Chapter 5 Measurement Evaluation[15]

5.1 Introduction

In this study, a research questionnaire was developed and used to obtain empirical data from Chinese manufacturing firms in order to test the theoretical models hypothesized in this study. In the questionnaire, there were two measurement instruments used to measure TQM implementation and overall business performance, respectively. Each instrument had some measurement scales (see Appendices 1 and 2). Before testing the theoretical models, it was necessary to first evaluate the reliability and validity of the instruments; it is only on the basis of reliable and valid measurement scales that hypothesis testing can be conducted. Section 5.2 provides methods for empirically testing and validating the reliability and validity of the measurement scales. Section 5.3 presents the results of testing and validating the reliability and validity of the TQM implementation instrument. Section 5.4 involves the evaluation of measurement instrument of overall business performance. Finally, a number of conclusions are presented in Section 5.5.

5.2 Methodology

5.2.1 Reliability

Reliability refers to whether you get the same answer by using an instrument to measure something more than once (Bernard, 2000). Reliability concerns the extent to which an experiment, test, or any measuring procedure yields the same results in repeated trials (Carmines and Zeller, 1979); it is a statistical measure of how reproducible the survey instrument’s data are (Litwin, 1995). There are four methods commonly used for assessing reliability, namely, (1) the test-retest method, (2) the alternate-form method, (3) the split-halves method, and (4) the internal consistency method (Nunnally, 1967).

Test-retest reliability is measured by having the same set of respondents complete a survey at two different times to see how stable the responses are. It is a measure of how reproducible a set of results is. Correlation coefficients are then calculated to compare the two sets of responses. These correlation coefficients are collectively referred to as the survey instrument’s test-retest reliability. In general, if correlation coefficients equal or exceed 0.70, it is considered that the test-retest reliability is good (Litwin, 1995).

Alternate-form reliability is a method of evaluating the reliability of a survey instrument. It involves employing differently worded items to measure the same attribute. Questions and responses are reworded, or their order changed, to produce two items that are similar but not identical. Items are only different in their wording. Items or scales are administered to the

---

[15] Parts of this chapter were published in Zhang (1999b) and Zhang et al. (2000).

same population at different time points. In the same way, correlation coefficients are calculated. If they are high, the survey instrument is considered to have good alternate-form reliability (Litwin, 1995).

The split-halves method is a way to evaluate the reliability of the survey instrument. To use the split-halves method, the sample should be large enough to be divided in half and each alternate form administered to half of the group. Results from the two halves are then compared. When the split-halves method is used, the half-samples should be randomly selected. To do so, it can be ensured that no group differences exist (Litwin, 1995).

Internal consistency reliability is a commonly used psychometric measure in assessing survey instruments and scales. Internal consistency is an indicator of how well the different items measure the same concept. This is important because a group of items that purports to measure one variable should indeed be clearly focused on that variable. Internal consistency is measured by calculating a statistic known as Crobach’s coefficient alpha (Cronbach, 1951; Nunnally, 1967). Coefficient alpha measures internal consistency reliability among a group of items combined to form a single scale. It is a statistic that reflects the homogeneity of the scale. Generally, reliability coefficients of 0.70 or more are considered good (Nunnally, 1967).

Among the four methods mentioned above, it is clear that the first three have some limitations, particularly for field studies. These limitations include, for example, requiring two independent administrations of the instrument on the same group of people and requiring two alternate forms of the measuring instrument. In contrast, the internal consistency method does not require either the splitting or repeating of items. Instead, it requires only a single test administration and provides a unique estimation of reliability for the given test administration. It is the most general form of reliability estimation (Nunnally, 1967). Therefore, the internal consistency method was used in evaluating the reliability of the survey instruments in this research.

