Testing Cross-Group Measurement Invariance of Quality Management Model

C. H. Ng, Xiande Zhao, T. S. Lee, Jeff Yeung
Department of Decision Sciences and Managerial Economics
Chinese University of Hong Kong
Email: chng@baf.msmail.cuhk.edu.hk

Abstract

This paper presents a rigorous approach to identify the various kinds of disagreements between top and middle management of PRC manufacturing firms towards quality management (QM). It represents the first attempt in QM to empirically tackle the disagreements between different tiers in organizations. Indeed, the criticality of consensus between tiers in organizations to the implementation of quality management is beyond doubt but has never been empirically explored in QM research. We use confirmatory factor analysis (CFA) to conduct cross-group comparison by evaluating the Measurement Equivalence/Invariance (ME/I) of a quality management instrument. By doing this, we are able to discover different types of conceptual and psychometric disagreements between the groups and their impacts on perceived organizational performances. These findings are however cannot be detected by traditional mean comparison methods (e.g., ANOVA, *t* test).

Our study reveals that top and middle management in PRC manufacturing firms have a general consensus towards the framework of quality management but have significant disagreements on the perceived importance of each quality practice related to Leadership, Customer & Market Focus and Human Resource Focus. This exposes a serious problem of QM in PRC manufacturing firms that while the hard factors of QM are conceptually consented, its soft factors are however un-harmonized. Besides, different response patterns to the QM survey instrument (psychometric disagreements) between top and middle management are found significantly affecting the survey results. Without knowing and deducting these response artifacts, the conclusions drawn relating to QM performances and its impact on organizations are at best ambiguous and at worst erroneous.

1. Introduction

It's not uncommon that the Critical Success Factors (CSFs) of Quality Management (QM) are identified solely based on the responses from top management [2][10][14][26][27] whose perspectives are assumed to be representative of others in the organization. However, we argue that solely relying on the responses of single respondent would probably provide distorted or incomplete pictures of the status quo of QM. The assumption that top managers thoroughly grasp the details of QM is always skeptical, particularly when the organization is huge and top managers are too busy to know the operations in details [cf. 6].

It is widely accepted that QM emphasizes the involvement and collaboration of all tiers in organizations. In general, top management is responsible for the setting up the strategic directions of CSFs of QM while middle management plays the role of strategy execution. The lack of conceptual consensus between top and middle management however, will definitely be detrimental to the QM deployment and implementation. On the other hand, Cheung & Rensvold [9] argue that the observed conceptual differences may be due to one or more measurement artifacts unrelated to the construct of interest. It is therefore important to identify and interpret such artifacts, particularly in cross-group work.

The objective of this study is two-fold: (1) to identify the conceptual and psychometric differences and invariance between top and middle management towards QM; and (2) to evaluate the impact of these differences on QM implementation. Conceptual differences refers to raters are using different conceptual framework to assess performances. They may be using different measurement items to represent the same performance dimension or assign different weights to the items. On the other hand, psychometric differences imply that different groups may have consensus on the conceptual framework but respond to the measurement scale differently [6][13].

According to Vandenberg & Lance [32], traditional mean comparison methods are usually used to identify the difference of perceptions between the two groups. However, it does not help to identify the conceptual and psychometric differences. Although a variety of techniques have been used to assess the extent of difference or invariance of a measurement instrument across groups [cf. 19], multi-group confirmatory factor analysis model [20] represents the most powerful and versatile approach [32]. Confirmatory factor analysis has advantages over the traditional methods (e.g. ANOVA) including firstly, if substantial measurement non-invariance exists across groups, it is inappropriate to directly compare mean group differences based on non-invariant items; secondly, traditional methods do not take into account of the dis-attenuation of by measurement errors; and finally, it is not possible for traditional methods to control partial measurement invariance, which is however critical in cross-group comparison [34].

Evaluation of Measurement Equivalence/ Invariance (ME/I) is used to identify conceptual and psychometric differences. ME/I is a logical prerequisite to

conduct cross-group comparison. But it has never been studied in QM research. The importance of identifying ME/I lies in the desire to make meaningful inferences regarding the status of different groups on the measures and to draw conclusions as to how this group difference may affect organizational functioning [34]. If evidence supporting a measure's invariance is lacking, conclusions based on that scale are at best ambiguous and at worst erroneous [32]. On the other hand, testing of ME/I also helps to understand the characteristics of different populations. The identified ME/I will provide directions for improving the consensus between top and middle management towards the QM implementation.

This article is organized in the following manner: Firstly, we discuss the concept of measurement equivalence/invariance (ME/I). Secondly, we introduce our QM measurement model and the series of hypothesis that can identify the conceptual and psychometric ME/I across the two groups: top and middle management. Thirdly, we undertake cross-group confirmatory factor analysis to test the series of hypothesis for determining ME/I. Fourthly, we discuss the findings of tests. Finally, we present implications and directions for further research.

2. Measurement Equivalence/Invariance

According to Horn & Mahrdle [18, p.117], ME/I refers to "whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute". Generally speaking, ME/I can be categorized into configural invariance, metric invariance, scalar invariance, factor variance/covariance invariance, random measurement error invariance, and factor means invariance [6][8][32]34].

Vendenberg & Lance [34] suggest that the basic assumptions underlying cross-groups comparison are firstly the conceptual equivalence of latent variable (ξ) in each group; secondly, the equivalence associations (λ) between each measurement item and ξ across groups; and finally, the extent to which the measurement items are influenced to the same degree by the same unique factors (δ) across groups. The authors further argue that it does not make sense to compare mean group differences based on measures with substantial measurement in-equivalence.

