CHAPTER 4: INSTRUMENT DEVELOPMENT AND ITEM GENERATION

This chapter presents the measuring scales either adopted from past literature or developed for this study. The scales to measure supplier integration, customer integration and internal integration were adopted from previous studies (Stank et al., 2001; Narasimhan and Kim, 2002; Flynn et al., 2010; Wong et al., 2011) with some minor modifications to suit the context. The instruments to measure supplier related performance outcome, customer related performance outcome and manufacturing related performance outcome were developed and used.

4.1 Instrument Development

The instrument development for supplier related performance outcome, customer related performance outcome and manufacturing related performance outcome constructs have been carried out in three stages. In the first stage, potential items were generated through an extensive literature review and from the construct definitions. The initial pool of items was pre-tested with two academicians and two practitioners. The respondents were requested to provide feedback about the overall clarity and readability of the items. Based on their feedback, items were either discarded or modified to ensure content validity.

In the second stage, the scales were validated using the Q - sort method (Stephenson, 1953). The items placed in a common pool were subject to two rounds of sorting by the judges into four categories. The objective was to pre-assess the convergent and discriminant validity of the scales by examining how the items are categorized into various construct categories. The analysis of inter-judgement scores identified both weak and non-relevant items.
In the third stage, reliability and validity of the instrument were tested using the data from a large scale survey. Third stage is described in detail in the next chapter.

4.2 Item Generation

The initial pool of items for supplier related performance outcome, customer related performance outcome and manufacturing related performance outcome was generated through extensive review of supply management, supply chain performance and supply chain integration literature (Frohlich et al., 2001; Flynn et al., 2010; Vickery et al. 2003; Wong et al. 2011; Droge et al. 2004; Stank et al. 2001; Li et al., 2009; Lee et al. 2007; Kannan 2010; Sezen 2008; Prajogo et al. 2012; Schoenherr et al. 2012; Swink et al. 2007; Narasimhan et al. 1998; Richey et al. 2009; Chen et al., 2004, Gunasekaran et al., 2004).

The initial pool of items generated through the literature review was reviewed by two academicians to check the clarity of the items and the relevance of the construct’s definition. Based on their feedback, some of the items were either deleted or modified. Some new items were also added wherever required necessary. Table 4.1 presents the list of the three constructs and their corresponding number of items. The details of items corresponding to constructs are presented in Appendix I.
<table>
<thead>
<tr>
<th>Construct</th>
<th>Number of items</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRPO</td>
<td>7</td>
</tr>
<tr>
<td>CRPO</td>
<td>9</td>
</tr>
<tr>
<td>MRPO</td>
<td>10</td>
</tr>
</tbody>
</table>

4.3 Scale development: Q-Sort Method

Q-sort method was first introduced by Stephenson (1953) to pre-assess the convergent and discriminant validity of an instrument. It is done by examining the instrument items that are sorted by knowledgeable people in the field of study into corresponding categories defined by the researchers. It is an iterative process in which the degree of agreement between respondents forms the basis of assessing construct validity and improving the reliability of the constructs (Nahm et al., 2002).

Two industry experts and two academicians were invited to participate in the sorting process. The industry experts involved in the sorting process held positions as production manager and supply chain manager in manufacturing firms. The academicians involved in the sorting process were professors of supply chain management in a management institute. All 26 items related to three different constructs were mixed in a common pool. Before the start of sorting process the respondents received an explanation of the meaning and the definition of each of the constructs. Then they were asked to sort the 26 items into four groups, one each corresponding to the three constructs and a non-applicable (NA) category. The non-applicable category was included to prevent the respondents from forcing the placement of any item into a
particular category. The sorting result was evaluated using inter-judge raw agreement score, Moore and Benbasat’s hit ratio (Moore and Benbasat, 1991) and Cohen’s Kappa (Cohen, 1960).

**Inter-judge raw agreement score:** It was calculated by counting the number of items that both the judges (one academician and one practitioner) agreed to place in one category. An item was treated as an item with agreement, if both the judges sorted the item in same category.

**Moore and Benbasat’s hit ratio (Moore and Benbasat, 1991):** Item placement ratios were calculated by counting all the items that were correctly sorted into the theoretical category by each of the judges and dividing this amount by twice the total number of items.

**Cohen’s Kappa (Cohen, 1960):** This index is a measure of agreements between two judges.

**4.4 Results of First Sorting Round**

In the first round, a total of $N$ items are independently classified into $K$ mutually exclusive categories by two judges.

Let the categories be denoted by $1, 2, \ldots, k$.

Let $X_{ij}$ be the number of items placed in category $i$ by Judge One and the same items are placed in category $j$ by Judge Two; $i, j = 1, 2, \ldots, k$ and $i \neq j$.

Let $X_{ii}$ denote the number of items placed in category $i$ by both the judges which indicates agreement.

