A New Analytical Approach for Continuous Improvement Evaluation: A National Presentation Case

Chyong Ling Judy Chen

Feng Chia University, Taichung, Taiwan

Abstract—**This paper demonstrates a new approach for analyzing evaluation data to determine a team's (sometimes called circle) improvement ability and the characteristics of evaluators by using the data from the national Quality Control Circle (QCC) presentations in Taiwan. Most evaluations of the QCC presentation use the itemized rating method or the duplex-pole evaluation method. The former has the negative effects of canceling out the higher and lower while the latter cannot distinguish the ability type of a team. This study takes the average of the highest rating "A" as an index for superiority. Based on these indices and the K-mean method, similar ability groups and evaluator groups are generated, respectively.**

According to a team's improvement ability level, which is determined by the similar group to which it belongs, a proper regenerated input can be applied to sustain continuous improvement activity. In addition, knowing the characteristics of evaluators, proper solutions can be applied to improve the evaluation credibility. This paper also shows that some of the evaluation data can provide valuable information through further study, thus increasing "evaluation use".

Index Terms—**Ability type, Continuous improvement, Evaluation analysis, Evaluator characteristic.**

I. INTRODUCTION

In the era of globalization and fast changing economies, Continuous Improvement (CI) is thought to play an important role in maintaining a company's competitiveness. The ultimate goal of CI is to create an environment conducive to learning and growth through company-wide involvement in gradual improvement in process performance and innovation. CI generally takes into account the activities performed under the names of Statistical Quality Control (SQC), Quality Control Circle (QCC), Quality Improvement Team (QIT), Project Management (PM), 6 Sigma, etc.

A. The Problem with the Continuity of CI Activity

A company's overall capability can be improved through persistent CI activity. A CI activity will be sustained provided it produces good results. However, maintaining the original performance is difficult, not to mention creating further effective results. Imai [11] described the problem as, "All systems are destined to deteriorate once they have been established unless continuing efforts are made to maintain it and then to improve on it." In other words, performance may diminish even upon repetition of the original activity.

In general, there are three key components to a CI program: theme/problem, model and tool, and promotion [20]. These construct the three bases of a pyramid structure. With solid bases and strong linkages among them, a company's potential capability in improvement and innovation can be reached. Unfortunately, most companies focus on promotion as a way to boost the employees' involvement or improve some factors such as training, communication, or management involvement and commitment [5]. Nevertheless, increasing the participation rate or a few factors in the absence of solid structure or overall capability cannot guarantee good results. [19] analyzed the performance of the QCC program run by Uni-President Enterprises Corporation[1]. They found that the correlation coefficient between participation rates and monetary benefits was very low. This indicates that an increase in the participation rate cannot assure an increase in financial benefit. On the other hand, this could also imply that the participation rate in the company has reached a level which cannot be improved anymore. In this case, the financial benefit can be generated only if the company's capability is upgraded.

The other possible reason why a CI activity cannot be sustained is because there is no complete roadmap for continuity to move forward. A company may not have a clear idea about what its next step should be and how to get there.

B. Possible Solution for CI Continuity

Senge [15] proposed the concept of learning organization. However, a learning organization cannot be established in one day. By observing many companies' behaviors, [2] summarized five evolutionary CI levels: from no improvement activity (Level 0) to trying out the idea (Level 1), structural and systematic CI (Level 2), strategic CI (Level 3), autonomous CI (Level 4), and ultimately becoming a learning organization (Level 5).

Although the evolutionary CI model can be treated as a roadmap for companies to move toward a learning organization, it is not easy for management to implement it so as to upgrade a firm's CI level. [20] proposed an integrated structural model to enhance the behavioral model of [2]. Corresponding to the five evolutionary CI levels, the integrated structural model constructed the respective pyramid structures for each level [4]. A

---

[1] Uni-President Enterprises Corporation is the oldest and largest company in the food industry in Taiwan. They have run QCC programs for many years and have generated good results.

company can eventually become a learning organization by building up the structure stage by stage. Each stage requires certain improvement ability.

Based on the progress of the associated improvement cases, [20] classified the improvement ability into 4 levels. They are Level 1, case cannot be completed; Level 2, case is completed but not selected for presentation; Level 3, case is selected but is classified as a general class that is possessing basic improvement ability; Level 4, case is selected and classified as a process class that is possessing the ability to solve the root cause or to innovate [20]. For example, moving from Level 3 to Level 4, a company requires the basic ability of solving problems, while from Level 4 to Level 5, a company should possess the process and innovation ability. Only if equipped with certain ability can a company move up to a higher improvement level.

