Libri, 2003, vol. 53, pp. 282–289
Printed in Germany · All rights reserved

Libri
ISSN 0024-2667

Ranking University Libraries with A Posteriori *Weights*

CHIANG KAO AND HSI-TAI HUNG
Department of Industrial and Information Management, National Cheng Kung University, Tainan, Taiwan, Republic of China

In aggregating the wide range of resources and services provided by the university libraries to produce a single composite index for comparison, two types of weights representing the relative importance of the criteria have been discussed in the literature. One is *a priori* weights assigned by the experts in a subjective manner before the scores for the criteria are measured. The other is *a posteriori* weights determined objectively from the scores measured for the criteria. Taking into account the opinion of experts, this paper proposes an *a posteriori* approach based on the concept of least squares to compare university libraries. Since the characteristics of the libraries as revealed by their performance scores on each criterion have been considered in calculating the composite indices for comparison, the results are more representative and are thus more acceptable to the libraries being compared. To illustrate this approach, university libraries in Taiwan are ranked using the proposed approach. The composite indices calculated from this approach are greater and more concentrated than those calculated from the *a priori* approach. Psychologically, this is very attractive to the libraries being ranked, although most libraries have similar ranks determined from these two approaches.

Introduction

University libraries bear the responsibility for providing sufficient resources and services for teaching and research purposes. Much effort has been devoted to establishing standards and performance measures to ensure the quality of the libraries (see, for example, Lynch 1987, Kania 1988, Hendrickson 1989, Lines 1989, Cotta-Schønberg and Line 1994, ACRL 1995, Chen 1997, Kao and Lin 2001). Different methods also have been devised to determine how well the library meets the educational and information needs of its clients. One type of method is to measure the library size in terms of the resources provided (ARL 1995, Kao et al. 1998, Kao and Lin 1999), and the libraries are ranked according to their sizes.

Based on standards or performance measures, the resources and services can be measured quantitatively. Although quantitative measures have the advantage of objectivity, lucidity, and easy calculation, they are confined to assessing individual factors. Since the university library involves a wide range of resources and services, any attempt to consolidate measures of those individual factors to a single index requires a decision about the relative importance of those factors (Van House 1989). With the relative importance of the factors, usually expressed as cardinal weights, the weighted sum can be calculated for each library as a composite index for comparison. In order to yield meaningful composite indices, the process used to obtain the relative importance of the factors is a key issue.

The weights can be elicited from two fundamentally different ways (Zeleny 1982). One is direct explication, in that the weights are collected from the experts via the procedure of interviews, questionnaire surveys or the like. The other is indirect explication, in that the weights are determined via regression analysis or mathematical programming based on the observed circumstance. The study of Kao et al. (1998) is a typical example of the former. The directors of twenty-four university libraries in Taiwan served as the experts. Their opinions regarding the importance

Chiang Kao, PhD., Professor. Department of Industrial and Information Management, National Cheng Kung University, Tainan, Taiwan, Republic of China. Tel: (+886) 6-275 3396, Fax: (+886) 6-236 2162. E-mail: ckao@mail.ncku.edu.tw

of each factor, or criterion in the jargon of multiple criteria decision analysis were gathered and averaged to represent the weight. This is a human-based approach that produces *a priori* weights in a subjective manner. Kao et al. (1998) also formulated a linear program to find the weights that would produce the largest weighted sum for each library. This approach is a data-based approach that produces *a posteriori* weights in an objective manner. Both approaches have pros and cons.

This paper proposes a two-stage approach keeping the pros and eliminating the cons of the two approaches. The university libraries in Taiwan are adopted for illustration. Following the study of Kao et al. (1998) and Kao and Lin (1999), five primary criteria, namely, collections, personnel, expenditures, buildings, and services, are considered. Basically, the approach is a data-based *a posteriori* approach. However, the opinion of the experts is also taken into account. Therefore, the rankings resulted from the two-stage approach will suitably reflect the real status of the libraries and are more representative to the libraries to be compared.

This paper is the fourth paper in a series on the ranking of university libraries. The first paper ranks university libraries (using *a priori* weights) without considering university size (Kao et al. 1998). The second takes university size into account (Kao and Lin 1999). The third establishes standards for university libraries (Kao and Lin 2001). The fourth ranks the university libraries using *a posteriori* weights.