A 2-way contingency table can be constructed as $(X_{ij})$.

Let $X_i$ denote total of row $i$; $i = 1, 2, \ldots, k$

Let $X_j$ denote total of column $j$; $j = 1, 2, \ldots, k$
The Cohen’s Kappa $K$ is calculated as

$$K = \frac{N \sum_{i=1}^{k} X_{ii} - \sum_{i=1}^{k} X_i \cdot X_i}{N^2 - \sum_{i=1}^{k} X_i \cdot X_i}$$

Table 4.2 gives the inter-judge raw agreement scores resulting from the first sorting round.

Let the row/column categories be represented as:

1: SRPO
2: CRPO
3: MRPO
4: NA- Not applicable

<table>
<thead>
<tr>
<th>JUDGE</th>
<th>1: SRPO</th>
<th>2: CRPO</th>
<th>3: MRPO</th>
<th>4: NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>JUDGE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1: SRPO</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2: CRPO</td>
<td></td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3: MRPO</td>
<td></td>
<td></td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>4: NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Total item placement: 26
No. of matched agreement: 21
Agreement ratio: 0.80

$$K = \frac{(26) \cdot (21) - 208}{208} = 0.72$$
The result of first round of sorting shows that the inter-judge raw agreement score was 80. The Cohen’s Kappa score averaged 72%. A Cohen’s Kappa score of 0.9 is considered excellent. However, a score greater than 0.65 has been considered acceptable by several researchers (Landis & Koch, 1977; Todd & Benbasat, 1993).

The item placement ratios of the first Q-sort round are presented in Table 4.3. As evident from the table the initial overall placement ratio of items within the target constructs was 82%.

| THEORETICAL | ACTUAL CATEGORIES | | | | |
|-------------|-------------------|---|---|---|---|---|
| 1: SRPO     | 14                |   |   |   | 14 | 100% |
| 2: CRPO     | 1 14 3           |   |   |   | 18 | 77%  |
| 3: MRPO     | 1 2 15 2         |   |   |   | 20 | 75%  |
| 4: NA       | 0 0 0 0          |   |   |   | 0  0% |

Table 4.3 Items Placement Ratios: First Sorting Round
In order to improve the Cohen’s kappa measure of agreement, the off-diagonal entries in the placement matrix (Table 4.3) was examined. The items classified in a construct different from the target constructs were identified and reworded. Two items were considered vague and classified into ‘not applicable category’ by judges, and these items were deleted. Further, feedback from both the judges was obtained on each item and incorporated. Finally, two items were deleted and four items were reworded after the first round of Q-sort.

4.5 Results of Second Sorting Round

Two judges who are different from the judges of first round were involved in the second round of Q-sort which included the modified items after the first sorting round. The procedure explained in first round of sorting was repeated to calculate inter-judge raw agreement score, Cohen’s Kappa and item placement ratios of the second round of Q-sort. The inter-judge raw agreement score and the item placement ratios of the second Q-sort round are presented in Table 4.4 and Table 4.5 respectively.
Table 4.4 Inter-Judge Raw Agreement Scores: Second Sorting Round

<table>
<thead>
<tr>
<th>JUDGE 1</th>
<th>1: SRPO</th>
<th>2: CRPO</th>
<th>3: MRPO</th>
<th>4: NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: SRPO</td>
<td></td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2: CRPO</td>
<td></td>
<td>8</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3: MRPO</td>
<td></td>
<td></td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>4: NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Total item placement: 24
No. of matched agreement: 22
Agreement ratio: 0.91

\[
K = \frac{(24) \times (22) - 194}{(24) \times (24) - 194} = 0.87
\]
The second round sorting yielded an excellent inter-judge raw agreement score of 91%. The initial overall placement ratio of items within the target constructs was 95%. The Cohen’s Kappa score averaged 87%. The value of Cohen’s kappa greater than 0.76 is considered sufficient (Nahm et al., 2002), hence, the Cohen’s kappa score of 87% indicated a satisfactory result. The above mentioned inter rater reliability scores indicated a high level of reliability and validity of
constructs. Hence, the Q-sort method was terminated after round two. The details of resulting measurement scales for SRPO, CRPO and MRPO are presented in Appendix II.

The survey instrument for this study was designed which included the newly developed measurement scales for SRPO, CRPO, MRPO and adopted scales for supply chain integration and financial performance constructs.

Apart from the items to measure the above mentioned constructs, the survey instrument is also designed to collect information about the characteristics of the respondents and organizations in which they are currently employed. The characteristics of respondents are in terms of their job designation and number of years in the current organization. While, the characteristics of respondent’s organizations are in terms of industry category and the number of employees in the organization. The survey instrument is presented in Appendix III. The next chapter presents the details of large scale survey and instrument validation methodology.