C. The Objectives of This Study

For a company to upgrade its capability, it needs to know where it stands, that is, what ability it currently possesses. This information can be found from the evaluation results of its improvement cases. In general, an improvement case in a CI activity will go through several evaluations including self-assessment, field evaluation (site-examination), and presentation evaluation. It may also be selected for a national presentation.

One of the objectives of this paper is to introduce an analytical method which, by analyzing the national evaluation data, is able to find a participating team's improvement ability. Based on the results, a company can apply a workable regenerating input to its system, so that its improvement activity can be sustained.

Undoubtedly, the characteristics and qualifications of evaluators affect the fairness of the evaluation results.

Thus, using the same data and analytical method, this paper also analyzes the evaluators' characteristics. From the results, the shortcomings of the evaluation system and the weaknesses of the evaluators may be found. Solutions can be applied to make the evaluation more fair and trustworthy.

Moreover, one of the core subjects on evaluation in the past 10 years is the evaluation use [7][9][13][17]. The evaluation data used in this study are somewhat useless to the sponsor after the presentation activity is over. This study, by chance, demonstrates that some of the evaluation materials can provide valuable information through further study, which increases the use of the evaluation.

II. EVALUATION OF THE NATIONAL PRESENTATION IN TAIWAN

The development of CI in Taiwan has a long history. QCC, SQC, CWQC (Companywide Quality Control), etc. are the core of CI. For example, the reason why Phillips (Taiwan) won the Japanese Deming award in 1991 was because it achieved an excellent performance through SQC activities. The Association of Pioneer Quality Control Research ("the Association") has sponsored the QCC national presentation in Taiwan for many years. The rank of teams is determined based on the evaluation results. Once finalized, the evaluation data have met their goal.

The evaluation form includes 16 items (Table I). These are developed from each step, or category, in the QC-STORY model. Every step contains 2 to 4 items. There are 5 grades for each item, A, B, C, D, and E. The meaning of each grade and the score with which it is associated are shown in Table II.

TABLE I.
EVALUATION TABLE

Step	No.	Item Description	A	B	C	D	E
Theme selection	1	Is the selected theme proper?					
	2	Does the theme properly indicate the problem?					
Analytical process	3	Is the cause & effect analysis sufficient and precise?					
	4	Are the necessary data sufficient?					
	5	Are the QC techniques properly applied throughout the analytical processes?					
	6	Are the root causes of the problem fully captured?					
Countermeasure	7	Is the countermeasure based on the analytical results?					
	8	Is the countermeasure executable?					
	9	Is the countermeasure creative and does it fully exploit brainstorming?					
	10	Is the effort sufficient enough to execute the countermeasure?					
Confirmation of results	11	Has every result of countermeasure been confirmed?					
	12	Have the results of countermeasure before and after execution been compared?					
Maintenance of results	13	Has the standard derived from the countermeasure been established?					
	14	Has the effectiveness of improvement result been fully maintained?					
Plan	15	Has the plan for next activity been established?					
Presentation	16	Are presentation, graphs, and Q&A appropriate?					

TABLE II.
MEANING AND SCORE OF GRADES

Grade	Meaning	Score
A	Indicates that the item is especially valuable	9
B	Indicates that the item is valuable or better	7
C	Indicates that the item is fair	5
D	Indicates that the item is incomplete or lagging behind	3
E	Indicates that the item is especially poor	1

In the first five years, the Association adopted an itemized rating method, that is, it decided the ranks of each team based on their total scores which were generated by summing up the individual score of each item. This system caused two complications: 1) the participant whose score is slightly behind others doubted the evaluation results; 2) the results can hardly be objective because the individual evaluator has his/her subjective view. For example, the evaluation results for one event from a technician and from a manager can be different. The difference may cancel each other out and, thus, the strength of a team is not able to be recognized. Also, according to [10], linked sets of grades from scales not belonging to the same categories may bring about errors of rank and absurd averaging.

To cope with this problem, the Association then adopted the duplex-pole evaluation method [16]. First, the standard of different ranks was set up. This standard, instead of scores, only considers the percentage of each level of grade that a team obtained from evaluators (Table III). The concept for the duplex-pole method is that the teams that have been given many "D"s and "E"s will not be included in the excellent or outstanding group, and the teams that have been given adequate "A"s and "B"s will be included.

