*ICIC Express Letters*
*Part B: Applications*
Volume **10**, Number **2**, February **2019**

ICIC International ©2019 ISSN 2185-2766
pp. 113–120

# QUALITY OF CHINA'S PRE-SCHOOL RESOURCE MANAGEMENT EXPLAINED BY CLUSTER ANALYSIS

Dian-Fu Chang[1] and Wenshan Zhu[2,*]

[1]Graduate Institute of Educational Policy and Leadership
[2]Doctoral Program of Educational Leadership and Technology Management
Tamkang University
No. 151, Yingzhuan Road, Tamsui District, New Taipei City 25137, Taiwan
140626@mail.tku.edu.tw; *Corresponding author: 806764014@s06.tku.edu.tw

Received July 2018; accepted October 2018

Abstract. *This study explores the quality issues of China's pre-school resource management. The effective allocation of educational resources has become an important strategy for promoting the quality and ameliorating the equity of pre-school education. This study defines the resources in pre-school level including human resources, physical resources, and financial resources of 31 provinces in China. Specifically, there are four indicators in human resources, four indicators in physical resources, and three indicators in financial resources. Cluster analysis was used to classify the resources in current pre-school system. According to Minitab, the key outputs include the similarity and distance values, the dendrogram, and the final partition. The priority areas have been listed based on their disadvantaged resource allocation in pre-school education. The findings suggest cluster analysis can be used to detect the effect of resource management by selected indicators. Finally, the related suggestions for policy makers are made.*

**Keywords:** China education, Cluster analysis, Pre-school education, Resource management, Quality of education

1. **Introduction.** Quality resource management has been considered as the priority strategy in different management systems. While different settings provide various consequences when confronted with the limited resources allocation. In China, the pre-school education has become an emerging movement due to the new economic development. Furthermore, most of parents have realized that getting start earlier learning can lead to success in next levels of education which has driven the numbers increasing. According to China's "Planning Guideline for the Middle and Long term Development 2010-2020", the government has initiated that target enrollment in pre-school education should reach 70% gross entrance ratio (GER) in 2020. In 2016, there are 44 million and 139 thousand kids enrolled in 240 thousands kindergartens. It implies 77.4% GER counted in 3 years pre-school education [1].

Even though the demand has shown increasing, the quality of pre-school education did not fit the parents' basic needs. Disparity of resources allocation is a crucial influencing factor in pre-school education level. This study considered the resource management issue by using cluster analysis to detect which provinces should be listed in the priority areas to reallocate resources. Previous studies pointed out the disparity of human resources in kindergartens, for example, rural areas are confronted with teacher shortage [2]. The physical resources which include numbers of kindergartens, space of kindergarten, facilities, related equipment are also faced to a wide gap between rural and urban areas [3]. Song et al. indicated there is no any guarantee of the financial investment for kindergartens in rural areas [4]. Moreover, the government's policy is an important impact factor to promote the quality of kindergartens. Therefore, how to balance between marketization and

DOI: 10.24507/icicelb.10.02.113

government intervention has become an emergent issue for effective investing resources in pre-school education [4]. From managerial viewpoint, previous study raised opening market or providing multiple investment approaches for kindergartens may cause the resource allocation differently [5]. With regard to the resource issues, previous studies do not fully take the province differences into consideration. This is why this study takes this point to explore the issue.

This study selected China as a target to explore the quality issues of pre-school resource management based on province basis data. In addition, the province-based data collection is not so difficult in China. The findings are easy to implement in current province government level. Therefore, the purposes of this study are as follows: a) to realize the pre-school resource management issues in the system; b) to identify the areas that need to reinforce their specific resources. Given these purposes, the structure of this paper is as follows. The first section described the development of pre-school education and the issue of disparity of pre-school resources allocation in China. The method section provides a brief description of the cluster analysis to classify the resources in current pre-school system. Then, this study conducts Minitab to verify the classification to provide suggestions for policy makers. Finally, the conclusions present the implications of this study.