3. Measurement Model and Hypotheses

The measuring instrument of the seven-factor QM model derived from Lau, et al.[22] is examined in this study. Since we split the samples into top and middle management, the original instrument is re-purified by factor analysis. Redundant items, items with factor loadings less than 0.45 and/or cross-loadings with difference less than 0.1 between the largest and second largest will also be eliminated [24]. Except Leadership shows a marginal Cronbach's alpha for top manager group: 0.40, all have values over 0.60. For each group, the construct validity of each construct is evaluated by correlating items within the construct and items across constructs. Results show that most of the items, except those of Leadership, have a higher correlation with items within construct than across constructs. This testifies the constructs' acceptable convergent and discriminant validity [25]. The relatively weaker reliability and construct validity of Leadership, however, signify its needs for improvement of items in future. The seven-factor model (Model 1) is shown in Figure 1, with each item measured on a 9-point scale as shown in Appendix I. Figure 2 depicts the test flow of hypotheses. In the following paragraphs, we discuss different forms of disagreements based on the examination of ME/I.

3.1 Conceptual Disagreement

3.1.1 Differences in Factor Form - Configural Differences (Test of Model 1 & 2)

Differences in factor form arise from different number of factors, and different number and locations of non-salient and salient factor loadings under the same factor [32]. To prove configural invariance, the same configuration of salient and non-salient factor loadings should exist in the measuring instrument across different populations [18] but it is unnecessary to constraint the salient factor loadings to be equal across groups [6].

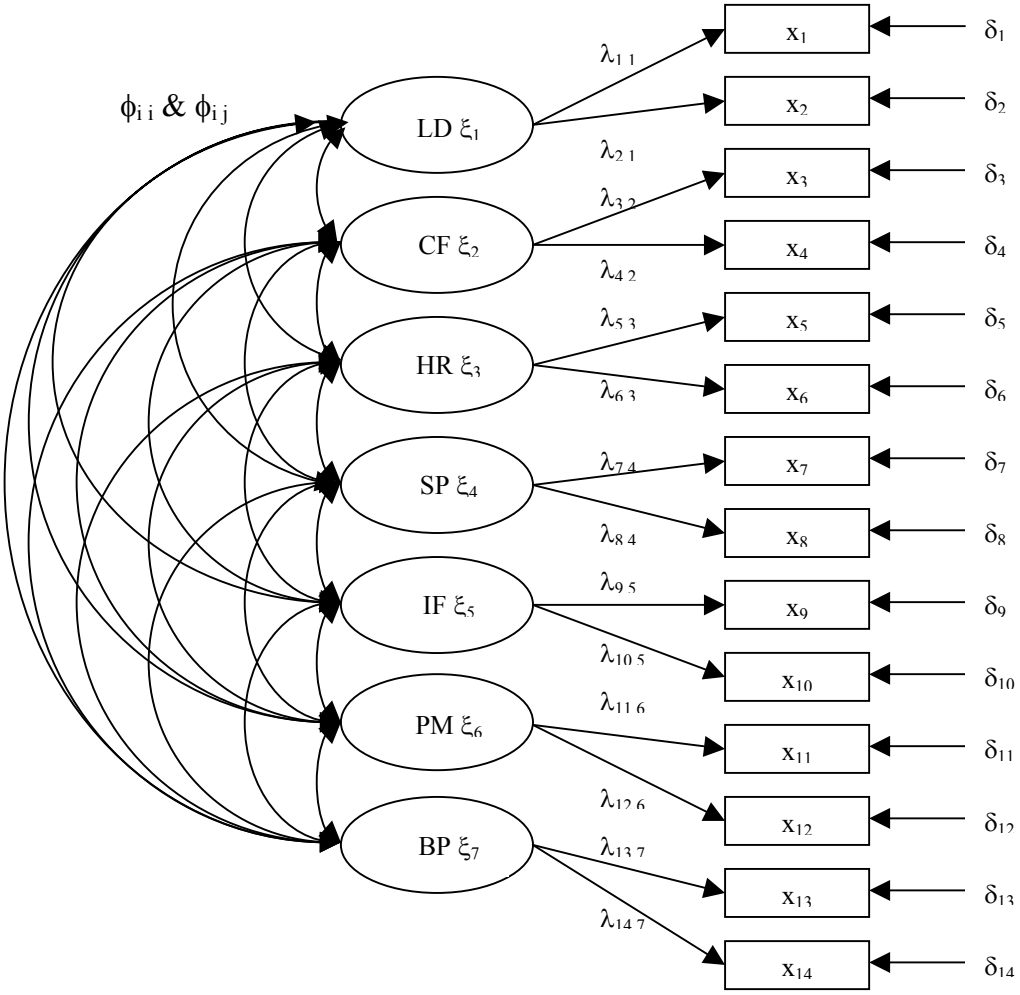
If top and middle managers are holding different concepts about QM, it implies that firstly, one party may do something enthusiastically but is deemed as superfluous by the counter party and may results in sarcasm. Secondly, both parties will prioritize their efforts in an un-concerted manner based on their different perceptions and this will result in conflicting relationship. Finally, there will be serious deployment gap occurring. Lack of consensus on QM implies the strategy established by top would become an empty talk to bottom. To test the existence of configural invariance, a best-fit model will be established separately for top and middle management (Model 1A and Model 1B respectively) by fitting our QM model (Model 1) with the two sets of data separately using CFA.

H_{O1}: Model 1 fits the two sets of data

After the fitted models are established, CFA is then run again to establish a baseline model with Model 1A stacked with Model 1B [3]. Factor loadings are not constrained to be equal across groups [6]. i.e.,

$$H_{O2}: \Lambda_{form}^{(top\ mgr)} = \Lambda_{form}^{(mid\ mgr)}$$

If H_{O2} is not rejected, it implies that the baseline model (Model 2) which constraints Model 1A and Model 1B having the same number of factors and same pattern of salient and non-salient factor loadings, will not be rejected.