The procedures of duplex-pole evaluation are as follows: During the presentation, evaluators have to make judgments as to whether a certain item is outstanding. If so, a mark, such as "V", is made in the corresponding column A in the evaluation form. If an evaluator feesl that a certain item is valuable or better than other member's presentation, he/she marks "V" in the corresponding column B. If an evaluator feels that a certain item in the presentation is incomplete or relatively poor, he/she marks "V" in the corresponding column D. And if an evaluator feels that a certain item of the presentation is relatively inferior, he/she marks "V" in the corresponding column E. Blank columns are considered fair and a "V" mark will be made in column C. Summing up the number of each level of grades, and comparing the results with the standard set by the Association, the final rank will be determined for each team.

TABLE III.
STANDARD OF RANKS [16]

Rank	Basis for ranking
Excellent Team	1) A $\geq$ 25%K A+B $\geq$ 80%K D+E $\leq$ 10%K; or2) A $\geq$ 50%K D+E $\leq$ 10%K
Outstanding Team	1) A+B $\geq$ 60%K D+E $\leq$ 26%K; or 2) A $\geq$ 20%K A+B+C $\geq$ 70%K
Good Team	1) A+B+C $\geq$ 60%K
Industrious Team	1) A+B+C $\geq$ 10%K

Note: K indicates the number of evaluating items. In this paper, it is 16. This standard may be adjusted slightly in different presentations.

A. The Superiority Index: the Average of "A"s

Basically, the participating teams are good teams since they had undergone a selection process. However, people want to know in which fields the team excel. If the presentation results can only tell their scores or ranks, ability types will be unclear. For example, if one team obtains 5 "A"s and 5 "C"s, while the other obtains 10 "B"s, they may receive the same total score or same rank. Yet, we know the first team has 5 outstanding items and the latter has none.

According to the Association's concept, that is, giving an "A" to the team that impressed the evaluator the most, "A" indicates that a team is very outstanding in this item. Since one of the goals of this paper is to find a team's improvement level and ability type from the evaluation outcomes, this paper proposes a new method which takes the average of "A" as an index of superiority for an item for a team. This average is calculated by dividing the number of "A"s of a item by the number of evaluators who evaluated this team. For example, if 7 out of 10 evaluators give an "A" to a team for Item #1, then this team receives 0.7 as its superiority index on Item #1. These indices will be used for cluster analysis to generate similar groups (or clusters).

B. Cluster Analyses

Euclidian nearest centroid distance method or K-mean method was used for clustering. This is the most popular clustering tool used in scientific and industrial applications [1]. The main advantage of this method is that it is easily understood and implemented, and the foundation of analysis of variances is firm. Possible shortcomings are that there is no concrete method to choose the number of clusters (K), the computed local optimum may be quite different from the global one, and the process is sensitive to the outliers [1][14].

The procedures of K-mean method are briefly described as follows [3]:

(1) Arbitrarily divide data into K groups. In this study, K=10 is predetermined. Calculate the centroid or weighted average for each group.

(2) Calculate the distance of each data point to the 10 centroids. Relocate each data point to the group whose centroid is the closest to this data point.

For example, in a two-dimensional space, there are two groups with their respective centroids, $A(a_1, a_2)$ and $B(b_1, b_2)$. For a data point $C(x, y)$, its distances to A and B are calculated as

$$d_1 = (x - a_1)^2 + (y - a_2)^2$$
$$d_2 = (x - b_1)^2 + (y - b_2)^2.$$

If $d_1 < d_2$, then data point $C(x,y)$ will be classified into group A; if $d_1 > d_2$, then data point $C(x,y)$ will be classified into group B.

(3) Recalculate the centroid of each group. (The Centroid Method in SPSS was used to calculate centroids in this study)

(4) Repeat steps (2) and (3) until there is no more necessity to relocate the data point.

To ensure that the classification is proper, the variances of data within each group were calculated and the F-value of the variances between 2 groups was derived. All F-values are significant. This confirms that the classification is appropriate. Finally, radar charts were drawn based on the cluster mean in each group. Radar charts provide information about the ability type of each cluster.