## 2. Method.

**2.1. Research framework.** The main resource variables are human resources (H), physical resources (P), and financial resources (F). The definitions of the related resources are displayed as follows:

Human resources (H) refer to capacities of pre-school (kindergartens), staff, ratio of kids by teachers, and ratio of kids by caretakers;

Physical resources (P) refer to kindergartens, classes, class size, and space;

Financial resources (F) refer to total investment, educational expenditure (EE) by kids, EE by kindergartens.

Total resources are equal to H+P+F. The framework of research is presented in Figure 1.

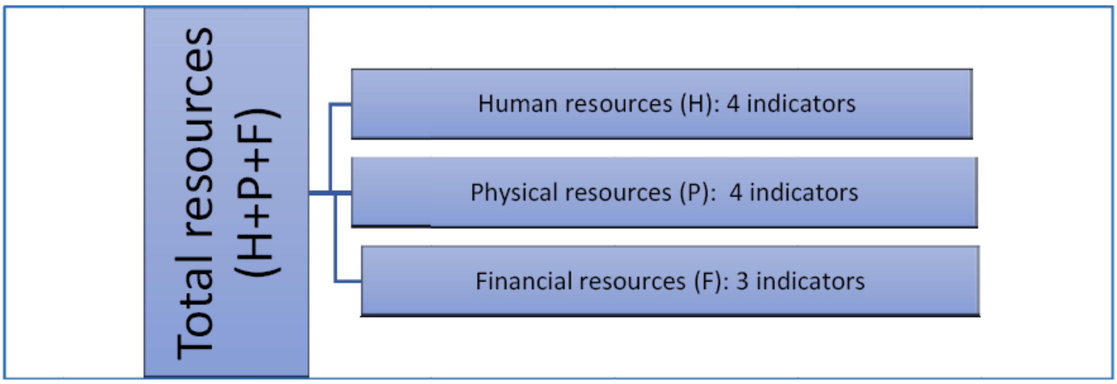


Figure 1. The framework of research

**2.2. Data collection.** The main data set comes from Educational Statistics in 2015, Department of Development Planning, Ministry of Education of the People's Republic of China. Part of data are selected from China Educational Finance Statistical Yearbook, 2016. The data have been converted as the indicator formats by specific 31 provinces (including autonomous regions and municipalities) in China to fit the cluster analysis. The original data are defined as follows.

a) Staff refers to caretakers and teachers mean full-time teachers and caretakers. The data were collected based on province basis which will be transformed to 4 indicators and applied in human resource domain.

b) Per-child floor area means floor area divided by the amount of kindergarten children; per-child refers to numbers of kindergartens, classes, class size. The related indicators will be classified in physical resource domain.

c) Per-child average educational expenditure means kindergarten education investment divided by amount of in-kindergarten children. This is one of indicators in financial resource domain.

d) Per-kindergarten average educational expenditure means kindergarten education investment divided by the amount of kindergartens. This indicator was used in financial resource domain.

2.3. **Cluster analysis.** Basically, cluster evaluation determines the optimal number of clusters for the data using different evaluation criteria in diverse settings. In this study, the data selection based on China's provinces is the first step; then this study transformed the data into related indicators in terms of human, physical, and financial resources for analyzing. In the cluster analysis section, this study applied cluster observations in Minitab to determining the fittest clusters with current data of indicators. We put the indicators according to the different domains. The cluster algorithms are as follows [6-9]:

1) Select $k$ point as initial centroids,
2) Repeat,
3) From $k$ clusters assign each point to its closest centroids,
4) Re-compute the centroids of each cluster,
5) Until centroids do not change.

According to Minitab, the key outputs include the similarity and distance values, the dendrogram, and the final partition. In this study, the cluster analysis process is displayed as follows.

Step 1: Examine the similarity and distance levels. At each step in the amalgamation process, view the clusters that are formed and examine their similarity and distance levels. The higher the similarity level is, the more similar the observations are in each cluster. The lower the distance level is, the closer the observations are in each cluster. Ideally, the clusters should have a relatively high similarity level and a relatively low distance level.

Step 2: Get key results. Check similarity level and distance level. It is important to balance that goal with having a reasonable and practical number of clusters. At each subsequent step, as new clusters are formed, the similarity level decreases and the distance level increases. At the final step, all the observations are joined into a single cluster.