In the following, the two-stage approach which combines the subjective opinion of the experts and the objective information contained in the data is described. Then, the university libraries in Taiwan are ranked using the method proposed in this paper. Finally, the results obtained from the *a posteriori* approach are compared with those of the *a priori* approach with analyses and discussions.

The two-stage method

Suppose there are m criteria to be considered in ranking n university libraries. The measures in all criteria must be aggregated to form an overall measurement for comparison. Due to the difference in the units measured from the criteria, the scores may need standardization. Standardization is done by dividing each score by the largest score observed in that criterion. The standardized scores are dimensionless and range from 0 to 1, which makes them comparable among different criteria. The direct explication method starts with obtaining the relative importance of the criteria. Some famous methods include ratio-scale weights, rank-order centroid weights and rank-sum weights (Jia et al. 1998). The weights determined by experts are then used to yield a single value. In most cases, the average of those weights solicited from the experts is used to represent the relative importance of each criterion. Let S_{ij} be the standardized score of the ith library in the jth criterion and w_j be the average weight of the jth criterion. Then the weighted sum $\Sigma_{j=1}^{m} w_j S_{ij}$ is the overall performance measure of the ith library. Note that the sum of the weights $\Sigma_{j=1}^{m} w_j$ is equal to 1. The study of Kao et al. (1998) is an example of this method for ranking university libraries.

Using the average weights solicited from the experts to represent the relative importance of the criteria has some problems. One is subjectivity, in that not everyone agrees with the weights, especially when some library performs very well in a specific criterion while the experts assign a low weight to that criterion. Another problem is that the average weights are solicited from the experts before the data are observed. The experts may be able to specify the weights according to their perceptions. Sometimes, however, the experts may have limited information about the real status of the libraries. Due to these problems, an *a posteriori* method is proposed in this paper.

The method has two stages in generating the weights. In the first stage, the sets of weights most favourable to individual libraries in terms of the composite index are determined from a set of linear programs. From the most favourable indices for different libraries found in the first stage, a set of general consensus weights based on the concept of least squares is derived in the second stage.

Denote S_{ij} as the standardized score of the ith library in the jth criterion. From questionnaire surveys, the weights for all criteria perceived by the experts are collected. Let l_j and u_j denote the smallest and largest weights, respectively, observed in the jth criterion. Since range restriction is the most straightforward way to incorporate

the experts' opinion into the weighting procedure (Halme and Korhonen 2000), in the first stage we allow each library to select the weight for each criterion within the range of $[l_j, u_j]$, to produce the largest weighted sum. This idea can be formulated as the following linear program:

$$\begin{aligned} &\max I_i = \sum_{j=1}^{m} w_{ij} S_{ij} \\ &\text{s.t.} \sum_{j=1}^{m} w_{ij} = 1 \\ &l_j \le w_{ij} \le u_j,\ j=1,...,m, \end{aligned} \tag{1}$$

where the variables w_{ij}, $j=1,...,m$, are the weights to be assigned to the jth criterion by the ith library. From the optimal weights w_{ij}, the optimal overall performance index for the ith library is calculated as $I_i^*=\Sigma_{j=1}^{m} w_{ij}^* S_{ij}$, which is the optimal objective value of (1).

Each library is allowed to select their own weights within the bounds imposed by the experts in the first stage. Consequently, the optimal overall performance index I_i^* is the largest possible score that the ith library could obtain. Any other set of weights would result in an index smaller than I_i^*. The idea of this stage is identical to the first part of the method of multi-attribute choice proposed by Doyle (1995). In the next step of that method, the average of those optimal sets of weights found in the first step is calculated to be the set of weights for calculating the weighted sum for all alternatives. This mechanical calculation of the average does not fully account for the trend of the data. The second stage of the method proposed in this paper is to find a set of weights which will yield the composite indices closest to the most favourable indices I_i^* found in the first stage. The idea adopted in this paper for compromising the overall performance indices calculated from the common set of weights and the individual sets of weights is least squares.