III. THE EMPIRICAL STUDY

A. Data

A total of 2306 data sets were used in this study[2]. They were taken from the original evaluation forms in the national presentation sponsored by the Association of Pioneer Quality Control Research from 1985 to 1987. There was much interest and involvement in QCC activities and presentations during these three years in Taiwan. There were 23 presentation sessions in these three years, with 190 teams participating. Those teams were selected as outstanding and were from different companies. There were 6-14 evaluators evaluating each team. A total of 249 evaluators were involved. These evaluators included QC experts and CI activity promoters.

B. Reliability Analysis

First, the internal reliability analysis was applied to each team. The results (Table IV) show that the Cronbach's α of 163 teams (86%) is higher than 0.7; and that only 9 teams (4.74%) had Cronbach's α below 0.35. Among them, 8 were from the same session. According to [8], data are unreliable if the Cronbach's α is below 0.35 while [18] suggested 0.3 as a critical number. Thus, our data in terms of internal consistency should be acceptable.

C. Team Improvement Ability

From the data, the superiority index of every item in each team was calculated. Ten clusters were identified by the K-mean method. The cluster mean of each item was

---

[2] Data were provided by one of the researchers in the Association.

listed in Table V. Radar charts were drawn for each cluster using the cluster means (Table VI). Radar charts provide an easier way to identify the characteristics and the ability type of each group. From the radar charts, 10 clusters were summarized into 3 classes: the general class (improvement ability Level 3), the process class (improvement ability Level 4), and the extraordinary class. The details of the 3 classes are described as follows:

TABLE IV.
DISTRIBUTION OF CRONBACH'S α

Cronbach's α	No. of team	percentage
α< 0.3	8	4.21%
α< 0.35	9	4.74%
0.3 α< 0.4	5	2.63%
0.4 α< 0.5	5	2.63%
0.5 α< 0.6	9	4.74%
0.6 α< 0.7	22	11.58%
0.7 α	141	74.21%

- General class (85% of total=161 teams): In this class, a group generally does not have outstanding performance in most of the items but could possibly be good in one particular area. Both the superiority index and total scores in this class are relatively low. Among this class, three features are observed. They are the common type (Cluster #1), which shows no excellence in any item (59%=112 teams); the presentation type (Cluster #2), which is outstanding in presentation techniques (17%=33 teams); and the technology-excellence type (Cluster #3), which is good in countermeasure techniques (8%=16 teams). None in the common type were ranked in the excellent group in the national presentation by the Association, while 11 teams of the presentation type (33%) and 6 teams of the technology-excellence type (37%) were ranked as excellent.

- Process class (12% of total=22 teams): In this class, a group has an above-average performance in each item and is excellent in a few. The superiority index and total scores in this class are relatively high. Among this class, three different features have also been found. They are the presentation type (Cluster #4) (3%=6 teams), the SQC type (Cluster #5), which is considered good in the application of SQC techniques (5%=10 teams), and the confirmation type (Cluster #6), which is considered outstanding in identifying and maintaining the effectiveness of the performance (3%=6 teams). Five teams were of the presentation type (83%), 8 teams were of the SQC type (80%), and 6 teams were of the confirmation type (100%), and all were ranked as excellent in the national presentation.

- Extraordinary class (4% of total=7 teams): In this class, one particular item is outstanding, while others are not as good. (There are 4 types in this class, Clusters #7 to #10.) This class cannot become the prototype for learning. Only 1 team (14%) was ranked as excellent in the national presentation.

From the percentage rate of excellent teams, it seems that the results from this study are consistent with those of the duplex-pole method adopted by the Association. That is, many teams in the higher ability level, such as those in

the process class, were ranked as excellent teams in the national presentation.

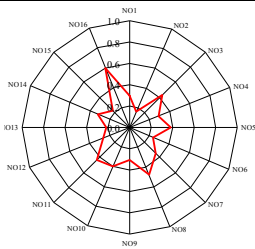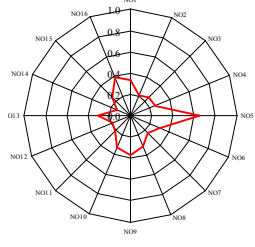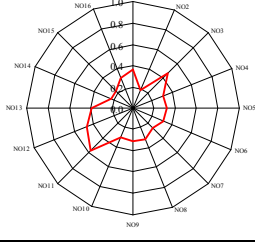A team of a company can find its ability feature from the group to which it belongs. The government may also obtain a picture of the ability of companies within the nation