5.2.2 Item Analysis

Nunnally (1967) developed a method of evaluating the assignment of items to scales that considers the correlation of each item with each scale. Specifically, the item-score to scale-score correlations are used to determine whether an item belongs to the scale as assigned, to some other scales, or should be eliminated. The scale-score is obtained by computing the arithmetic average of the scores of the items that comprise that scale. The values of item to scale correlations should be greater than 0.50; those lower than 0.50 do not share enough variance with the rest of the items in that scale. Therefore, it is assumed that the items are not measuring the same construct and it should be deleted from the scale (Kemp, 1999). Saraph et al. (1989) used this method to evaluate the assignment of items to scales in developing their instrument for measuring the critical factors of quality management. It was judged that item analysis should be performed in order to understand whether items were assigned appropriately.

5.2.3 Validity

Validity is defined as the extent to which any instrument measures what it is intended to measure. The three most popular methods of evaluating the validity of a measurement instrument are content validity, criterion-related validity, and construct validity (Carmines and Zeller, 1979). However, due to limitations of some instruments that are known to be valid, many researchers did not evaluate the criterion-related validity of their instruments (e.g., De Jong, 1999; Kemp, 1999). In this study, only content validity and construct validity were conducted in order to evaluate the measurement instruments.

Content Validity

Content validity depends on the extent to which an empirical measurement reflects a specific domain of content. It cannot be evaluated numerically-it is a subjective measure of how appropriate the items seem to various reviewers with some knowledge of the subject matter. The evaluation of content validity typically involves an organized review of the survey's contents to ensure that it includes everything it should, and does not include anything it should not. Strictly speaking, content validity is not a highly scientific measure of a survey instrument's accuracy. Nevertheless, it provides a solid foundation on which to build a methodologically rigorous assessment of a survey instrument's validity. In this research, however, it was argued that the 11 scales for measuring TQM implementation constructs and the four scales for measuring overall business performance had content validity since the development of these measurement items was based mainly on an extensive review of the literature and detailed evaluations by academicians and practitioners. The references list the literature reviewed by the author during the period of conducting this research, and the research methodology section addresses the detailed process of developing the research questionnaire.

Construct Validity

Construct validity measures the extent to which the items in a scale all measure the same construct (Flynn et al., 1994), and can be evaluated by the use of factor analysis. Factor analysis addresses the issue of analyzing the interrelationships among a large number of items[16] and then explaining these items in terms of their common underlying dimensions (factors). In fact, the general purpose of factor analysis is to find a way of condensing or summarizing the information into a smaller set of new composite dimensions (factors) with a minimum loss of information (Hair et. al., 1992). There are two forms of factor analysis, namely, exploratory factor analysis and confirmatory factor analysis. According to Hair et al. (1992), there is continued debate concerning the appropriate role of factor analysis. Many researchers consider it only exploratory, useful in searching for structure among a set of variables, or as a data reduction method. In this study, two instruments were developed in order to measure TQM implementation and overall business performance, respectively. These instruments had never been used before. Therefore, factor analysis in this context was

---

[16] In some literature, the term "variable" is often used. In this context, the term "item" is the same as "variable".

exploratory in nature rather than confirmatory; thus, exploratory factor analysis was adopted in this study.

According to Hair et al. (1992), there are two methods of exploratory factor analysis: Principal component analysis and common factor analysis. Principal component analysis is appropriate when researchers are primarily concerned about the minimum number of factors needed to account for the maximum portion of the variance represented in the original set of items. In contrast, common factor analysis is appropriate when the primary objective is to identify the latent dimensions or constructs represented in the original items. According to the aim of conducting factor analysis in this study, principal component analysis was selected as it can determine how and to what extent items are linked to their underlying factors (Byrne, 1998). Principal component analysis can help to identify whether selected items cluster on one or more than one factor. Particularly, three or more items are selected for measuring a latent construct. Factor loadings are used to present these relations. Factor loadings greater than 0.30 are considered significant; loadings of 0.40 are considered more important; if the loadings are 0.50 or greater, they are considered very significant (Hair et. al, 1992). In this study, a factor loading of 0.50 was used as the usual cut-off point.