Let w_j denote the weights which will yield an index $I_i=\Sigma_{j=1}^{m} w_j S_{ij}$ for the ith library such that the total squared-difference between I_i^* and I_i is minimum. This idea can be cast into the following quadratic program:

$$\begin{aligned} &\min \sum_{i=1}^{n} \left(I_i^* - \sum_{j=1}^{m} w_{ij} S_{ij} \right)^2 \\ &\text{s.t.} \sum_{j=1}^{m} w_j = 1 \\ &l_j \le w_j \le u_j,\ j=1,...,m. \end{aligned} \tag{2}$$

This program can be considered as a constrained least-squares problem, where I_i^* is the observed value and $\Sigma_{j=1}^{m} w_j S_{ij}$ is the expected value. The weights w_j are the parameters to be estimated under the unity constraint $\Sigma_{j=1}^{m} w_j=1$ and bound constraints $l_j \le w_j \le u_j$, $j=1,...,m$. The optimal solutions w_j^* are the general consensus weights for all libraries. The compromised composite index of each library as calculated from the general consensus weights: $CI_i^*=\Sigma_{j=1}^{m} w_j^* S_{ij}$, is then the index for comparison.

In minimizing the objective function of Model (2), larger weights would be assigned to the criteria that are more advantageous to the libraries because they will bring the subsequently calculated composite indices $\Sigma_{j=1}^{m} w_j S_{ij}$ closer to the theoretical ceiling value I_i^*. This is similar to regression analysis where variables with higher explanatory power will have relatively larger regression coefficients in the statistical sense. On the contrary, the weights will be small for those criteria with low scores. Notably, criteria with higher scores are the ones emphasized by the libraries while those with lower scores are the ones considered by the libraries as less important. Since the weights w_j^* take into account both the opinion of the experts before the data are observed and the information contained in the data collected from the libraries, they are more representative and are thus more acceptable to the libraries to be compared.

University libraries in Taiwan

As reported in Kao et al. (1998), Taiwan had twenty-four universities in 1998, excluding independent colleges and institutes. In that study five primary criteria for comparing the university libraries, namely, *collections*, *personnel*, *expenditures*, *buildings*, and *services*, were determined from a thorough discussion of eleven library directors. Each primary criterion had several subordinate secondary criteria while some of the secondary criteria were further classified into third-level criteria. The deepest level was the fourth. The left part of Table 1 shows the criteria of different levels.

To determine the weights of the criteria at all levels, Kao et al. (1998) asked the twenty-four directors to assign *a priori* weights subjectively. For the criteria under the same criterion of higher

Table 1. Criteria and weights for ranking the university libraries in Taiwan.

Criteria of different levels	Weight range	*A priori* wt. (*A posteriori*)
Collections	[0.18 0.35]	0.23 (0.30)
Volumes held	[0.20 0.94]	0.65 (0.26)
Books	[0.19 0.40]	0.30 (0.40)
Serials	[0.20 0.50]	0.31 (0.24)
Current titles	[0.20 0.50]	0.37 (0.50)
Back titles	[0.04 0.40]	0.23 (0.09)
Total titles	[0.10 0.40]	0.20 (0.10)
Bound volumes	[0.00 0.35]	0.20 (0.31)
Microforms	[0.00 0.22]	0.10 (0.11)
Audio-visual works	[0.00 0.18]	0.10 (0.15)
Database	[0.10 0.30]	0.19 (0.10)
Volumes added	[0.06 0.80]	0.35 (0.74)
Books	[0.19 0.40]	0.30 (0.40)
Serials	[0.20 0.50]	0.31 (0.48)
Current titles	[0.20 0.50]	0.37 (0.20)
Back titles	[0.04 0.40]	0.23 (0.05)
Total titles	[0.10 0.40]	0.20 (0.40)
Bound volumes	[0.00 0.35]	0.20 (0.35)
Microforms	[0.00 0.22]	0.10 (0.00)
Audio-visual works	[0.00 0.18]	0.10 (0.00)
Database	[0.10 0.30]	0.19 (0.12)
Personnel	[0.12 0.35]	0.20 (0.12)
Tenure-tracked staff	[0.40 0.80]	0.63 (0.61)
Professional	[0.40 0.90]	0.73 (0.45)
Parallel professional	[0.10 0.60]	0.27 (0.55)
Non-tenure-tracked staff	[0.00 0.40]	0.11 (0.00)
Professional	[0.40 0.90]	0.73 (0.54)
Parallel professional	[0.10 0.60]	0.27 (0.46)
Student assistants	[0.05 0.35]	0.16 (0.29)
Unclassified staff	[0.00 0.25]	0.10 (0.10)
Expenditures	[0.16 0.35]	0.21 (0.16)
Operating expenditures	[0.10 0.50]	0.33 (0.21)
Capital expenditures	[0.40 0.80]	0.57 (0.78)
Special expenditures	[0.00 0.20]	0.10 (0.01)
Buildings	[0.06 0.23]	0.15 (0.07)
Main library area	[0.20 0.75]	0.49 (0.48)
Main library seats	[0.10 0.50]	0.30 (0.50)
Branches area	[0.00 0.30]	0.10 (0.02)
Branches seats	[0.00 0.30]	0.10 (0.00)
Services	[0.11 0.35]	0.22 (0.35)
Range of services	[0.10 0.65]	0.18 (0.65)
Communication channels	[0.05 0.15]	0.11 (0.10)
Remote access	[0.05 0.20]	0.14 (0.05)
Operating hours	[0.05 0.15]	0.11 (0.07)
Attendance	[0.04 0.15]	0.11 (0.04)
Circulation	[0.05 0.20]	0.13 (0.05)
Amount of services	[0.02 0.15]	0.10 (0.02)
Database searching	[0.02 0.15]	0.09 (0.02)
Secondary users	[0.00 0.10]	0.05 (0.00)