TABLE V.
CLUSTER MEANS OF SUPERIORITY INDEX

Cluster \ Item	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
1	0.21	0.34	0.24	0.30	0.33	0.36	0	0.37	0.33	0.17
2	0.07	0.12	0.09	0.16	0.20	0.19	0	0	0.44	0.17
3	0.11	0.18	0.21	0.44	0.24	0.47	0.17	0.49	0.89	0.33
4	0.09	0.16	0.12	0.29	0.25	0.31	0.33	0.18	0.33	0.33
5	0.11	0.16	0.22	0.38	0.65	0.32	0.17	0.05	0.33	0.33
6	0.09	0.20	0.27	0.24	0.30	0.31	0.50	0.07	0.33	0
7	0.13	0.22	0.31	0.35	0.23	0.26	0.67	0.41	0.22	0.50
8	0.15	0.27	0.39	0.48	0.30	0.32	0.83	0.25	0	0
9	0.1	0.17	0.46	0.31	0.37	0.31	0.17	0.30	0.33	0
10	0.14	0.24	0.38	0.40	0.33	0.30	0.17	0.14	0.33	0.50
11	0.12	0.22	0.23	0.43	0.20	0.56	0	0.11	0.11	0
12	0.10	0.15	0.15	0.26	0.18	0.47	0	0.14	0.11	0
13	0.12	0.21	0.22	0.22	0.31	0.39	0.33	0.09	0	0
14	0.08	0.15	0.16	0.32	0.13	0.22	0	0.35	0	0.17
15	0.10	0.16	0.07	0.21	0.25	0.22	0	0.14	0	0
16	0.14	0.62	0.36	0.60	0.38	0.30	0	0.35	0.22	1.00
Average	0.12	0.18	0.22	0.30	0.29	0.33	0.21	0.21	0.25	0.22
No. of teams	112	33	16	6	10	6	1	4	1	1

TABLE VI.
ABILITY TYPES AND RADAR CHARTS

General class A group generally does not have outstanding performance in most items but could possibly be good in a certain area (Total 161teams). Three types are found according to their outstanding items.	
The common type (Cluster #1) Indicates that there is no outstanding characteristic in any item (0%).	
The presentation type (Cluster #2) Indicates excellence in presentation technique (33%).	
The technology-excellence type (Cluster #3) Indicates excellence in countermeasure technique (37%).	

Process class A group that has above-average performance in each item but is excellent in a few (Total 22 teams). Three types are found according to their outstanding items.	
The presentation type (Cluster #4) Indicates excellence in presentation technique (83%).	
The SQC type (Cluster #5) Indicates excellence in the application of analytical tools (80%).	
The performance confirmation type (Cluster #6) Indicates excellence in identifying and maintaining the effectiveness of the performance (100%).	

Extraordinary class: Indicates that one particular item is outstanding, while others are not so good (Total 7 teams). This class cannot become the prototype for learning. There are 4 types in this class (Clusters #7 to #10). (radar charts are omitted) (14%)

Note: The percentage in the parentheses (.) is the percentage of the participants of this ability type who won the excellence award in the national presentation.

D. Evaluator Characteristics

In order for an evaluation to generate useful results, the evaluators must be qualified. If no one in a class receives "A"s, does this imply that there is no outstanding student in this class, or did the evaluator/teacher set the standard too high? The answer could depend on the characteristics of the evaluators.