Step 3: Examine the final partition. After determining the final groupings in step 2, this study reruns the analysis and specifies the number of clusters (or the similarity level) for the final partition. Minitab displays the final partition table, which shows the characteristics of each cluster in the final partition. Based on the final partition, the average distance from the centroid provides a measure of the variability of the observations within each cluster. In this step, it may need to examine the clusters in the final partition to determine whether the grouping seems logical for the application. If it is still unsure, the repeated analysis is necessary to decide which final grouping is the most logical for the data.

In this study, the Euclidean distance with Ward linkage is the fittest model for cluster analysis with these data set. The Euclidean distance, between points $\mathbf{p}$ and $\mathbf{q}$, is the length of the line segment connecting them. If $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ and $\mathbf{q} = (q_1, q_2, \ldots, q_n)$ are two points in Euclidean $n$-space, then the distance ($d$) from $\mathbf{p}$ to $\mathbf{q}$, or from $\mathbf{q}$ to $\mathbf{p}$ is given as [10]:

$$d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2} = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}.$$

We carried on like this into 4 or more dimensions, in general $n$ dimensions, where $n$ is the number of variables.

In this study, the first step of the Ward method is the normalization. It is applied usually because of the possible scale differences among the variables; thus, the data should be normalized. In this study, the Ward linkage conducts as follows:

$$d_{A,B} = n_A\|\bar{x}_A - \bar{\bar{x}}\|^2 + n_B\|\bar{x}_B - \bar{\bar{x}}\|^2$$

By applying Ward method the aim was to join elements into clusters so that the variance within clusters is minimized [11].

## 3. **Results.**

The results have been displayed by visualized and statistical formats. In this study, the analyses include province differences in human resources, physical resources, financial resources, and total resources. The related dendrograms and their final partition will be demonstrated in the following sections.

### 3.1. **Verification of human resources.**

Based on the cluster analysis of observations with capacity, staff, ratio of kids by teachers, and ratio of kids by caretakers, the result reveals that Euclidean distance and Ward linkage can be classified the human resources data of 31 provinces into two significant groups. The details of dendrogram and final partition are presented in Figure 2 and Table 1.