LD: Leadership

CF: Customer and market focus

HRM: Human resource focus

SP: Strategic Planning

IF: Information and analysis

PM: Process management

BP: Business Performance

Figure 1 Measurement Model

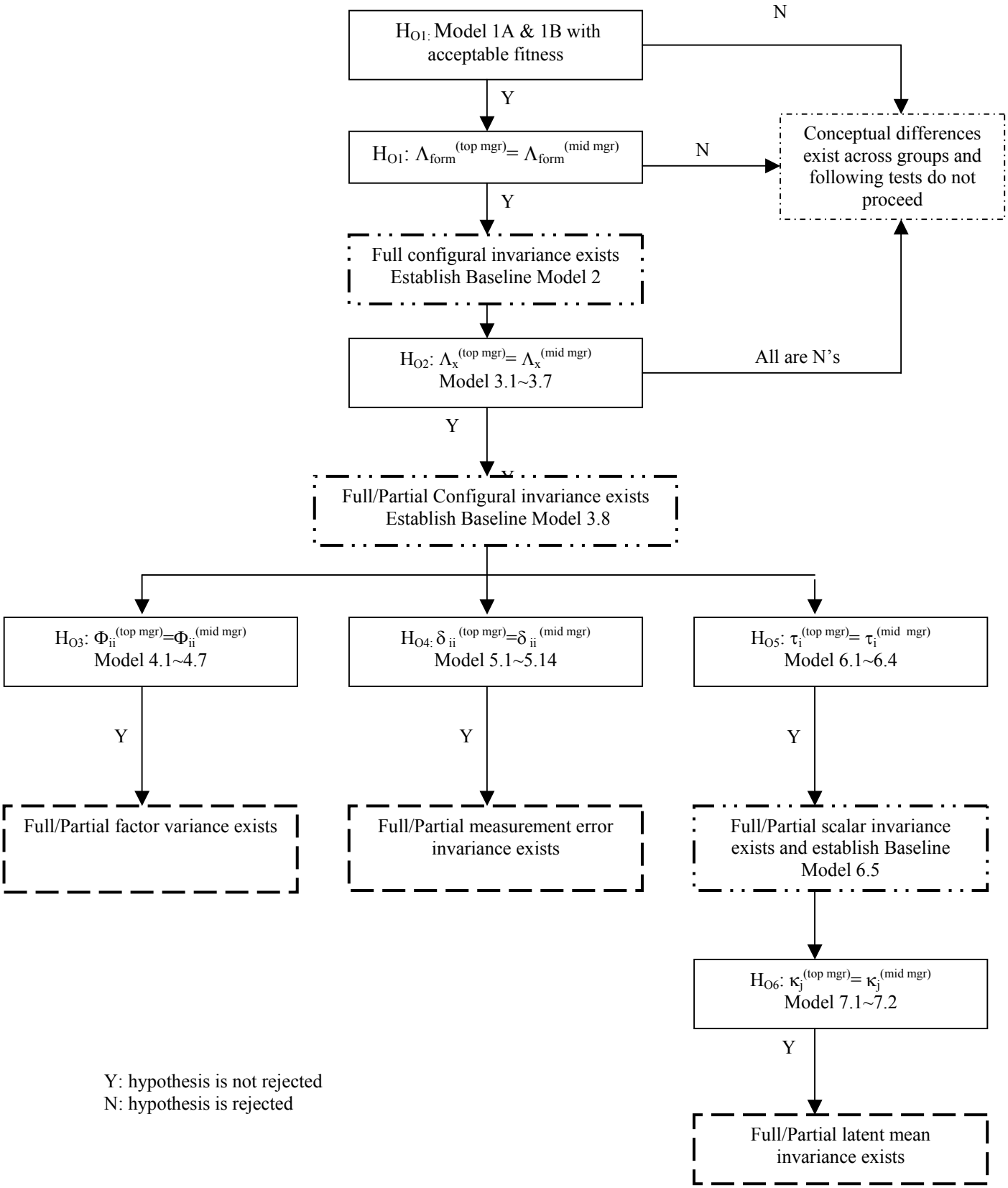


Figure 2 Hypothesis Testing Flow-chart

3.1.2 Differences in Factor Loadings – metric/factorial differences (Test of Model 3's)

Factor loadings (λ_{ij}) are the regression slopes relating X_i to their corresponding latent variables, ξ_j [3][34]. The differences in the magnitude of factor loadings reflect that both parties have the same conceptual framework (i.e., configural invariance) about the seven-factor QM model but differ in acknowledging the strength of relationships (λ_{ij}). Difference in λ_{ij} might result in conflict of efforts and waste of resources in implementing QM because different parties will have their own rank of importance and work in their own agendas. By using χ^2 difference test, a constrained model established for each construct will be compared with Model 2 separately by requiring all the factor loadings of that construct to be equal across groups and the factor loadings of other constructs to be freely estimated, i.e.,

$$H_{O3}: \Lambda_x^{(\text{top mgr})} = \Lambda_x^{(\text{mid mgr})}$$

If H_{O3} is rejected, it implies that the tested construct is non-invariant and the two corresponding measurement items are thus non-invariant too.* Upon completion of identification of invariant and non-invariant construct(s), a new baseline model (Model 3.8) will be established by setting all λ's of invariant constructs equal across groups while those of non-invariant constructs as freely estimated.

*if the construct has more than two items, identification of non-invariant items should follow the procedure suggested by Cheung & Rensvold [8].

3.2 Psychometric Disagreement

3.2.1 Difference in Latent Factor Variability (Test of Model 4's)

Range restriction is operationalized as the degree of latent factor variance [6]. Lower variance indicates higher range restriction. Factor variance (Φ_{ii}) represents the dispersion of the latent variable (ξ_j) within groups [34]. According to Cheung [6], the difference in latent factor variability is caused by the difference in range of response intervals between different parties [16]; different norms of responding [33]; raters' unwillingness to justify high or low ratings [23]; or disagreement concerning scale intervals and definitions of effective and ineffective performance [36]. Halo effect may also contribute to the difference of latent factor variability and produce low factor variability [6]. Rejection of equality of factor invariance indicates that the group with a smaller factor variance is using a narrower range of the construct continuum in responding to the survey items [34] even though the performance of the measured construct should be much more diverse to the group. As a result, the associations between the QM factors may be overstated as their variances become small. The difference in latent factor variability can be tested separately for each construct (Model 4.1~4.7) by adding the constraints of Φ_{ii} of target construct to be equal across groups on Model 3.8 while the variances of other constructs will be set freely estimated:

$$H_{O4}: \Phi_{ii}^{(\text{top mgr})} = \Phi_{ii}^{(\text{mid mgr})}$$

3.2.2 Difference in Random Measurement Error (Test of Model 5's)

The difference in random measurement error, according to Cheung [6], may be due to unfamiliarity with scale items; inexperience with the rating format; informational constraints; errors in marking responses; and carelessness. All these errors can be attributed to the reliability problem of the raters. The difference in measurement error variance is tested separately item-by-item (Model 5.1~5.14). Each time, the constraint of equality of θ_{ii} (the variance of measurement error δ_i of item i) across groups [21] is added to Model 3.8, i.e.,

$$H_{O5:}\, \theta_{ii}^{(\text{top mgr})} = \theta_{ii}^{(\text{mid mgr})}$$

3.2.3 Difference in Latent Means (Test of Model 6's & Model 7's)

The overall rating that is given to a latent factor can be operationalized as the latent mean [6]. Byrne [5] states that the means of latent variables are unobservable but are derived indirectly from the indicator variables. Therefore, testing latent mean differences is in fact testing the associated mean structures including the test of scalar (intercept) invariance and latent mean invariance. The scalar invariance has to be proved before latent mean invariance can be examined [7][34]. Difference in latent mean is usually due to leniency or stringency [6] and unnecessarily implies that the groups have different concepts of interest. There are two consecutive tests to be conducted. Test(s) of scalar/intercept invariance has to be conducted prior to the test(s) of latent mean invariance. On the other hand, test(s) in scalar invariance can only be done for those items having factorial invariance in priori [7][32][34]. The hypothesis is as follows:

$$H_{O6}: \tau_i^{(\text{top mgr})} = \tau_i^{(\text{mid mgr})}$$

This model is established by constraining intercept τ_i of the equation $x_i = \tau_i + \lambda_{ij}\xi_j + \delta_i$. equals across groups. This constrained model will be compared with the Model 3.8. Finally, a consolidated baseline model capturing all the identified invariant and non-invariant factor loadings and scalars, will be established for the test of latent mean invariance. Test of latent mean difference is achieved by additionally constrain the equality of latent mean "κ_j" across groups. Latent mean comparison can only be done on those constructs satisfied the conditions of factorial invariance and scalar invariance. The hypothesis is as follows:

$$H_{O7}: \kappa_j^{(\text{top mgr})} = \kappa_j^{(\text{mid mgr})}$$

Comparing this model with the baseline model established in scalar invariance tests, will provide conclusion on the significance of latent means invariance.

4. Sampling and Data Collection

Data was jointly collected by China Association of Quality and a research team at the Chinese University of Hong Kong [22]. Two tiers of people including top and middle management from 600 organizations were selected for our study. Finally, a group of top managers (N=103) and middle managers (N=433) are selected for evaluation.

5. Results and Discussion

5.1 Conceptual Disagreement – Configural and Factorial Difference

The test results are shown in Table 1. The results show that there are significant disagreements in some of the testing areas. Our results reveal that basically, both top and middle management agree that QM is composed of seven dimensions. This reflects that PRC manufacturing enterprises have wide acceptance on the modern quality management concepts that quality is no longer a narrow focus on statistical process control but encompasses a variety of technical and behavioral methods for improving organizational performance [10]. Moreover, both parties also have consensus on how to measure each dimension. This indicates that they do not only have agreement on the conceptual definition of QM but also have consensus on the detailed activities that should be undertaken under QM.

Nevertheless, this study reveals a potential problem that disagreements of importance significantly occur in all the detailed activities of QM soft factors including Leadership, Customer & Market Focus and Human Resource Focus. It seems that the implementation of QM soft factors are more difficult than that of hard factors. At least, this study reveals that even though both parties have conceptual consensus on the dimensions and the compositions of each dimension, disagreement on the importance of each measurement item implies that both parties are working on their own agendas with different priorities. These un-concerted efforts may elicit waste of resources, generation of conflict, and more importantly, the ambiguity of strategic directions. The existence of these disagreements also reflects that top management overlooked the QM implementation issues such as communication and visibly promotion to achieve a common understanding across tiers of the prioritization of QM activities.

Table 2 reveals that top and middle management hold opposite views towards the rank of λ's of Leadership but similar views towards the rank of λ's of Customer & Market Focus and Human Resource Focus. Examination of Leadership at the item level reveals that top management concern more about the social liability of the organization while middle management believe that effective QM leadership should emphasize more on product and service quality. It can then be imagined that the enthusiastic efforts of top management in participating in social services may be perceived by middle management as a waste of time and a knock at an open door. In contrast, top management would be less supportive to those activities related to improving product and service quality. This may results in several vicious concomitants. Firstly, the sustainability of QM is vulnerable for sure as top management do not take the lead. Numerous studies find that top management leadership is the prime enabler for QM success [2][14][27][35]. Secondly, the enthusiasm of middle management will die away due to the lack of synchronized support from top management. Finally, QM will not be able to foster cultural change and will result in superficial and perfunctory implementation.

On the other hand, even though top management rank a higher importance on customer management system than middle management do, both parties deem that customer management system is more important than benchmarking to Customer & Market Focus. This may reflect that firms in PRC are adopting a reactive rather than proactive orientation towards customer focus [cf. 30]. Table 2 also indicates that top and middle management disagree on the importance of the measurement items of Human Resources Focus. However, both groups perceive a higher importance in providing training for employees to improve their work competency than providing a safe and healthy work environment. This reflects that workers in PRC have not yet been cherished as valuable assets. Even worst, productivity improvement may be pursued at the expense of workers' safety. Ironically, QM should be a humanistic and systematic approach to management [4].

5.2 Psychometric Disagreement – Measurement Error Differences

Table 3 shows that top management's ratings bear more measurement errors in five of the fourteen measurement items while middle management have three of them possess more errors. This suggests that top management who do not have time to observe their operations, would generate more biased pictures about QM when they respond to the survey items. Besides, top management may not be able to access the detailed information about various QM performances in their organizations. Unequal measurement errors will sway the generalizibility of relationships between the QM factors. One would find the self-ratings can produce one set of significant relationships, while use of others' ratings can produce another set, leaving the researcher in a conceptual quandary [6].