Using the same evaluation data and analytical methods, the characteristics of the 249 evaluators in the national presentation were analyzed. First, the average of "A" for each item and each evaluator is generated, that is, dividing the number of "A"s given by an evaluator by the number of teams such evaluator had ever evaluated. Call this average the grading index, which is equivalent to the superiority index for ability. Thus, 1 indicates that this particular evaluator gave "A" on this item to all teams he/she evaluated; 0 indicates that no "A" was given by this particular evaluator on this item; 0.1 indicates that one out of ten teams received "A"s from this particular evaluator, and so on. Next, based on these grading indices, the same clust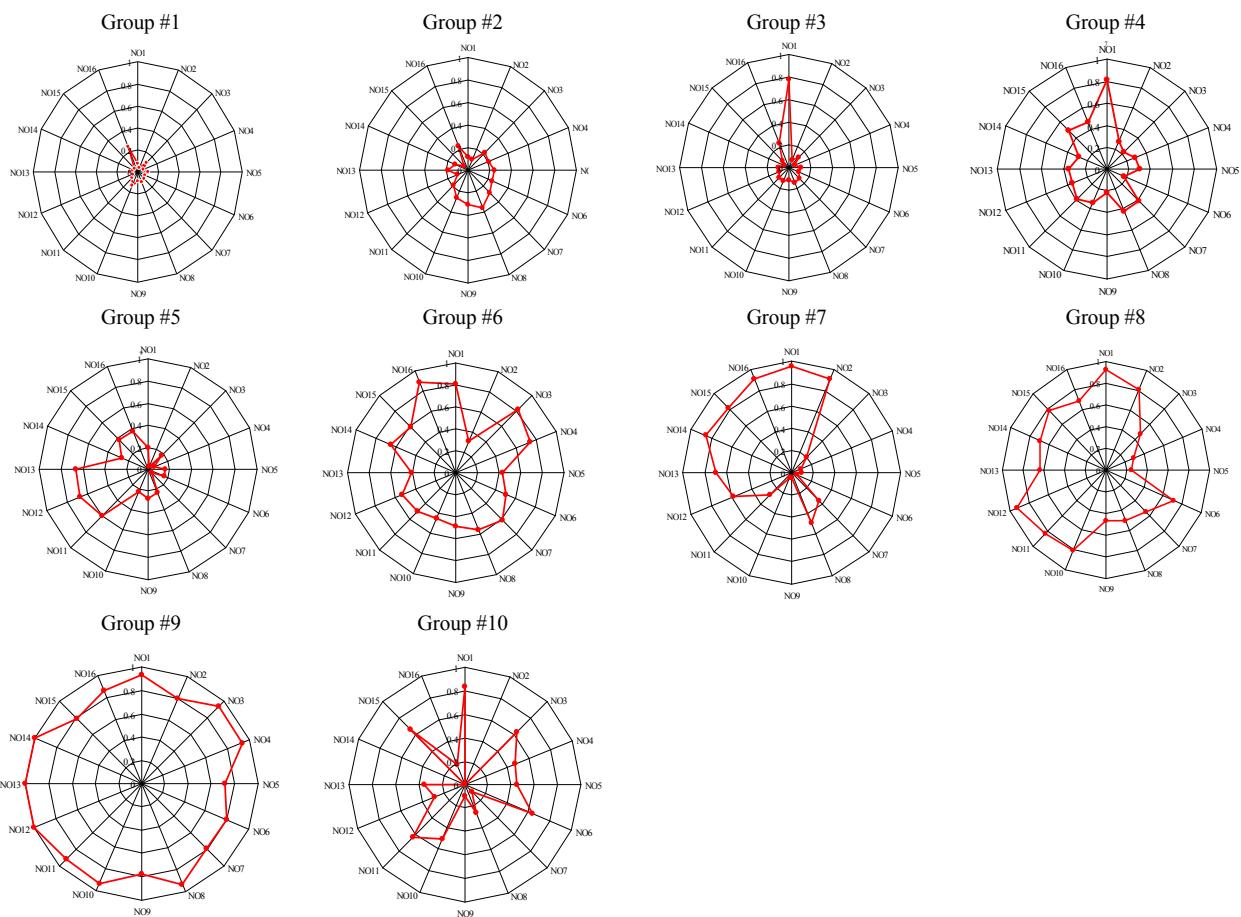er analysis was applied to identify 10 clusters. The cluster means and their associated radar charts are shown in Table VII.

Since the first four groups cover 232 evaluators (93%), only these four groups are discussed. Table VII shows that for Group #1, except for Item #16, the cluster means are relatively low. The average of 16 items is 0.087. This indicated that the evaluators in this group seldom gave an "A". The cluster means in Group #3 are slightly higher than those in Group #1, except for Item #1, which is extremely high. Without Item #1, the average values of these two groups are quite close (0.088 vs. 0.112). The cluster means of Group #2 are consistently higher than their counterparts in Group #1. The cluster means of Group #4 are also much higher than those of their counterparts in Group #1, and that of Item #1 is particularly high.

For individual items, there are a few findings: the high (0.816) and low (0.074) cluster means of Item #1 are significant; the cluster means of Item #2 among groups are consistently low; the cluster means are high for Items.