According to Hair et al. (1992), the most commonly used method of determining whether items are loading on one factor is the latent root criterion. Only the factors having latent (eigenvalues) greater than 1 are considered significant; those with eigenvalues less than 1 are considered insignificant and are disregarded.

5.3 TQM Implementation Instrument

5.3.1 Reliability

There were 11 scales for measuring the 11 TQM implementation constructs for Chinese manufacturing firms. For each scale, there were a number of items to measure it (see Appendix 1). Appendix 3 lists the relative frequency distributions and means of respondents' responses to items that measure TQM implementation constructs. After all data were entered into a computer, the SPSS 7.5.2 reliability program was performed separately for the items of each scale. Table 5.1 lists Cronbach's alpha for different TQM implementation scales. This table shows that the reliability coefficients ranged from 0.838 to 0.925, indicating that some scales were more reliable than others. Accordingly, the instrument developed for measuring TQM implementation constructs was judged to be reliable.

Table 5.1 Internal Consistency Analysis

Scales	Number of items	Cronbach's alpha
1. Leadership	8	0.892
2. Supplier quality management	6	0.838
3. Vision and plan statement	8	0.914
4. Evaluation	10	0.890
5. Process control and improvement	8	0.883
6. Product design	8	0.839
7. Quality system improvement	5	0.925
8. Employee participation	8	0.883
9. Recognition and reward	6	0.857
10. Education and training	6	0.885
11. Customer focus	6	0.875

5.3.2 Item Analysis

Table 5.2 lists the correlation matrix for the 11 scales of the TQM implementation (Scales 1-11) and their measurement items. This table shows that all values of item to scale correlations were greater than 0.50. The complete correlation matrix presented in Appendix 4 further shows the items correlated highly with the scales they intended to measure (shown in bold in Appendix 4). For instance, Item 1 in Scale 1 had correlations of 0.784, 0.581, 0.631, 0.587, 0.555, 0.560, 0.525, 0.559, 0.466, 0.540, and 0.538 with the 11 scales of the TQM implementation, respectively. Since the value of Scale 1 (Leadership) was the average of the eight items, the high correlation between Scale 1 and its Item 1 was expected. In addition, since Item 1 showed relatively smaller correlations with the other scales, it was concluded that Item 1 in Scale 1 had been assigned appropriately to this scale. All other items were similarly examined.

Table 5.2 Item to Scale Correlation Matrix (Pearson Correlation)

	Item number									
Scales	1	2	3	4	5	6	7	8	9	10
Scale 1	.784	.810	.709	.702	.851	.624	.751	.816	--	--
Scale 2	.741	.787	.816	.690	.683	.753	--	--	--	--
Scale 3	.844	.861	.553	.861	.859	.863	.652	.823	--	--
Scale 4	.741	.738	.768	.703	.656	.772	.769	.627	.653	.692
Scale 5	.725	.666	.755	.713	.800	.754	.790	.721	--	--
Scale 6	.688	.649	.642	.726	.636	.570	.802	.765	--	--
Scale 7	.782	.913	.914	.939	.844	--	--	--	--	--
Scale 8	.765	.760	.804	.698	.741	.777	.639	.761	--	--
Scale 9	.756	.718	.801	.766	.745	.800	--	--	--	--
Scale 10	.828	.863	.846	.680	.703	.854	--	--	--	--
Scale 11	.813	.737	.838	.849	.649	.813	--	--	--	--

Notes: Item number in this table is the same as the item number in the instrument.
The symbol "--" means not available.

From the figures in Appendix 4, it was obvious that all items had relatively high correlations with the scales to which they were originally assigned, compared with all the other scales. Therefore, it was concluded that all items had been appropriately assigned to scales.