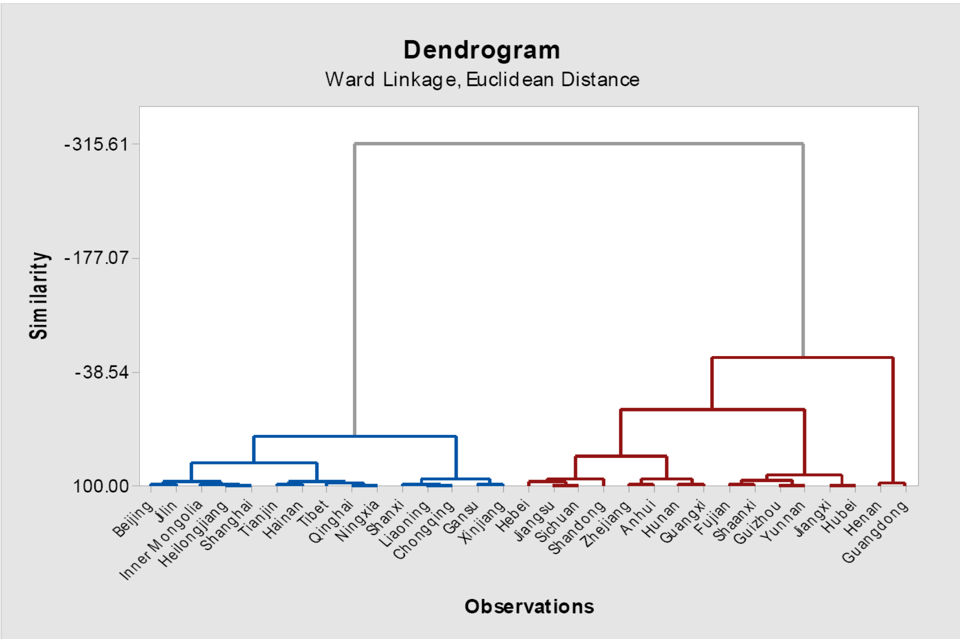


Figure 2. Dendrogram of human resources in 31 provinces

Table 1. Final partition of human resources in cluster analysis

| Cluster analysis | Number of observations | Within cluster sum of squares | Average distance from centroid | Maximum distance from centroid |
|---|---|---|---|---|
| Cluster1 | 15 | 1.19253E+12 | 239037 | 458560 |
| Cluster2 | 16 | 1.04726E+13 | 621743 | 1868208 |
| Indicators | | Cluster1 | Cluster2 | Grand centroid |
| Capacity | | 525190 | 2173152 | 1375751 |
| Staff | | 47637 | 173828 | 112767 |
| Ratio of kids by teachers | | 20 | 23 | 21 |
| Ratio of kids by caretakers | | 99 | 75 | 87 |

This dendrogram was created using a final partition of 2 clusters, which occurs at a similarity level of approximately −40. The first cluster (far left) is composed of 15 observations (Beijing, Jilin, Inner Mongolia, Heilongjiang, Shanghai, Tianjin, Hainan, Tibet, Qinghai, Ningxia, Shanxi, Liaoning, Chongqing, Gansu, Xinjiang). The second cluster, directly to the right, is composed of 16 observations (Hebei, Jiangsu, Sichuan, Shandong, Zhejiang, Anhui, Hunan, Guangxi, Fujian, Shaanxi, Guizhou, Yunnan, Jiangxi, Hubei, Henan, Guangdong). In this case, the provinces in the first cluster are with small capacity and better ratio of kids by teachers, while the provinces in the second cluster are with larger capacity and better ratio of kids by caretakers.

3.2. **Verification of physical resources.** Based on the cluster analysis of observations with kindergartens, classes, class size and space, the result shows that Euclidean distance and Ward linkage can be classified the human resources data of 31 provinces into three significant groups. The details of dendrogram and final partition are presented in Figure 3 and Table 2.

This dendrogram was created using a final partition of 3 clusters, which occurs at a similarity level of approximately −100. The first cluster (far left) is composed of 14 observations (Beijing, Shanghai, Chongqing, Gansu, Qinghai, Ningxia, Tianjin, Hainan, Tibet, Inner Mongolia, Shanxi, Liaoning, Jilin, Heilongjiang). The second cluster, in

