 0.621 | A |
| 10 | 0.223 | 0.573 | 0.199 | 0.618 | A |
| 11 | 0.567 | 0.095 | 0.334 | 0.152 | C |
| 12 | 0.195 | 0.178 | 0.623 | 0.418 | B |
| 13 | 0.195 | 0.178 | 0.623 | 0.669 | A |
| 14 | 0.119 | 0.330 | 0.546 | 0.525 | B |
| 15 | 0.119 | 0.330 | 0.546 | 0.298 | C |
| 16 | 0.195 | 0.178 | 0.623 | 0.255 | C |
| 17 | 0.195 | 0.178 | 0.623 | 0.308 | C |
| 18 | 0.195 | 0.178 | 0.623 | 0.520 | B |
| 19 | 0.195 | 0.178 | 0.623 | 0.427 | B |
| 20 | 0.195 | 0.178 | 0.623 | 0.345 | C |
| 21 | 0.195 | 0.178 | 0.623 | 0.308 | C |
| 22 | 0.119 | 0.330 | 0.546 | 0.353 | C |
| 23 | 0.119 | 0.330 | 0.546 | 0.391 | C |
| 24 | 0.195 | 0.178 | 0.623 | 0.225 | C |
| 25 | 0.223 | 0.573 | 0.199 | 0.102 | C |
| 26 | 0.195 | 0.178 | 0.623 | 0.224 | C |
| 27 | 0.223 | 0.573 | 0.199 | 0.227 | C |
| 28 | 0.195 | 0.178 | 0.623 | 0.540 | A |
| 29 | 0.119 | 0.330 | 0.546 | 0.705 | A |
| 30 | 0.223 | 0.573 | 0.199 | 0.147 | C |
| 31 | 0.195 | 0.178 | 0.623 | 0.439 | B |
| 32 | 0.119 | 0.469 | 0.407 | 0.175 | C |
| 33 | 0.195 | 0.178 | 0.623 | 0.414 | B |
| 34 | 0.195 | 0.178 | 0.623 | 0.548 | A |
| 35 | 0.119 | 0.330 | 0.546 | 0.260 | C |
| 36 | 0.195 | 0.178 | 0.623 | 0.224 | C |
| 37 | 0.195 | 0.178 | 0.623 | 0.391 | B |
| 38 | 0.119 | 0.330 | 0.546 | 0.271 | C |
| 39 | 0.195 | 0.178 | 0.623 | 0.422 | B |
| 40 | 0.195 | 0.178 | 0.623 | 0.503 | B |
| 41 | 0.164 | 0.239 | 0.591 | 0.108 | C |
| 42 | 0.119 | 0.330 | 0.546 | 0.138 | C |
| 43 | 0.195 | 0.178 | 0.623 | 0.388 | C |
| 44 | 0.119 | 0.330 | 0.546 | 0.234 | C |
| 45 | 0.195 | 0.178 | 0.623 | 0.571 | A |
| 46 | 0.195 | 0.178 | 0.623 | 0.205 | C |
| 47 | 0.195 | 0.178 | 0.623 | 0.363 | C |
| Average | | | | 0.396 | |
| Stdev | | | | 0.184 | |

## Footnotes

[1] In some cases, performance indicators take the form of inputs and outputs as in productive efficiency analysis while in other cases they reflect different aspects of performance as in the construction of composite indicators.

[2] The distance-based method has been employed previously by Wu, Sun and Liang (2012) for aggregating cross efficiencies. It may be considered as a modification of Wang and Wang (2013) weighted least-square deviation approach.

[3] According to Kao *et al.* (2008), when the values of the performance indicators are normalized in the range of [0,1] then it is necessary to have the sum of the resulting BoD weights to be greater than or equal to one.

[4] If evaluation scores and ranks vary considerably across possible ordering schemes then an overall measure may not be so useful but it is still preferable than relying on the results of a particular ordering scheme.

[5] Somewhat different approaches are used by Ladhari, Babai and Lajili (2016), i.e. a constructive order classification algorithm, and Li *et al.* (2018), i.e., stochastic multi-criteria acceptability analysis, based on most and least favorable evaluation under each descending order scheme.

[6] This normalization presumes that all indicators are "good" in the sense that higher values mean better performance. If some indicators are "bad" in the sense that lower values mean better performance, one should then used the following normalization: $y_j^k = \left(\max_k I_j^k - I_j^k\right) / \left(\max_k I_j^k - \min_k I_j^k\right)$.

[7] If however a DMU performs well on all indicators then the weighting scheme does not really matter.

[8] Ramanathan (2006) initially considered four classification criteria but Ng (2007) disregarded the criterion of critical factor (1, 0.5 and 0.01 for very-, moderate- and non-critical) because as a categorical variable is not suitable for linear optimization models.

[9] Notice that each of the aforementioned efficient items gets the score of one under two ordering schemes.

[10] We have only small differences in the estimated overall measure but no differences in the classification status compared to Zheng *et al.* (2017).

[11] Each method identifies nine different average criteria weights profiles; see Tables 7, 9, 10 and 11. Four of them are assigned to a single item, namely items #2, #8, #32

and #41, one to three items, namely items #1, #4 and #5, two to four items each, namely items #3, #6, #7 and #11 and items #10, #25, #27 and #30, and one to almost half of the items (23 out of 47), namely items #9, #12, #13, #16, #17, #18, #19, #20, #21, #22, #24, #26, #28, #31, #33, #34, #36, #37, #39, #40, #43, #45, #46 and #47. The values of average criteria weights in these profiles differ across methods but the items in each profile remain the same in all methods. For example, the 23 items with average criteria weights of (0.218, 0.177, 0.601) in the Shannon entropy method are the same 23 items with average criteria weights of (0.193, 0.193, 0.610) in the partial average cross-weight method and the 23 items with average criteria weights of (0.173, 0.288, 0.534) in the distance-based method and the 23 items with average criteria weights of (0.195, 0.178, 0.623) in the maximizing deviations method.

[12] In the last two columns we reproduce the results of Fu et al. (2016).

[13] We have only small differences in the estimated overall measure but no differences in the classification status compared to Wu *et al*. (2018).