5.3.3 Construct Validity

The program of SPSS 7.5.2 was used to perform factor analysis, each scale being factor analyzed separately. The detailed results are listed in Table 5.3. From this table, it was clear that all of the items had high factor loadings greater than 0.50 on Factor 1. When the items in a scale loaded on more than one factor, the rotated (varimax, quartimax if necessary) solution was examined. The factor analysis showed that the items in 9 of the 11 scales formed a single factor, except for Scales 4 (Evaluation) and 5 (Process control and improvement). In the cases of these scales, two factors emerged according to the rule that the eigenvalues are greater than 1, which are listed in Tables 5.4 and 5.5. According to Hair et al. (1992), the latent root criterion (eigenvalue) is the most commonly used method of judging whether items are loading on one factor. In principal component analysis, only the factors having eigenvalues greater than 1 are considered significant; all factors with eigenvalues less than 1 are considered insignificant and disregarded. It should be noted that percentage of variance and scree test can also be used as criteria for judging whether items in a scale load on one factor. However, these two criteria are not easy to use in practice. For example, in the social sciences, where information is often less precise, it is not uncommon for the analyst to consider a solution that accounts for 60% of the total variance (and in some instances even less) as a satisfactory solution (Hair et. al., 1992). Further, it is not easy to make the judgement based on the scree test.

Table 5.6 lists the unrotated factor and rotated factor matrix for Scale 4 (Evaluation). From this table, it was clear that Items 5, 6, 7, and 8 constituted a factor, which can be interpreted as the factor "Use of quality-related information". The other items in Scale 4 were in the other factor, which formed that of "Audit". Therefore, the construct of evaluation has two dimensions, namely, audit (Items 1, 2, 3, 4, 9, and 10) and use of quality-related information (Items 5, 6, 7, and 8).

Table 5.7 lists the unrotated factor and rotated (varimax and quartimax) factor matrix for Scale 5 (Process control and improvement). After orthogonal and oblique factor rotation was done, it was not easy to decide whether Item 5 belonged to Factors 1 or 2 as it loaded very significantly on both factors. After its content was examined, it was decided that Item 5 (Our processes are designed to be "foolproof" in order to minimize the chance of employee error) should be deleted from this scale. Thus, Items 1, 2, 3, and 4 formed a factor that can be interpreted as that of "Process control". Similarly, Items 6, 7, and 8 constituted a factor that can be interpreted as that of "Use of quality management methods". Therefore, the construct of process control and improvement has two dimensions, namely, process control (Items 1, 2, 3, and 4) and use of quality management methods (Items 6, 7, and 8).

Table 5.3 Results of Exploratory Factor Analysis for the Eleven TQM Implementation Scales

Scales	Factor number	Eigenvalues	Factor loadings										% of variance
			Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10	
1	1	4.620	0.784	0.802	0.725	0.677	0.839	0.644	0.770	0.816			57.750
2	1	3.394	0.748	0.808	0.817	0.696	0.665	0.737					55.820
3	1	5.095	0.848	0.861	0.531	0.875	0.873	0.873	0.630	0.815			63.692
4	2	5.106 (Factor 1)	0.737	0.679	0.731	0.751	0.701	0.671	0.797	0.785	0.647	0.627	51.062
		1.009 (Factor 2)	0.223	0.170	0.269	0.020	-0.094	-0.379	-0.294	-0.589	0.273	0.444	10.095
5	2	4.405 (Factor 1)	0.727	0.677	0.765	0.725	0.793	0.751	0.777	0.714			55.060
		1.019 (Factor 2)	0.313	0.492	0.395	0.138	-0.080	-0.408	-0.333	-0.469			12.734
6	1	3.797	0.709	0.665	0.661	0.712	0.637	0.572	0.789	0.745			47.469
7	1	3.878	0.795	0.904	0.902	0.940	0.855						77.561
8	1	4.443	0.754	0.740	0.806	0.705	0.754	0.781	0.637	0.773			55.539
9	1	3.514	0.761	0.699	0.759	0.776	0.747	0.810					58.572
10	1	3.834	0.833	0.858	0.844	0.674	0.718	0.848					63.895
11	1	3.710	0.819	0.748	0.825	0.842	0.657	0.813					61.883

Note: Eigenvalue greater than 1 was used as criterion for factor extraction.