TABLE VII.
CLUSTER MEANS AND RADAR CHARTS OF EVALUATOR GRADING INDEX

Cluster \ Item	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
1	.074	.117	.775	.816	.193	.810	.952	.929	.929	.833
2	.036	.104	.073	.272	.020	.314	.905	.807	.786	0
3	.113	.223	.125	.225	.173	.819	.190	.464	.929	.633
4	.062	.209	.060	.270	.060	.743	.095	.286	.929	.467
5	.102	.247	.109	.296	.160	.429	.095	.243	.714	.450
6	.075	.241	.094	.171	.167	.505	.048	.714	.786	.633
7	.088	.284	.143	.414	.040	.610	.352	.543	.786	.083
8	.097	.362	.141	.408	.227	.562	.486	.507	.929	.250
9	.086	.302	.114	.216	.267	.486	.048	.464	.774	.100
10	.126	.270	.124	.337	.220	.457	.048	.807	.929	.500
11	.076	.196	.129	.383	.601	.495	.286	.829	.917	.633
12	.053	.106	.095	.342	.673	.533	.581	.929	1	.283
13	.074	.188	.093	.346	.667	.410	.695	.636	1	.350
14	.049	.126	.054	.279	.269	.657	.857	.686	1	0
15	.044	.055	.089	.492	.373	.590	.819	.779	.786	.667
16	.243	.232	.231	.458	.369	.886	.905	.686	.857	.183
Average	.087	.204	.153	.358	.280	.582	.460	.644	.878	.379
Average w/o item#1	.088	.210	.112	.327	.286	.566	.427	.625	.875	.349
No. of evaluators	119	68	27	18	5	3	3	2	2	2
Percentage of evaluators	47.78	27.31	10.84	7.23	2.01	1.20	1.20	0.80	0.80	0.80

Group #1   Group #2   Group #3   Group #4

Group #5   Group #6   Group #7   Group #8

Group #9   Group #10

IV. INTERPRETATION AND SUGGESTIONS

A. Team Improvement Ability

Four conclusions can be summarized from the results: 1) the ratio of the process class is relatively low; 2) presentation type is shown in each class, even in the extraordinary class; 3) there is no technology-excellence type in the process class; 4) the theme value was not shown.

Further examination can be done on the above four conclusions. For example, why does the theme value not show? Was the item description not clear enough? Item #2 is a good example since its cluster mean is low in most groups. The description "Is the substitution proper?" (Note that this is a direct translation from Chinese which is different from the description shown in Table I) seems vague. The real meaning of this item is "Does the theme properly indicate the problem?" This could be the reason why evaluators could not give a high score.

Next, why is there no technology-excellence type in the process class? This is because the innovation ability of teams is either not sufficient, or the problem came from the quality of evaluators. A proper solution, being either finding a suitable regenerating input or improving the evaluation system, cannot be applied unless the true reason is determined.

B. Characteristics of Evaluators

Table VII shows that Groups #1, #2, and #3 account for 86% of all evaluators. Yet, those evaluators rarely gave "A"s, especially Group #1, which seldom gave an "A" on any item. Group #3 has a high grading index (0.775) on Item #1, which tells us that most of the evaluators in this group appreciated the themes presented. The two evaluators in Group #9 gave every item a high grade. Group #1 and Group #9 show completely opposite results.

Items #1 and #2 are the only two items that are related to theme. The cluster mean of Item #1 has mixed high and low values. This is evidence that the evaluators' perspectives are diverse and the itemized rating method is not proper to identify a team's outstanding ability. The cluster means of Item #2 are very low for all groups. A reasonable explanation of this finding is that the description is vague, as mentioned in the previous section.

Most of the cluster means of Items #7 through #10 are high, which indicates that most evaluators appreciated the teams' performance in terms of "countermeasure." Of course, performance is closely related to the degree of difficulty of the theme/problem. Since the ratio of technology-excellence type is low, it may be reasonable to infer that the selected themes/problems are not difficult.

Factoring out personal preferences such as personal relationships among evaluators and participants, extreme conservatism (the one who is hard to please, giving unreasonably low scores), or extreme optimism (the one who is easy to please, giving high scores indiscriminately), the difference between the above two disparate evaluators (Group #1 and Group #9) tells us that either they lost the ability to distinguish the superiority of teams or that the description of the items evaluated was too vague to be understood. To cope with the first difficulty, perhaps the obvious thing to do is to re-educate the evaluator with a purpose toward certification. For the second problem, a redesigning of the evaluation method and items can help.