level, their weights are required to sum to 1. The average of the weights solicited from the twenty-four directors for each criterion is designated as the weight of that criterion. Starting from the lowest level, the weighted sum of the criteria under the same criterion of higher level, that is, the sum of the scores multiplied by their corresponding weights, is calculated. This weighted sum becomes the score of that higher-level criterion. We then calculate the weighted sum for the higher-

level criteria under the same criterion of still higher level. This process is continued until the top level of the five primary criteria is reached, where the composite index of a university library is calculated for comparison. Kao et al. (1998) have a detailed description of this *a priori* approach. Table 1 also shows the average weights for the criteria of all levels.

As opposed to the *a priori* approach, the *a posteriori* approach described in the preceding section has two stages to produce the general consensus weights for calculating the composite indices. Firstly, the smallest and largest weights for each criterion solicited from the twenty-four directors are summarized in Table 1 under the heading of "Weight range". Similar to the *a priori* approach, for each primary criterion we start from the lowest level to calculate the weighted sum for each library by applying Model (1). From the weighted sums of the libraries calculated from Model (1), a set of general consensus weights for the criteria under the same criterion of higher level can be determined from Model (2). The new weighted sum is then calculated from the general consensus weights, which becomes the score of the higher-level criterion. This process is repeated from the bottom of the tree structure of the criteria to the top until the final composite indices of all libraries are obtained.

As an illustration, consider the primary criterion *collections*. The lowest level criteria are *current titles*, *back titles*, *total titles*, and *bound volumes* under the higher-level criterion *serials*. Note that there are two criteria named *serials*, one under the secondary criterion *volumes held* and the other under *volumes added*. For the *serials* under *volumes held*, the associated Model (1) for calculating the most favourable weighted sum for the *i*th library is:

$$\max I_i = w_{i1}CT_i + w_{i2}BT_i + w_{i3}TT_i + w_{i4}BV_i$$

$$\text{s.t. } w_{i1} + w_{i2} + w_{i3} + w_{i4} = 1$$

$$0.37 \le w_{i1} \le 0.50$$

$$0.04 \le w_{i2} \le 0.40$$

$$0.10 \le w_{i3} \le 0.40$$

$$0.00 \le w_{i4} \le 0.35,$$

where CT_i, BT_i, TT_i, and BV_i are the standardized scores of *current titles*, b*ack titles*, t*otal titles*, and *bound volumes*, respectively, for the *i*th library. The bounds for the weights are taken from Table 1. The optimal objective value I_i^* is the most favourable weighted sum of the criterion *serials* under the secondary criterion *volumes held* for the *i*th library. Since the sum of the four weights is equal to 1 and all standardized scores lie between 0 and 1, the weighted sum I_i also lies between 0 and 1.