Table 1 Measurement Model Invariance Tests Summary

Test No.	Nature of model	χ^2	df	$\Delta\chi^2$	Δdf	Reject if p<0.05	RMSEA	TLI	CFI	Reject if ΔCFI< -0.01	A/R*
1A	Top management	108.38	56				0.0751	0.9309	0.9575		A
1B	Middle management	177.39	56				0.0711	0.9553	0.9725		A
2	Baseline Model	285.77	112				0.0719	0.9500	0.9692		A
3.1	Baseline + LD invar	291.81	113	6.04	1	0.01399	0.0724	0.9490	0.9683	-0.0009	R
3.2	Baseline + SP invar	285.78	113	0.01	1	0.92034	0.0713	0.9507	0.9694	0.0002	A
3.3	Baseline + CS invar	290.07	113	4.30	1	0.03809	0.0720	0.9495	0.9686	-0.0006	R
3.4	Baseline + IF invar	286.65	113	0.88	1	0.34902	0.0715	0.9505	0.9692	0.0000	A
3.5	Baseline + HR invar	295.21	113	9.44	1	0.00212	0.0734	0.9480	0.9677	-0.0015	R
3.6	Baseline + PM invar	286.31	113	0.54	1	0.46243	0.0715	0.9505	0.9693	0.0001	A
3.7	Baseline + BP invar	287.09	113	1.32	1	0.25149	0.0715	0.9503	0.9692	0.0000	A
3.8	PFI (SP, IF, PM, BP invar)	288.61	116	2.84	4	0.58452	0.0702	0.9520	0.9694	0.0002	A
4.1	PFI+ PH(LD invar)	289.43	117	0.82	1	0.36533	0.0698	0.9525	0.9695	0.0003	A
4.2	PFI+ PH(SP invar)	290.40	117	1.79	1	0.18123	0.0700	0.9522	0.9693	0.0001	A
4.3	PFI+ PH(CS invar)	293.47	117	4.86	1	0.02753	0.0700	0.9514	0.9687	-0.0005	R
4.4	PFI+ PH(IF invar)	294.74	117	6.13	1	0.01331	0.0713	0.9510	0.9685	-0.0007	R
4.5	PFI+ PH(HR invar)	292.43	117	3.82	1	0.05072	0.0709	0.9517	0.9689	-0.0003	A
4.6	PFI+ PH(PM invar)	292.78	117	4.17	1	0.04121	0.0714	0.9516	0.9689	-0.0003	R
4.7	PFI+ PH(BP invar)	288.62	117	0.01	1	0.93099	0.0696	0.9527	0.9696	0.0004	A
5.1	PFI+TD 1 1 invar	292.64	117	4.03	1	0.04476	0.0711	0.9516	0.9689	-0.0003	R
5.2	PFI+TD 2 2 invar	294.32	117	5.71	1	0.01689	0.0720	0.9511	0.9689	-0.0003	R
5.3	PFI+TD 3 3 invar	455.99	117	167.38	1	0.00000	0.0968	0.9066	0.9399	-0.0293	R
5.4	PFI+TD4 4 invar	292.23	117	3.62	1	0.05717	0.0707	0.9517	0.9690	-0.0002	A
5.5	PFI+TD 5 5 invar	289.54	117	0.93	1	0.33551	0.0693	0.9525	0.9694	0.0005	A
5.6	PFI+TD 6 6 invar	456.89	117	168.28	1	0.00000	0.0974	0.9063	0.9397	-0.0297	R
5.7	PFI+TD 7 7 invar	304.10	117	15.49	1	0.00008	0.0740	0.9484	0.9669	0.0272	R
5.8	PFI+TD 8 8 invar	292.06	117	3.45	1	0.06324	0.0704	0.9518	0.9690	0.0021	A
5.9	PFI+TD 9 9 invar	288.63	117	0.02	1	0.89476	0.0696	0.9527	0.9695	0.0005	A
5.10	PFI+TD 10 10 invar	299.41	117	10.80	1	0.00102	0.0712	0.9497	0.9678	-0.0017	R
5.11	PFI+TD 11 11 invar	288.66	117	0.05	1	0.83109	0.0697	0.9537	0.9696	0.0018	A
5.12	PFI+TD 12 12 invar	289.07	117	0.46	1	0.49772	0.0699	0.9526	0.9695	-0.0001	A
5.13	PFI+TD 13 13 invar	293.82	117	5.21	1	0.02252	0.0708	0.9513	0.9687	-0.0008	R
5.14	PFI+TD 14 14 invar	292.75	117	4.14	1	0.04189	0.0699	0.9498	0.9689	0.0002	R
6.1	PFI+TX 3 TX 4 invar (SP)	288.96	117	0.34	1	0.55724	0.0697	0.9526	0.9695	0.0006	A
6.2	PFI+TX 7 TX 8 invar (IF)	288.76	117	0.15	1	0.70287	0.0697	0.9527	0.9696	0.0001	A
6.3	PFI+TX 11 TX 12 invar (PM)	290.11	117	4.34	1	0.03723	0.0699	0.9523	0.9693	-0.0002	R
6.4	PFI+TX 13 TX 14 invar (BP)	300.37	117	14.60	1	0.00013	0.0717	0.9495	0.9675	-0.0018	R
6.5	PSI (TX3 TX 4 TX 7 TX 8 invar)	289.10	118	0.49	2	0.78368	0.0692	0.9532	0.9697	0.0004	A
7.1	PSI+KA 2 (SP) invar	289.11	119	0.01	1	0.92034	0.0687	0.9539	0.9699	0.0002	A
7.2	PSI+KA 4 (IF) invar	290.88	119	1.78	1	0.18215	0.0690	0.9534	0.9695	-0.0004	A