V. CONCLUSION

In the past, people have taken different approaches toward CI activities. Some have focused on problems, some have focused on models and tools, and some have focused on promotions. Everyone thinks that his or her approach is the most effective one. As Deming expressed, what is best for a sub-system may not be the best for the entire company [12]. Focusing on one area worked well in the past because the evolution of CI was still at the lower level. However, when it reaches a higher level, the ability has to be upgraded in order to sustain the CI activity.

This study applied a new analytical method on the evaluation data from the national presentation in Taiwan. The results provide two kinds of information. One is the improvement levels of companies and their ability types. Individual companies can realize their strengths and weaknesses and thus inject the proper regenerated input for future activity. The other is the characteristics and quality of evaluators. The sponsor can modify the system accordingly to make the presentation more fair and reliable.

This introduced method can be applied to various fields, especially to public issues, such as education, healthcare, social welfare, etc., so that resources can be used effectively. The other contribution of this paper is that it increases the use of evaluation data.

REFERENCES

[1] P. Berkhin, "A survey of clustering data mining techniques," *Accrue Software*, 2002, http://citeseer.ist.psu.edu/berklin02survey

[2] J. Bessant, S. Caffyn, and M. Gallagher, "An evolutionary model of continuous improvement behavior," *Technovation*, vol. 21, pp.67-77, 2001.

[3] C.B. Chang, *Multivariate analysis*, Taipei: Wen Fu Press, 1993 (In Chinese).

[4] C.L.J. Chen and C.W. Wu, "An integrated structural model and its diagnostic and analytical approach for continuous improvement activity," in a book entitled, *Taiwan's Economic Transformation in Evolutionary Perspective: Entrepreneurship, Innovation Systems and Government*, edited by Tony Yu, New York: Nova Science Publishers, 2007, pp.1-18.

[5] R.B. Coronado and J. Antony, "Critical success factors for the successful implementation of six sigma projects in organizations," *The TQM Magazine*, vol.14, No. 2, pp.92-99, 2002.

[6] W.E. Deming, *Out of the Crisis*, Cambridge: MIT Press, 1986.

[7] A. Ginsburg and N. Rheit, "Building a better body of evidence: new opportunities to strength evaluation utilization," *American Journal of Evaluation*, vol. 24, pp.489-498, 2003.

[8] J.P. Guildford, *Fundamental Statistics in Psychology and Education,* New York: McGraw-Hill, 1965.

[9] G.T. Henry and M.M. Mark, "Beyond use: understanding evaluation's influence on attitudes and actions," *American Journal of Evaluation*, vol. 24, pp.293-314, 2003.

[10] A. Ieta, G. Silberberg, and Z. Kucerovsky, and W.D. Greason, "On scales and decision-making based on arithmetic mean," *Quality & Quantity*, vol. 38, pp.559-575, 2004.

[11] M. Imai, *Kaizen: the Key to Japan's Competitive Success*, New York: McGraw-Hill, 1986.

[12] W.J. Latzko and D.M. Saunders, Four days with Dr. Deming-A strategy for modern methods of management, New York: Addison-Wesley, 1996.

[13] M.Q. Patton, *Utilization-focused evaluation: The new-century text*, Thousand Oaks, CA: Sage Publications, 1997.

[14] L.R. Peck, "Using cluster analysis in program evaluation," *Evaluation Review*, vol. 29, pp.178-196, 2005.

[15] P. Senge, *The Fifth Discipline: the Art and Practice of the Learning Organization*, New York: Doubleday, 1990.

[16] C.S. Tsong, C.C. Liu, and S.L. Chen, "A study on the duplex-pole evaluation method for QCC presentation," *ICQCC83*, Taipei, 1983 (In Chinese).

[17] C.H. Weiss, "Have we learned anything new about the use of evaluation?" *American Journal of Evaluation*, vol. 19, pp.21-33, 1998.

[18] T.X. Wu, "Reliability and validity on attitude and behavior research: theory, application, and rethinking", *Public Opinion Monthly*, vol. 101, pp.29-53, 1985 (In Chinese).

[19] C.W. Wu and C.L. Chen, "The evaluation and analysis in the continuous improvement integrated structural model," *CIIE2005*, Hsinchu, Taiwan, 2005 (In Chinese).

[20] C.W. Wu and C.L. Chen, "An integrated structural model toward successful continuous improvement activity," *Technovation*, vol. 26, pp.697-707, 2006.

Author

C. L. J. Chen is with the Department of Economics, Feng Chia University, Taichung, Taiwan (e-mail: clchen@fcu.edu.tw).