Model (1) is executed once for each library. After all twenty-four libraries are enumerated, the most favourable weighted sums for *serials* are obtained for all libraries. We then apply Model (2) to determine the general consensus weights as follows:

$$\min \sum_{i=1}^{24} [I_i^* - (w_1CT_i + w_2BT_i + w_3TT_i + w_4BV_i)]^2$$

$$\text{s.t. } w_1 + w_2 + w_3 + w_4 = 1$$

$$0.37 \le w_1 \le 0.50$$

$$0.04 \le w_2 \le 0.40$$

$$0.10 \le w_3 \le 0.40$$

$$0.00 \le w_4 \le 0.35.$$

The weighted sum $(w_1^*CT_i + w_2^*BT_i + w_3^*TT_i + w_4^*BV_i)$ is the score of serials under volumes held for the ith library. Together with the scores of books, microforms, audio-visual works, and database, the score of volumes held can be calculated by applying Models (1) and (2) in sequence. Likewise, the score of volumes added can be calculated. The weighted sum of volumes held and volumes added is then calculated to represent the score of collections. Finally, the composite index for the ith library is calculated from the scores of collections, personnel, expenditures, buildings, and services in a similar way.

The general consensus weights for all criteria are recorded in Table 1 under the heading of "*A posteriori* weight". The scores of the five primary criteria and the final composite index for each library calculated from both the *a priori* weights and *a posteriori* weights are listed in Table 2. The letter *N* appearing in some universities is an abbreviation for *National*, indicating that the associated university is public. In contrast, a university without *N* in front is private.

Table 2. Scores and rankings of the twenty-four university libraries in Taiwan calculated from the two-stage method and the average-weight method (in parentheses).

University	Collections	Personnel	Expenditures	Buildings	Services	Composite Index	Rank
N Taiwan U	0.85 (1.00)	1.00 (1.00)	1.00 (1.00)	0.73 (0.86)	0.95 (0.98)	0.92 (0.98)	1 (1)
N Cheng Kung U	0.99 (0.67)	0.79 (0.81)	0.77 (0.74)	0.30 (0.36)	1.00 (0.92)	0.89 (0.72)	2 (2)
N Chung Cheng U	1.00 (0.57)	0.29 (0.22)	0.51 (0.50)	1.00 (1.00)	0.78 (0.55)	0.76 (0.54)	3 (5)
N Cheng Chi U	0.71 (0.59)	0.43 (0.51)	0.66 (0.63)	0.66 (0.69)	0.96 (1.00)	0.75 (0.69)	4 (3)
N Tsing Hua U	0.68 (0.42)	0.30 (0.34)	0.63 (0.58)	0.33 (0.34)	0.94 (0.76)	0.69 (0.50)	5 (6)
Tamkang U	0.51 (0.33)	0.36 (0.41)	0.51 (0.62)	0.81 (0.79)	0.99 (0.87)	0.68 (0.59)	6 (4)
N Taiwan Normal U	0.83 (0.70)	0.35 (0.37)	0.18 (0.21)	0.35 (0.41)	0.88 (0.71)	0.65 (0.49)	7 (8)
Feng Chia U	0.42 (0.29)	0.67 (0.46)	0.49 (0.45)	0.66 (0.59)	0.86 (0.70)	0.63 (0.49)	8 (7)
Tunghai U	0.37 (0.27)	0.55 (0.37)	0.32 (0.30)	0.71 (0.62)	0.95 (0.71)	0.61 (0.44)	9 (10)
N Central U	0.49 (0.32)	0.25 (0.20)	0.43 (0.41)	0.50 (0.48)	0.88 (0.79)	0.59 (0.44)	10 (11)
N Chung Hsing U	0.34 (0.31)	0.41 (0.31)	0.53 (0.49)	0.37 (0.44)	0.89 (0.71)	0.57 (0.45)	11 (9)
Soochow U	0.39 (0.26)	0.35 (0.30)	0.55 (0.52)	0.32 (0.32)	0.87 (0.71)	0.57 (0.43)	12 (12)
Chinese Culture U	0.58 (0.45)	0.34 (0.27)	0.40 (0.39)	0.21 (0.18)	0.71 (0.47)	0.54 (0.36)	13 (17)
Fu Jen Catholic U	0.66 (0.43)	0.39 (0.31)	0.10 (0.15)	0.56 (0.54)	0.68 (0.53)	0.54 (0.39)	14 (14)
Providence U	0.47 (0.26)	0.37 (0.36)	0.21 (0.19)	0.29 (0.28)	0.84 (0.72)	0.53 (0.37)	15 (16)
N Sun Yat-Sen U	0.44 (0.28)	0.24 (0.23)	0.45 (0.40)	0.51 (0.52)	0.75 (0.56)	0.53 (0.39)	16 (13)
N Chiao Tung U	0.46 (0.33)	0.11 (0.13)	0.50 (0.44)	0.22 (0.21)	0.80 (0.68)	0.53 (0.37)	17 (15)
Chung Yuan U	0.37 (0.24)	0.32 (0.28)	0.31 (0.28)	0.43 (0.40)	0.78 (0.59)	0.50 (0.36)	18 (18)
N Chi Nan U	0.79 (0.36)	0.04 (0.04)	0.11 (0.11)	0.06 (0.05)	0.63 (0.41)	0.48 (0.21)	19 (23)
N Changhua U of Edu.	0.44 (0.25)	0.20 (0.23)	0.24 (0.30)	0.37 (0.36)	0.71 (0.52)	0.47 (0.33)	20 (19)
N Taiwan Ocean U	0.15 (0.11)	0.13 (0.11)	0.26 (0.23)	0.26 (0.24)	0.81 (0.64)	0.40 (0.27)	21 (20)
N Yang Ming U	0.13 (0.08)	0.11 (0.11)	0.14 (0.13)	0.14 (0.14)	0.91 (0.65)	0.40 (0.23)	22 (22)
N Kaohsiung Normal U	0.24 (0.23)	0.14 (0.14)	0.21 (0.21)	0.15 (0.16)	0.75 (0.53)	0.39 (0.26)	23 (21)
N Dong Hwa U	0.17 (0.11)	0.06 (0.08)	0.18 (0.15)	0.03 (0.03)	0.84 (0.57)	0.38 (0.20)	24 (24)
Average	0.52 (0.34)	0.34 (0.29)	0.40 (0.37)	0.42 (0.40)	0.84 (0.66)	0.58 (0.44)	