✶ A: Accept R: Reject

✶ PFI: Partial Factorial Invariance PSI: Partial Scalar Invariance

✶ PH: ϕ TD: θ TX: τ KA: κ

Table 2 Scalar and Latent Mean Comparison

	Top Management	Middle Management
$\lambda_{1\ 1\ (LD)}^{\#}$	1.000	1.000
$\lambda_{2\ 1(LD)*}$	1.487	0.992
$\lambda_{3\ 2\ (SP)}^{\#}$	1.000	1.000
$\lambda_{4\ 2\ (SP)}$	1.174	1.174
$\lambda_{5\ 3\ (CS)}^{\#}$	1.000	1.000
$\lambda_{6\ 3\ (CS)}^{*}$	0.736	0.946
$\lambda_{7\ 4\ (IF)}^{\#}$	1.000	1.000
$\lambda_{8\ 4\ (IF)}$	1.043	1.043
$\lambda_{9\ 5\ (HR)}^{\#}$	1.000	1.000
$\lambda_{10\ 5\ (HR)*}$	0.628	0.857
$\lambda_{11\ 6\ (PM)}^{\#}$	1.000	1.000
$\lambda_{12\ 6\ (PM)}$	0.991	0.991
$\lambda_{13\ 6\ (BP)}^{\#}$	1.000	1.000
$\lambda_{14\ 6\ (BP)}$	0.937	0.937
$\tau_1^{\#}$	0.000	0.000
τ_2^{*}	-5.220	-0.498
$\tau_3^{\#}$	0.000	0.000
τ_4	-2.659	-2.659
$\tau_5^{\#}$	0.000	0.000
τ_6^{*}	2.180	0.480
$\tau_7^{\#}$	0.000	0.000
τ_8	-0.431	-0.431
$\tau_9^{\#}$	0.000	0.000
τ_{10*}	3.341	1.241
$\tau_{11}^{\#}$	0.000	0.000
τ_{12*}	0.460	0.642
$\tau_{13}^{\#}$	0.000	0.000
τ_{14*}	0.967	1.374
κ_1^{*}	8.615	8.424
κ_2	8.192	8.192
κ_3^{*}	8.063	8.039
κ_4	7.930	7.930
κ_{5*}	8.067	8.103
κ_6^{*}	7.471	7.364
κ_7^{*}	7.567	7.030

#: parameters of marker items *: parameters with value differ across groups at $p<0.05$

Table 3 Difference of variance and measurement errors between groups

	Top Management	Middle Management
$\Phi_{1\,1}$	0.806	0.806
$\Phi_{2\,2}$	1.513	1.513
$\Phi_{3\,3}$*	2.047	1.343
$\Phi_{4\,4}$*	2.511	1.687
$\Phi_{5\,5}$*	1.477	1.477
$\Phi_{6\,6}$	3.670	2.686
$\Phi_{7\,7}$	1.914	1.914
$\theta_{1\,1}$*	0.473	0.292
$\theta_{2\,2}$*	2.185	1.488
$\theta_{3\,3}$	0.437	0.294
$\theta_{4\,4}$*	1.724	1.724
$\theta_{5\,5}$	0.936	0.936
$\theta_{6\,6}$*	0.939	0.651
$\theta_{7\,7}$*	1.472	0.662
$\theta_{8\,8}$	0.606	0.606
$\theta_{9\,9}$*	0.366	0.366
$\theta_{10\,10}$*	0.361	0.687
$\theta_{11\,11}$	1.557	1.557
$\theta_{12\,12}$	0.324	0.324
$\theta_{13\,13}$	0.814	1.269
$\theta_{14\,14}$	0.285	0.524

* parameters differ across group at $p<0.05$

5.3 Psychometric Disagreement - Factor Variance Differences

Table 3 shows that there are significant differences between groups in latent factor variability of Customer & Market Focus, Information & Analysis, and Process Management. Middle management have lower values in all non-invariant factor variances indicating that they have narrower ranges of restriction than top management in responding to the measurement items [cf. 6]. According to Cheung [6], the difference in range of restriction may be due to the different norms of responding, for instance, unwillingness to justify high or low ratings; disagreement on differences between scale intervals; disagreement on the standard of effective/efficient performance; and halo effect that impression of high or low rating on one item may be generalized to the ratings of other items in the same factor. These potential causes will prohibit the respondents from significantly differentiate high and low factor performances and thus result in low associations among QM factors.

We run traditional variance comparison method by using F-test to evaluate the equality of population variances between top and middle management (variance ratio test) and find that no significant difference is found between two groups in all the seven QM factors.

5.4 Psychometric Disagreement – Scalar and Latent Means Differences

In testing latent mean invariance, we found that disagreement happens in Leadership (κ_1), Customer & Market Focus (κ_3), Human Resource Focus (κ_5), Process Management (κ_6) and Business Performance (κ_7) (refer to Table 2). The disagreements indicate the possibility of leniency or stringency, or the adoption of different performance standards across groups [6]. Except Human Resource Focus, top management in general rate higher in latent means than middle management do. The differences in intercept reflect systematic response biases (e.g. leniency) between groups [3][34]. Table 3 shows that five out of the seven QM factors have significant intercept differences across groups indicating that both parties may really influenced by leniency and stringency

in response to the survey items. For instance, the Leadership scale in the survey questionnaire in Appendix I, is basically measuring the top management performance. In this connection, it becomes a self-rating exercise for top management who tend to inflate their own performances [6][16][17][34].

We rank the seven factor performances in accordance with their latent means. The results reveal that both parties agree that those infrastructural soft factors of QM including LD, CS and HR are relatively performing better than those hard factors including IF and PM. The lower ranking of hard factors indicates that the investment in advancement of QM hard factors in PRC manufacturing organizations may be far behind those input in developing QM infrastructure. It is also surprised to note that both parties rate customer focus, which is one of the pillars of QM at a lower rank. This reveals that learning type of customer orientation has not been fully established in PRC manufacturing firms even after years of quality promotion [cf.30]. Lau et al.[22]'s findings support ours that only a low percentage of PRC manufacturing firms are systematically identify customer requirements and market trends.