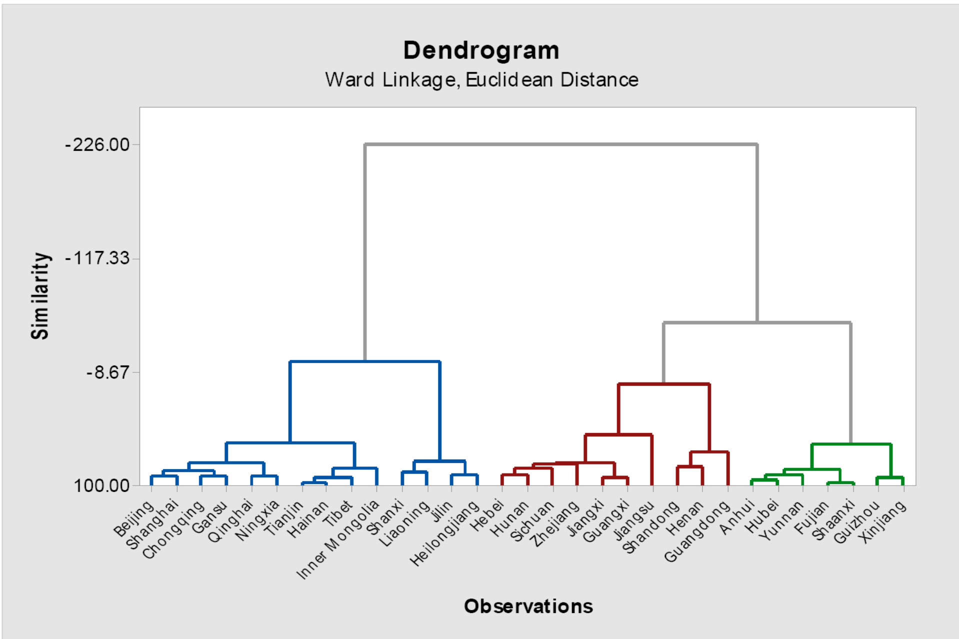


Figure 3. Dendrogram of physical resources in 31 provinces

Table 2. Final partition of physical resources in 31 provinces

| Cluster analysis | Number of observations | Within cluster sum of squares | Average distance from centroid | Maximum distance from centroid |
|---|---|---|---|---|
| Cluster1 | 14 | 18.8476 | 1.05097 | 1.93547 |
| Cluster2 | 10 | 21.8220 | 1.37135 | 2.55692 |
| Cluster3 | 7 | 3.2391 | 0.62796 | 1.07748 |
| Indicators | Cluster1 | Cluster2 | Cluster3 | Grand centroid |
| Kindergartens | −0.752420 | 1.14797 | −0.135118 | −0.0000000 |
| Classes | −0.794697 | 1.17021 | −0.082328 | −0.0000000 |
| Class size | −0.629140 | 0.23633 | 0.920672 | −0.0000000 |
| Space | −0.746162 | 1.11275 | −0.097315 | 0.0000000 |

the middle, is composed of ten observations (Hebei, Hunan, Sichuan, Zhejiang, Jiangxi, Guangxi, Jiangsu, Shandong, Henan, Guangdong). The third cluster, directly to the right, is composed of seven observations (Anhui, Hubei, Yunnan, Fujian, Shaanxi, Guizhou, Xinjiang). Based on the result in Table 2, the 10 provinces in cluster 2 should be listed as priority areas to improve their physical resources.

3.3. **Verification of financial resources.** Figure 4 and Table 3 demonstrate the cluster analysis of observations with total investment, EE by kids, and EE by kindergartens. The result displays that Euclidean distance and Ward linkage can be classified the financial resources data of 31 provinces into two significant groups.

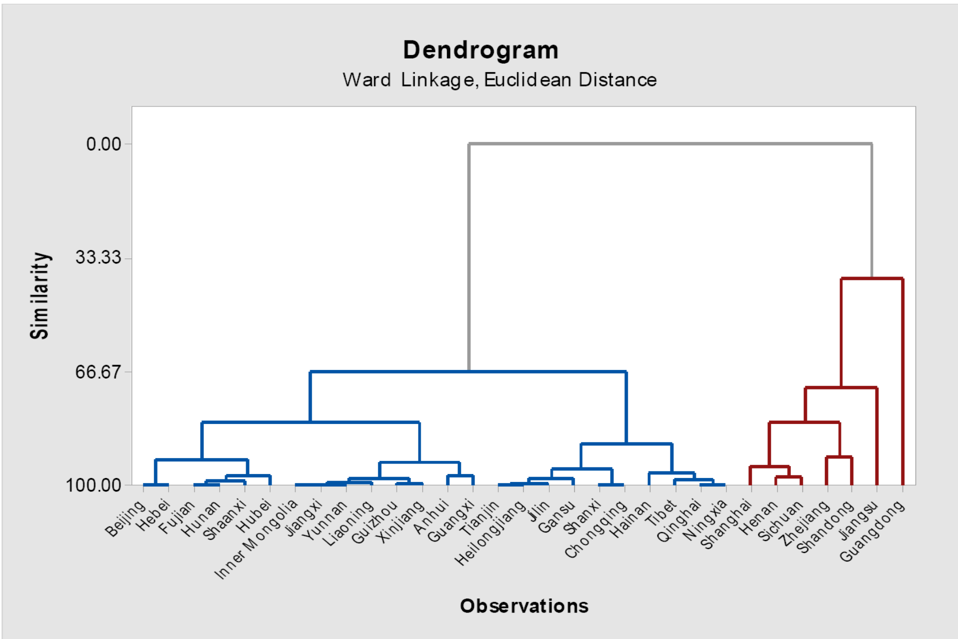


FIGURE 4. Dendrogram of financial resources in 31 provinces

TABLE 3. Final partition of financial resources in 31 provinces

| Cluster analysis | Number of observations | Within cluster sum of squares | Average distance from centroid | Maximum distance from centroid |
|---|---|---|---|---|
| Cluster1 | 24 | 1.55281E+14 | 2052982 | 4589897 |
| Cluster2 | 7 | 1.85342E+14 | 4011240 | 11162123 |
| Indicators | | Cluster1 | Cluster2 | Grand centroid |
| Total investment | | 5362373 | 16282498 | 7828208 |
| EE by kids | | 7 | 8 | 7 |
| EE by kindergartens | | 1249 | 2355 | 1498 |

Note. EE refers to educational expenditure.