Analyses

Regarding the *a priori* weights and *a posteriori* weights shown in Table 1, there are several remarks to make. Firstly, *a priori* weights are the average weights assigned subjectively by the experts whereas *a posteriori* weights are the objective weights obtained from the scores of the libraries. The primary criterion *collections* has two secondary criteria, viz., *volumes held* and *volumes added*. The weights subjectively assigned by the experts indicate that *volumes held* is more important than *volumes added*. However, the weights determined by the observed data show an opposite result. This is because data of the same criterion have been standardized by dividing by the largest observation to lie between 0 and 1. Most universities have larger standardized values in *volumes added* and smaller values in *volumes held*.

The secondary criteria *volumes held* and *volumes added* have the same subordinate third-level criteria of *books, serials, microforms, audio-visual works,* and *database*. And the fourth-level criteria under the third-level criterion *serials* are also the same. The weights subjectively assigned by the experts for the same third-level and fourth-level criteria under these two secondary criteria are the same. From the data collected from the libraries, it is observed that in relative not many microforms or audio-visual works are added, indicating they are not as important as other third-level criteria. The weights of these two criteria determined from the two-stage method reach the lower bound of 0. In contrast, the serials added has more amount relatively than that held in library. Therefore, the former has a larger weight than the latter, i.e., 0.48 versus 0.24. For criterion *serials*, the amount of current titles held is more important than that of current titles added. They differ by 0.3. Another point worthwhile to note is that the criterion *total titles* is the most important fourth-level criterion for *volumes added*, whereas it is one of the least important criteria for *volumes held*.

For primary criterion *personnel*, there are two secondary criteria: *tenure-tracked staff* and *non-*

tenure-tracked staff, which have the same third-level criteria. These two identical sets of criteria have the same *a priori* weights assigned by the experts yet different *a posteriori* weights calculated from the two-stage method. This is because the values under these two categories are of different proportion. In the experts' opinion, the professional staff plays a more important role in the library administration. However, the data collected from the libraries show that professional staff and parallel professional staff are of similar importance. For the third primary criterion *buildings*, its subordinate criteria *branches area* and *branches seats* have negligible *a posteriori* weights. The reason is because not many university libraries in Taiwan have branches. Finally, for *services*, its secondary criterion *range of services* has very large *a posteriori* weight compared with *a priori* weight, i.e., 0.65 versus 0.18. This is simply because most university libraries perform relatively well in this criterion.