We compare the ranking obtained from CFA with that measured by traditional mean comparison. The ranking pattern derived from traditional method is totally different from that established by CFA and is difficult to identify.

6. Implications and Conclusion

Nearly most of the previous QM studies adopt top management's view and seldom consider the perspectives of others at lower levels. Particularly, the view of middle management who are the key persons responsible for the dissemination and implementation of QM, was always overlooked. This research represents the first attempt to empirically tackle the perceptual differences between different tiers in organizations. Our study reveals that there are significant conceptual and psychometric disagreements between top and middle management in rating the QM performances of PRC manufacturing firms. By evaluating the ME/I of a QM measuring instrument, we are able to understand the different conceptual views between the two parties towards QM and their different response patterns that may contaminate the results of QM research. It is found that both groups reach conceptual consensus on the holistic framework of QM in terms of the number of dimensions and the corresponding measurement items. But they have disagreements about the importance of these measurement items, particularly, those of soft factors of QM including Leadership, Customer & Market Focus and Human Resource Focus. These soft factors are the components of QM infrastructure that enable the implementation of hard factors [2][14][26][35]. This study reveals that top management who are supposed to assume the prime role to drive the realization of soft factors, are found significantly less committed in quality than middle management. The sustainability of QM is thus doubtful.

This study also reveals that the implementation of Customer & Market Focus in PRC manufacturing firms is limited in its horizon because benchmarking with competitors is drawing less management attention. The focus of this QM factor is, in contrast, to establish customer management systems aiming at handling customer complaints. This narrow focus of customer orientation may attenuate the firm's responsiveness to the volatile and ever changing market environment. Moreover, more focus on customers' requirements, more convergent orientation that the firm will emphasize. This will drive the firm lingering on existing QM practices and hardly exhibiting exploratory behaviors [1][12][15].

The findings related to Human Resource Focus in PRC manufacturing firms make are worrying indeed. The less concern of employee safety reflects that in general, the implementation of QM in PRC manufacturing firms is pursuing full utilization of manpower resource without a synchronized concern for employees' welfare. In fact, the effectiveness of training to enhance competency is skeptical if lacking consideration of employees' welfare. Deming [11] strongly argues that driving out the fear of employees and let them work effectively is one of the major concerns in QM.

Our analysis discovers that psychometric differences are significantly affecting the QM survey. These response artifacts would cover up the real picture of QM implementation in organizations. In fact, the comparison of results from CFA with those of traditional mean and variance comparison methods (ANOVA) reveals that without controlling the measurement errors, traditional methods are unable to identify the disagreements found in this study. More importantly, they project a biased picture of QM implementation and may cause researchers in a conceptual quandary.

The findings of this study remind us that efforts should be devoted to foster consensus towards QM. Particularly, top management should involve more in promoting quality and spend more time in understanding the operations at lower level so that the quality views of two parties can be aligned. Furthermore, conclusions drawn based on single informant in QM studies are subject to bias. This practice is based on the assumption that the top manager in organizations is a reliable informant and the sampled organizations are homogenous that employees have no disagreement over the conceptual and psychometric issues. This study, however, reveals that these two assumptions are vulnerable. In fact, multiple rater approach should be encouraged. Furthermore, this study also sheds light on cross-group comparison that group differences may not be solely due to conceptual differences [6] but also arise from

psychometric disagreements, which should be identified and isolated before valid conclusions can be drawn.

7. Limitations and Future Research Directions

Limitations of this research include the reliance of perceptual data and the unequal sample size of top and middle management. Sample size requirements often limit the application of multiple group confirmatory factor analysis method [3]. Moreover, this technique is not a remedy to the problems of cross-cultural measurement equivalence but rather act as a diagnostic tool. Only care in the data-gathering process can prevent the occurrence of problems [29]. "Although problematic items can be removed from scales post hoc for subsequent testing, it should also be noted that this method can be counter to a theory-driven research agenda, and the true value of the methodology is on avoiding findings that may be skewed by measurement invariance across cultural groups. If this perspective is understood, this technique remains a useful tool for determining the value of measures within the cross-cultural context" [24, p.116].

This research represents the start of cross-group comparison in QM and much more work needs to be done. For instance, we have not yet explored the exact reasons why soft factors have a higher latent means than hard factors. We also have not investigated how to handle and/or improve the measurement items that are found cross-groups non-invariant. Besides, this study has not yet put focus on the inter-relationships among the QM factors. That is, we have not explored the ME/I on factor correlations and/or causal relationships among the factors. We believe that the exploration of disagreement(s) on these issues can provide further insights of QM implementation in PRC manufacturing firms and can shed light on the dynamics among the QM factors that are still controversial in quality management research.

In future, other studies may pursue to explore using the same research methodology for more comprehensive comparisons of QM model across different populations and more groups. It is also useful to extend the comparisons to include the responses of rank-and-file and front-line employees.

8. Appendix I

Leadership

X_1: senior executives take our product and service quality seriously

X_2: we participate enthusiastically in social and community services

Strategic Planning

X_3: In defining our strategic objectives, we carefully considered various potential factors such as market trends, competitive environment, and our capability

X_4: every employee in our organization understands our strategic objective and action plans

Customer and market focus

X_5: we have effective customer management system, which solves customer complaints or problems in a timely manner

X_6: we closely monitor our competitors' actions

Information and analysis

X_7: we have an effective system to assess our business performance

X_8: we adjust our performance indicators and appraisal systems according to the evolving internal and external business environment

Human resource focus

X_9: we provide training for our employees to improve their competency

X_{10}: we provide a safe and healthy work environment

Process management

X_{11}: when designing business processes, we carefully consider various factors, such as design quality, process cycle time, costs, new technology, and productivity

X_{12}: we continuously improve our production or delivery processes, enhancing the overall product and service quality

Business Performance

X_{13}: Employees are satisfied with the department for which they work

X_{14}: Our product quality has been improving steadily

References

[1] Benner, M., & Tushman, M. (2002), "Process management and technological innovation: a longitudinal study of the photography and paint industries" ***Administrative Science Quarterly***, 47:676-706.