This dendrogram was created using a final partition of 2 clusters, which occurs at a similarity level of approximately 33. The first cluster (far left) is composed of 24 observations (Beijing, Hebei, Fujian, Hunan, Shaanxi, Hubei, Inner Mongolia, Jiangxi, Yunnan, Liaoning, Guizhou, Xinjiang, Anhui, Guangxi, Tianjin, Heilongjiang, Jilin, Gansu, Shanxi, Chongqing, Hainan, Tibet, Qinghai, Ningxia). The second cluster, directly to the right, is composed of seven observations (Shanghai, Henan, Sichuan, Zhejiang, Shandong, Jiangsu, Guangdong). Based on educational expenditure, the provinces in cluster 1 should be considered as the priority areas to invest more financial resources.

3.4. **Types of total resource management.** Figure 5 and Table 4 demonstrate the cluster analysis of observations with total resources (H+P+F). The result displays that Euclidean distance and Ward linkage can be classified the total resources data of 31 provinces into two significant groups.

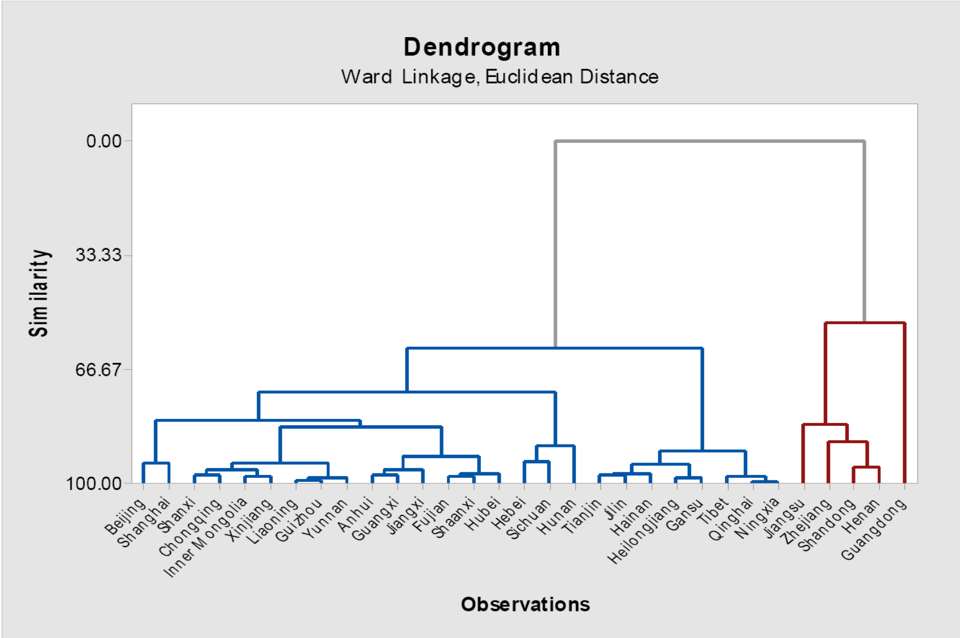


Figure 5. Dendrogram of total resources in 31 provinces

Table 4. Final partition of total resources in 31 provinces

| Cluster analysis | Number of observations | Within cluster sum of squares | Average distance from centroid | Maximum distance from centroid |
|---|---|---|---|---|
| Cluster1 | 26 | 5.47687E+14 | 3971192 | 8504728 |
| Cluster2 | 5 | 2.43656E+14 | 5808344 | 12891677 |
| Indicators | | Cluster1 | Cluster2 | Grand centroid |
| Capacity | | 1060194 | 3016647 | 1375751 |
| Staff | | 81763 | 273992 | 112767 |
| Caretakers and teachers | | 62195 | 213357 | 86576 |
| Ratio of kids by teachers | | 22 | 19 | 21 |
| Ratio of kids by caretakers | | 17 | 14 | 17 |
| Class size | | 28 | 30 | 29 |
| Space | | 5993760 | 20519719 | 8336657 |
| Space by kids | | 6 | 7 | 6 |
| Total investment | | 5868987 | 18016158 | 7828208 |
| EE by kids | | 8 | 6 | 7 |
| EE by kindergartens | | 1485 | 1568 | 1498 |