Table 5.4 Factor Extraction (Evaluation)

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5,106	51,062	51,062	5,106	51,062	51,062	3,322	33,222	33,222
2	1,009	10,095	61,157	1,009	10,095	61,157	2,794	27,935	61,157
3	,911	9,107	70,263						
4	,587	5,873	76,137						
5	,552	5,518	81,654						
6	,473	4,727	86,381						
7	,424	4,244	90,625						
8	,359	3,590	94,215						
9	,318	3,181	97,396						
10	,260	2,604	100,000						

Extraction Method: Principal Component Analysis.

Table 5.5 Factor Extraction (Process Control and Improvement)

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,405	55,060	55,060	4,405	55,060	55,060	2,722	34,025	34,025
2	1,019	12,734	67,794	1,019	12,734	67,794	2,702	33,769	67,794
3	,638	7,970	75,764						
4	,511	6,389	82,153						
5	,451	5,643	87,796						
6	,357	4,459	92,256						
7	,346	4,324	96,579						
8	,274	3,421	100,000						

Extraction Method: Principal Component Analysis.

Table 5.6 Factor Matrix for Scale 4 (Evaluation)

	Unrotated factor		Rotated factor (Varimax)	
Scale 4	1	2	1	2
Item 1	.737	.223	.701	.319
Item 2	.731	.170	.661	.354
Item 3	.751	.269	.742	.294
Item 4	.701	.020	.540	.447
Item 5	.671	-.094	.442	.513
Item 6	.797	-.379	.348	.811
Item 7	.785	-.294	.395	.739
Item 8	.647	-.589	.098	.870
Item 9	.627	.273	.652	.209
Item 10	.679	.444	.803	.114

Table 5.7 Factor Matrix for Scale 5 (Process Control and Improvement)

	Unrotated factor		Rotated factor (Varimax)		Rotated factor (Quartimax)	
Scale 5	1	2	1	2	1	2
Item 1	.727	.313	.295	.734	.753	.245
Item 2	.677	.492	.134	.826	.833	.077
Item 3	.765	.395	.264	.820	.836	.208
Item 4	.725	.138	.417	.609	.636	.375
Item 5	.793	-.080	.619	.503	.544	.583
Item 6	.751	-.408	.820	.240	.295	.802
Item 7	.777	-.333	.786	.311	.364	.763
Item 8	.714	-.469	.837	.171	.228	.823

It should be noted that two factors emerged for Scales 4 (Evaluation) and 5 (Process control and improvement). This result was obtained according to the rule of eigenvalues greater than 1. A careful examination of the two eigenvalues for the two scales was undertaken; the two eigenvalues for Factor 2 were only 1.009 and 1.019, just a little bit greater than 1. After the contents of items in Scales 4 and 5 were carefully examined, it was clear that the aim of using quality-related information is for evaluation. Similarly, the aim of using quality management methods is for controlling and improving process. Therefore, for subsequent data analysis, the construct of evaluation was not divided into two constructs. Similarly, neither was the construct of process control and improvement. Note that Item 5 in Scale 5 was deleted. Thus, Scale 5 consisted of 7 measurement items.

5.3.4 Summary

After the reliability analysis, item analysis, and validity analysis had been conducted, it was concluded that the TQM implementation instrument is reliable and valid. The data obtained through this instrument can be used for subsequent data analysis. The tested and validated TQM implementation instrument had 11 scales that consisted of 78 measurement items. One item in Scale 5 was deleted after factor analysis.