At the top level, the weights assigned by the experts show that *collections*, *personnel*, *expenditures*, *buildings*, and *services* are of similar importance with the weights of 0.23, 0.20, 0.21, 0.15, and 0.22, respectively. The two-stage method, on the other hand, indicates that *services* and *collections* are more important, with the weights of 0.35 and 0.30, respectively. The other three criteria have smaller *a posteriori* weights than *a priori* weights. *Buildings* is the least important criterion with a weight as low as 0.07. The largest difference in weights between the average-weight method and the two-stage method occurs at *services*, which is 0.13.

In sum, the criteria with better performance tend to have larger *a posteriori* weights and consequently larger composite indices. This characteristic makes this approach psychologically more attractive to the libraries to be compared.

Table 2 contains the scores of the primary criteria and the composite indices of the libraries calculated from both the *a priori* approach and the *a posteriori* approach. The first column shows the names of the twenty-four universities whose libraries are to be compared. The next five columns list the scores calculated from the two approaches. Scores of the *a priori* approach are copied from Kao et al. (1998) and are in parentheses. Intuitively, the *a posteriori* approach would result in higher scores. Although the averages shown in the last row of Table 2 give a rough idea of this intuition, we still need a statistical test to verify this intuition in a stringent manner.

The traditional *t*-test for comparing the means of two populations requires the observations to follow the normal distribution. In our case, each library is selecting the most favourable weights to push up their composite indices as close to 1 as possible. It is doubtful that the indices will be normally distributed. There is hardly any way to identify the probability distribution, either. Therefore, a nonparametric method is needed.

For cases of two related samples, the Wilcoxon matched-pairs signed-rank test is an appropriate one (Daniel 1978, Neter et al. 1993). The procedure for obtaining the numerical value of the test statistic is as follows.

1. For each library, obtain the difference between the two measures.
2. Rank the absolute values of these differences from smallest to largest.
3. Assign to each of the resulting ranks the signs of the difference whose absolute value yielded that rank.
4. Compute the sum of the ranks with positive signs and denote as T.

When n is greater than 20, the statistic

$$z = \frac{T - [n(n+1)] \ / \ 4}{\sqrt{n(n+1)(2n+1) \ / \ 24}} \qquad (3)$$

has an approximately standard normal distribution. Therefore, we reject the null hypothesis that the difference is less than or equal to 0 if $z \geq z_{(1-\alpha)}$. For detailed description of this method, please refer to Daniel (1978) or Neter et al. (1993).

From the data contained in Table 2, the Wilcoxon test shows that we cannot reject the null hypothesis that the two-stage method produces scores that are less than or equal to scores produced by the average-weight method in criteria *personnel* and *buildings* at the 95% significance level. In other words, the indices produced by the two-stage method are not greater than those produced by the average-weight method for *personnel* and *buildings* in a statistical sense, although the averages are larger. For the other three criteria, the scores produced by the two-stage method are significantly greater than those produced by the average-weight method. These higher scores naturally produce the higher final composite indices for the two-stage method, which is also

verified by the Wilcoxon test. The composite indices calculated from the two-stage method are not only greater than but also have smaller variance than those calculated from the average-weight method. Specifically, the former produces scores with a standard deviation of 0.1438, while the latter a standard deviation of 0.1774.

Since all libraries have a chance to express their opinion regarding the importance of the criteria through their scores in the criteria, the final composite indices and the associated rankings are more acceptable to them. The last column of Table 2 shows the ranks of the twenty-four libraries derived from the two methods. The difference is not much. Seven libraries have the same rank, nine have their ranks differ by 1, and five have their ranks differ by 2. Only three libraries have a rank difference of greater than 2. More specifically, Chinese Culture University has improved its rank by 4. This is mainly due to its large improvement in the score of *services*. National Chi Nan University also has improved its rank by 4. The scores in Table 2 reveal that this improvement is due to the large increase in the score of *collections*. As a matter of fact, this university is the most recently established university among the twenty-four universities; it does not have as much an amount of *volumes held*. However, it has relatively large amount of *volumes added*. Referring to Table 1, the average-weight method assigns weights of 0.65 and 0.35, respectively, to these two secondary criteria under the primary criterion *collections*, while the two-stage method produces the weights of 0.26 and 0.74, respectively, to these two secondary criteria. This big difference in weights improves its score in *collections* and consequently the rank.

The third library, which has a large change in ranking, is National Sun Yat-Sen University, whose rank has dropped by 3. If we compare the scores of this university with the averages of the twenty-four libraries for the five primary criteria, we find that this university does not make comparable improvement in each criterion when the two-stage method is applied. For the criterion *buildings*, its score has even decreased. Therefore, it is not surprising that its rank drops.