This dendrogram was created using a final partition of 2 clusters, which occurs at a similarity level of approximately 30. The first cluster (far left) is composed of 26 observations (Beijing, shanghai, Shanxi, Chongqing, Inner Mongolia, Xinjiang, Liaoning, Guizhou, Yunnan, Anhui, Guangxi, Jiangxi, Fujian, Shaanxi, Hubei, Hebei, Sichuan, Hunan, Tianjin, Jilin, Hainan, Heilongjiang, Gansu, Tibet, Qinghai, Ningxia). The second cluster, directly to the right, is composed of five observations (Jiangsu, Zhejiang, Shandong, Henan, Guangdong). Considering all the indicators in the cluster model, two clusters have been identified, while it is not easy to justify the appropriate resourcing strategies for the specific provinces.

4. **Conclusions.** This study has demonstrated current pre-school resource allocation in China by using cluster analysis. The results have presented the disparity existing among provinces in terms of human resource, physical resource and financial resource, which can provide useful information for policy makers to set a specific innovative plan.

Cluster analysis provides a practical solution with graphic format which is easy to detect the disadvantaged provinces or areas. However, the selecting variables for cluster analysis belong to a specific professional judgement. For further implication, it also depends on the policy purposes in different settings. This study provides an example to tackle the resource management issues in kindergartens based on different provinces. For future studies, the related designs can be modified to fit similar settings.

**REFERENCES**

[1] Ministry of Education, *National Education Development Statistical Communique of 2016*, http://www.moe.gov.cn/jyb_sjzl/sjzl_fztjgb/201707/t20170710_309042.html, 2017.
[2] X. Hong and L. Luo, Analysis of the differences in pre-school education development in urban and rural China from the perspective of education equity, *Journal of Educational Studies*, vol.8, no.5, pp.73-81, 2012.
[3] Y. Zhou and R. Yang, An analysis of the system factors of influencing preschool education equality and balanced development between urban and rural areas-taking Guangdong province as a case study, *Studies in Preschool Education*, no.5, pp.12-19, 2010.
[4] Z. Song, T. Zhang and Q. Ding, Current situation and rethinking on the sudy of preschool education resource allocation in China, *Journal of Shaanxi Xueqian Normal University*, vol.33, no.7, pp.133-137, 2017.
[5] Y. Xu and S. Chen, A study on the fairness of preschool education from the aspect of public resources constitution, *Early Childhood Education (Educational Sciences)*, no.4, pp.22-24, 2007.
[6] O. R. Battaglia, B. D. Paola and C. Fazio, A new approach to investigate students' behavior by using cluster analysis as an unsupervised methodology in the field of education, *Applied Mathematics*, vol.7, no.15, pp.1649-1673, 2016.
[7] M. J. Brusco, R. Singh, J. D. Cradit and D. Steinley, Cluster analysis in empirical OM research: Survey and recommendations, *International Journal of Operations & Production Management*, vol.37, no.3, pp.300-320, 2017.
[8] M. L. Crowe, A. C. LoPilato, W. K. Campbell and J. D. Miller, Identifying two groups of entitled individuals: Cluster analysis reveals emotional stability and self-esteem distinction, *Journal of Personality Disorders*, vol.30, no.6, pp.1-14, 2015.
[9] N. K. Lankton, D. H. McKnight and J. F. Tripp, Facebook privacy management strategies: A cluster analysis of user privacy behaviors, *Computers in Human Behavior*, vol.76, pp.149-163, 2017.
[10] Euclidean distance, *Wikipedia, the Free Encyclopedia*, https://en.wikipedia.org/w/index.php?title=Euclidean_distance&oldid=855266282.
[11] B. S. Everitt, S. Landau, M. Leese and D. Stahl, *Cluster Analysis*, 5th Edition, Wiley, 2011.