Compared to the other quality management instruments developed by Saraph et al. (1989), Flynn et al. (1994), and Ahire et al. (1996), the instrument developed in this study has the highest external validity for manufacturing industries in general and for Chinese manufacturing firms in particular. The reason for this is that the author used data from 212 Chinese manufacturing firms in nine industrial sectors for testing and validating this instrument. While its internal consistency falls behind the Flynn et al. and the Ahire et al. instruments, it is however better than that of Saraph et al. Saraph et al. (1989) used data from 162 general and quality managers in 89 divisions of 20 manufacturing and service firms. A main strength of their instrument is its highest level of external validity for manufacturing and service industries (Ahire et al., 1996). Flynn et al. (1994) employed data from 716 respondents at 42 plants in the transportation components, electronics and machinery industries. The focus of their instrument is more on manufacturing industries within limited sectors. Their instrument, therefore, has the second highest internal validity. Its external validity is, however, better than the Ahire et al. instrument. Ahire et al. (1996) utilized data from 371 manufacturing firms in a single industry–motor vehicle parts and accessories for validating their instrument. Their instrument therefore has the highest internal consistency, but its external validity is the lowest. Ahire et al. (1996) summarized the differences and similarities of the three instruments in their paper.

5.4 Overall Business Performance Instrument

In this study, four constructs of overall business performance were identified: Employee satisfaction, product quality, customer satisfaction, and strategic business performance. These constructs were measured by one, seven, two, and five items, respectively. For details, please refer to Appendix 2. For testing and validating the four scales of overall business performance, internal consistency analysis, item analysis, content validity, and exploratory factor analysis should be conducted. In fact, content validity has already been addressed in Subsection 5.2.3. It was assumed that the four scales for measuring overall business performance had content validity. In this section, however, only internal consistency analysis, item analysis, and exploratory analysis are presented. The detailed analysis results are listed in Tables 5.8, 5.9, and 5.10, respectively. Appendix 3 lists the relative frequency distributions and means of respondents' responses to items that measure overall business performance constructs.

5.4.1 Employee Satisfaction

In this study, only one item was used to measure perceived overall employee satisfaction. Therefore, it was not necessary to conduct internal consistency reliability analysis, item analysis, and factor analysis for this scale; it can be assumed that it is reliable and valid. For details, please refer to Chapter 3.

5.4.2 Product Quality

Table 5.8 shows that the reliability alpha was 0.931, which indicated that this scale is reliable. The coefficients of item to scale correlation were greater than 0.50, which indicated that the items had been assigned to the scale appropriately. Table 5.9 shows the results of exploratory factor analysis; it can be concluded that the construct of product quality had two dimensions according to the rule of eigenvalues greater than 1. Table 5.10 shows the results of rotated exploratory factor analysis. It was evident that Items 1, 2, 3, and 4 formed a factor, which could be interpreted as that of "Product performance". Similarly, Items 5, 6, and 7 formed another factor that could be interpreted as "Quality loss". Thus, the construct of product quality has two dimensions: Product performance and quality loss. However, Table 5.9 shows that all factor loadings on Factor 1 were greater than 0.50, while_factor loadings on Factor 2 were less than 0.50. In addition, the eigenvalue for Factor 2 was slightly greater than 1. Therefore, for subsequent data analysis, the construct of product quality was not divided into two constructs. In fact, this construct is a large concept that consists of the dimensions of product performance and quality loss. Thus, the scale for measuring product quality has good construct validity.

5.4.3 Customer Satisfaction

There were only two items for measuring the scale of customer satisfaction. Table 5.8 shows that the reliability alpha was 0.868, which indicates this scale is reliable. Table 5.8 also shows that the values of item to scale correlation were greater than 0.50. Table 5.9 suggests that factor loadings were greater than 0.50. Therefore, it can be concluded that this scale for measuring customer satisfaction is reliable and valid and can be used for subsequent data analysis.