From the scores listed in Table 2, each university library is able to identify its strong area and weak area and to make subsequent effort to improve its rank. For other universities not being ranked in Table 2, the weights contained in Table 1 can be utilized to calculate the composite indices for making comparisons with the twenty-four university libraries.

Conclusion

The major goal of comparing the university libraries is to let the top management of the libraries know how their libraries are performing in terms of the resources and services provided to the clients compared to other libraries. There are two essential factors, i.e., criteria and their associated importance, which must be considered in the multiple criteria analysis. This paper determines and discusses the relative importance of the criteria, assuming the criteria are given. A two-stage method, which takes into account both the experts' opinion and the individual libraries' opinion of the relative importance of the criteria, is proposed. The library's opinion is expressed through the scores in the criteria. Since the feature of this two-stage method is to determine the weights with focus on the observed data of the libraries, it belongs to *a posteriori* approach, and the weights derived are *a posteriori* weights.

There are several merits of the *a posteriori* weights determined from the two-stage method. One is they produce higher composite indices than those produced by the conventional *a priori* weights. Moreover, the variance of the former is smaller than that of the latter. This makes the libraries feel more comfortable because the indices are higher and are not far away from the leading libraries. Another merit is that the *a posteriori* weights are partially determined by the data collected from the libraries, which suitably reflect the reality. The example of the twenty-four university libraries in Taiwan confirms these merits. Due to these merits, the rankings determined by the two-stage method of this paper are more convincing and more acceptable to the libraries being compared.

References

Association of College and Research Libraries. 1995. Standards for college libraries, 1995 ed. *College & Research Libraries News*, 56 (4): 245–57.

Association of Research Libraries. 1995. Holdings of university research libraries in US and Canada. *The Chronicle of Higher Education*, September 1: 29.

Chen, T.Y. 1997. A measurement of the resource utilization efficiency of university libraries. *International Journal of Production Economics,* 53: 71–80.

Cotta-Schønberg, M. and M.B. Line. 1994. Evaluation of academic libraries: with special reference to the Copenhagen-Business-School-Library. *Journal of Librarianship and Information Science,* 26: 55–69.

Daniel, W.W. 1978. *Applied Nonparametric Statistics*. Boston, Mass.: Houghton Mifflin Co.

Doyle, J.R. 1995. Multiattribute choice for the lazy decision maker: let the alternative decide. *Organizational Behavior and Human Decision Processes,* 62 (1): 87–100.

Halme, M. and P. Korhonen. 2000. Restricting weights in value efficiency analysis. *European Journal of Operational Research,* 126: 175–88.

Hendrickson, K. 1989. Standards for university libraries: evaluation of performance. *College Research Libraries News,* 50 (8): 679–91.

Jia, J., G.W. Fischer, and J.S. Dyer. 1998. Attribute weighting methods and decision quality in the presence of response error: a simulation study. *Journal of Behavioral Decision Making,* 11: 85–105.

Kania, A.M. 1998. Academic library standards performance measures. *College Research Libraries,* 49 (1): 16–23.

Kao, C., Y.C. Lin, L.C. Liang, and S.C. Lo. 1998. Ranking university libraries: the Taiwan case. *Libri,* 48: 212–23.

Kao, C. and Y.C. Lin. 1999. Comparing university libraries of different university size. *Libri,* 49: 150–8.

Kao, C. and Y.C. Lin. 2001. Empirical standards for university libraries in Taiwan. *Libri,* 51: 17–26.

Lines, L. 1989. Performance measurement in academic libraries – a university perspective. *British Journal of Academic Librarianship,* 4: 111–20.

Lynch, B.P. 1987. Standards for university libraries. *IFLA Journal,* 13: 120–25.

Neter, J., W. Wasserman, and G.A. Whitmore. 1993. *Applied Statistics*. 4th ed. Boston, Mass.: Allyn and Bacon.

Van House, N.A. 1989. Output measures in libraries. *Library Trends,* 38 (2): 268–79.

Zeleny, M. 1982. *Multiple Criteria Decision Making*. New York, NY: McGraw-Hill.

Editorial history:
paper received 17 April 2002;
final version received 5 September 2002;
accepted 22 October 2002.