[2] Black, S., & Porter, L. J. (1996). Identification of the critical success factors of TQM. *Decision Sciences*, Vol. 27, No. 1, 1-21.

[3] Bollen, K. A. (1989). ***Structural equations with latent variables***, New York: Wiley.

[4] Brocka, B., & Brocka, M. S. (1992), ***Quality management: implementing the best ideas of the masters***. Homewoodm IL: Irwin.

[5] Byrne, B.M. (1998). ***Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming***, LEA.

[6] Cheung, G. W. (1999). Multifaceted conceptions of self-other ratings disagreement. ***Personnel Psychology***, 52, 1-36.

[7] Cheung, G. W. (2002). ***Cross-cultural comparison of construct means: the issue of partial measurement invariance***. paper presented at the 2002 Academy of Management Annual Meeting at Denver, Colorado, USA.

[8] Cheung, G. W. & Rensvold, R. B. (1999). Testing factorial variance across groups: a reconceptualization and proposed new method. ***Journal of Management***, Vol. 25, No.1, 1-27.

[9] Cheung, G. W. & Rensvold, R. B. (2000). Assessing extreme an acquiescence response sets in cross-cultural research using structural equations modeling. ***Journal of Cross-cultural Psychology***, Vol. 31 No. 2, March,187-212.

[10] Dean, J. W., Jr. & Bowen, D. E. (1994). Management theory and total quality improving research and practice through theory development. ***Academy of Management Review***, Vol. 19 No. 3, 324-428.

[11]Deming, W. E. (1986), "***Out of crisis***", Cambridge: Massachusetts Institute of Technology, Center of Advanced Engineering.

[12]DiMaggio, P. J., & Powell, W. W. (1983), "The iron cage revisited: institutional isomorphism and collective rationality in organizational field", ***American Sociological Review***,Vol.48; April 147-160.

[13] Edwards JR. (1993). Problems with the use of profile similarity indices in the study of congruence in organizational research, ***Personnel Psychology***, 46, 641-665.

[14]Flynn, B. B., Schroeder, R. G., & Sakakibara, S. (Sept/Oct. 1995). The impact of quality management practices on performance and competitive advantage. ***Decision Sciences*** Vol. 26, No.5, 659-691.

[15]Greenwood, R., & Hinings, C. R. (1996), "Understanding radical organizational change: bringing together the old and the new institutionalism", ***Academy of Management Review***, Vol. 21 No. 4, pp. 1022-1054.

[16] Harris M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. ***Personnel Psychology***, 41, 43-61.

[17] Holzbach, R. L. (1978). Rater bias in performance ratings: superior staff, self-, and peer ratings. ***Journal of Applied Psychology***, 63, 579-588.

[18] Horn, J. L., & Mchrdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. ***Experimental Aging Research,*** 18, 117-144.

[19] Hui, C. H., & Triandis, H. C. (1985). Measurement in cross-cultural psychology: A review and comparison of strategies. ***Journal of Cross-Cultural Psychology***, 16(June), 131-152.

[20] Jöreskog, K. G. (1971), Simultaneous factor analysis in several populations. ***Psychometrika***, 36 (Dec), 409-426.

[21] Jöreskog, K. G., & Sörbom D. (1993). LISREL 8: Structural equation modeling with the SIMPLIS command language. Chicago: Scientific Software International.

[22] Lau, R. S. M., Zhao, X. and Xiao, M. (2004). Assessing quality management in China with MBNQA criteria. ***International Journal of Quality and Reliability Management***, 21(7), 699-713.

[23] Murphy, K. R., & Cleveland, J. N. (1995). ***Understanding performance appraisal: social, organizational, and goal-based perspectives***. Thousand Oaks, CA: Sage.

[24] Myers, M. B., Calantone, R. L., Page, T. J. Jr., & Taylor, C. R. (2000). Academic insights: An application of multiple-group causal models in assessing cross-cultural measurement equivalence. ***Journal of International Marketing***, Vol. 8, No. 4, 108-121.

[25] Nunnally, J. C., & Bernstein, I. H. (1994). ***Psychometric theory***. 3rd ed., New York: McGrawHill.

[26] Samson, D., & Terziovski, M. (1999). The relationship between total quality management practices and operational performance. ***Journal of Operations Management*** 17, 393-409.

[27] Saraph, Jayant V.; Benson, P. G., & Schroeder, Roger G. (1989). An instrument for measuring the critical factors of quality management. ***Decision Sciences***, Vol. 20, 810-829.

[28] Schriesheim, C. A., & Neider, L. L. (2001). Equivalence in measurement. ***Research in Management***, Vol. 1. Connecticut: Information Age Publishing.

[29] Sekaran, U. (1992), "***Research methods for business***". Wiley 2nd Ed.

[30] Sitkin, S. B., Sutcliffe, K. M., & Schroeder, R. G. (1994), Distinguishing control from learning in total quality management: a contingency perspective. ***Academy of Management Review***, Vol. 19 No. 3.

[31] Steele, R. P., & Ovale, N. K. (1984). Self-appraisal based upon supervisory feedback. ***Personnel Psychology***, 37, 667-685.

[32] Steenkamp, J-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. ***Journal of Consumer Research***, Vol. 25, June.

[33] Tsui, A. S., & Ohlott, P. (1988). Multiple assessment of managerial effectiveness: Interrater agreement and consensus in effectiveness models. ***Personnel Psychology***, 41, 779-803.

[34]Vandenberg, R. J., & Lance, C. E. (2000). A review of synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. ***Organizational Research Methods***, Vol. 3 No.1 January, 4-69.

[35] Wilson, Darryl D., & Collier, David A. (2000). An empirical investigation of the Malcolm Baldrige National Quality Award Causal Model. ***Decision Sciences*** Vol.31(Spring), No.2., 361-390.

[36] Zalesny, M. D., & Kirsch, M. P. (1989). The effect of similarity on performance ratings and interrater agreement. ***Human Relations***, 42, 81-96.