5.4.4 Strategic Business Performance

Although the Cronbach's alpha for the scale of strategic business performance was 0.642 (see Table 5.8), it is common practice to consider 0.60 an acceptable value in management science research (Nunnaly, 1978). Therefore, this scale for measuring strategic business performance is reliable. According to item analysis, the coefficients of item to scale correlation were greater than 0.50 (see Table 5.8), which indicates that the items had been assigned to the scale appropriately. Table 5.9 shows the results of exploratory factor analysis. It can be concluded that the construct of strategic business performance has two dimensions according to the rule of eigenvalues greater than 1. It was clear that all factor loadings on Factor 1 were greater than those on Factor 2. However, the absolute values of the three factor loadings on Factor 2 were greater than 0.50. This result further indicated that this construct had two dimensions. Table 5.10 shows that Items 1, 4, and 5 formed a dimension that could be interpreted as the factor of "Firm size". Items 2 and 3 formed another dimension, which can be interpreted as the factor of "Profitability". These two dimensions of strategic business performance receive a great deal attention by the Chinese manufacturing firms and the Chinese government. In this study, the construct of strategic business performance was regarded as a complete concept. Therefore, for subsequent data analysis, this construct of overall business performance was not divided into two constructs.

Table 5.8 Reliability Analysis and Item Analysis

Scales	Items	CA	Item-total correlation						
			Item1	Item2	Item3	Item4	Item5	Item6	Item7
PQ	7	0.931	0.835	0.885	0.870	0.837	0.792	0.844	0.849
CS	2	0.868	0.936	0.945					
SBP	5	0.642	0.716	0.637	0.660	0.633	0.564		

Notes: CA means Cronbach's alpha; PQ for product quality; CS for customer satisfaction; SBP for strategic business performance.

Table 5.9 Unrotated Exploratory Factor Analysis

Scales	EV	Factor loadings							% of variance
		Item1	Item2	Item3	Item4	Item5	Item6	Item7	
PQ	5.005	0.846	0.897	0.886	0.853	0.770	0.826	0.834	71.497
	1.055	-0.348	-0.271	-0.325	-0.342	0.487	0.473	0.421	15.078
CS	1.770	0.941	0.941						88.497
SBP	2.084	0.732	0.659	0.704	0.611	0.493			41.677
	1.153	0.345	-0.579	-0.533	0.525	0.373			23.065

Notes: EV means Eigenvalue and Eigenvalues greater than 1 were used as criterion for factor extraction.
PQ means product quality; CS for customer satisfaction; SBP for strategic business performance.

Table 5.10 Rotated Exploratory Factor Analysis (Varimax)

Scales	Factors	Rotated (Varimax) factor loadings						
		Item1	Item2	Item3	Item4	Item5	Item6	Item7
Product quality	Factor 1	0.870	0.859	0.886	0.872	0.271	0.324	0.363
	Factor 2	0.283	0.374	0.326	0.292	0.870	0.896	0.860
Strategic Business performance	Factor 1	0.770	0.09	0.152	0.805	0.615		
	Factor 2	0.247	0.873	0.870	0.03	0.06		

5.4.5 Summary

In summary, the instrument for measuring overall business performance is reliable and valid. The data obtained from this instrument can be used in subsequent data analysis to test the theoretical models hypothesized in this research.

5.5 Conclusions

Reliability, item, and validity analyses were used to test and validate the measurement instruments. The procedures for testing the reliability and validity of the instruments (TQM implementation and overall business performance) have been described in greater detail in this chapter. After reliability, item, and factor analyses were conducted, the instruments for measuring TQM implementation constructs and overall business performance constructs were empirically tested and validated for their reliability and validity. One item in the scale of process control and improvement was deleted after factor analysis since it loaded significantly on two factors. Thus, for subsequent data analysis, only 78 items were used for measuring TQM implementation. Through their evaluation, it can be concluded that the instruments for measuring TQM implementation and overall business performance are reliable and valid. The data obtained from the two instruments can be used for testing the theoretical models hypothesized in this study. Practitioners can use the validated instruments for measuring their TQM implementation and overall business performance, and researchers can use them to study the effects of TQM implementation on overall business performance.