

5-1-2002

# The Q-Sort Method: Assessing Reliability And Construct Validity Of Questionnaire Items At A Pre-Testing Stage


Abraham Y. Nahm  
*University of Wisconsin*

S. Subba Rao  
*The University of Toledo*

Luis E. Solis-Galvan  
*Instituto de Empresa, Madrid, Spain*

T. S. Ragu-Nathan  
*The University of Toledo*

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

## Recommended Citation

Nahm, Abraham Y.; Rao, S. Subba; Solis-Galvan, Luis E.; and Ragu-Nathan, T. S. (2002) "The Q-Sort Method: Assessing Reliability And Construct Validity Of Questionnaire Items At A Pre-Testing Stage," *Journal of Modern Applied Statistical Methods*: Vol. 1: Iss. 1, Article 15.

DOI: [10.22237/jmasm/1020255360](https://doi.org/10.22237/jmasm/1020255360)

Available at: <http://digitalcommons.wayne.edu/jmasm/vol1/iss1/15>

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in *Journal of Modern Applied Statistical Methods* by an authorized administrator of DigitalCommons@WayneState.

## The Q-Sort Method: Assessing Reliability And Construct Validity Of Questionnaire Items At A Pre-Testing Stage

Abraham Y. Nahm  
 Management and Marketing Department  
 University of Wisconsin

S. Subba Rao  
 Department of Management  
 The University of Toledo

Luis E. Solis-Galvan  
 Operations and Technology Management Department  
 Instituto de Empresa

T. S. Ragu-Nathan  
 Department of Information Systems, Marketing, ECommerce & Sales  
 The University of Toledo

This paper describes the Q-sort, which is a method of assessing reliability and construct validity of questionnaire items at a pre-testing stage. The method uses Cohen's Kappa and Moore and Benbasat's Hit Ratio in assessing the questionnaire. Two examples are provided on how the method was applied in survey research.

**Keywords:** Q-sort, Surveys, Reliability, Construct validity, Kappa, Hit ratio

### Introduction

This paper describes the Q-sort method, which is a method of assessing reliability and construct validity of questionnaire items that are being prepared for survey research. The method is applied at the pre-test stage, which comes after the item generation through literature search and before the administering of questionnaire items as a survey. The method is cost efficient and simple, yet provides ample insight into potential problem areas in the questionnaire items that are being tested.

### Theoretical Basis

The Q-sort method is an iterative process in which the degree of agreement between judges forms the basis of assessing construct validity and improving the reliability of the constructs. The method consists of two stages. In the first stage, two judges are requested to sort the questionnaire items according to different constructs, based on which the inter-judge agreement is measured. In the second stage, questionnaire items that were identified as being too ambiguous, as a result of the first stage, are reworded or deleted, in an effort to improve the agreement between the judges. The process is carried out repeatedly until a satisfactory level of agreement is reached.

---

Abraham Y. Nahm, Management and Marketing Department, University of Wisconsin – Eau Claire, P.O. Box 4004, Eau Claire, Wisconsin, 54702-4004. E-mail: nahmay@uwec.edu. Luis E. Solis-Galvan, Operations and Technology Management Department, Instituto de Empresa, Madrid, Spain. S. Subba Rao, Department of Management, The University of Toledo, Toledo, Ohio. T. S. Ragu-Nathan, Department of Information Systems, Marketing, ECommerce and Sales, The University of Toledo, Toledo, Ohio.

The following example describes the theoretical basis for the Q-sort method and the two evaluation indices to measure inter-judge agreement level: Cohen's Kappa (Cohen, 1960) and Moore and Benbasat's "Hit Ratio" (Moore and Benbasat, 1991).

Assume two judges independently classified a set of N components as either acceptable or rejectable. After the work was finished the following table was constructed:

		Judge 1		Totals
		Acceptable	Rejectable	
Judge 2	Acceptable	$X_{11}$	$X_{12}$	$X_{1+}$
	Rejectable	$X_{21}$	$X_{22}$	$X_{2+}$
Totals		$X_{+1}$	$X_{+2}$	N

*Note:*  $X_{ij}$  = the number of components in the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column, for  $i, j = 1, 2$ .

The above table can also be constructed using percentages by dividing each numerical entry by N. For the population of components, the table will look like:

		Judge 1		Totals
		Acceptable	Rejectable	
Judge 2	Acceptable	$P_{11}$	$P_{12}$	$P_{1+}$
	Rejectable	$P_{21}$	$P_{22}$	$P_{2+}$
Totals		$P_{+1}$	$P_{+2}$	100

*Note:*  $P_{ij}$  = the percentage of components in the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column.

This table of percentages is used to describe the Cohen's Kappa coefficient of agreement. The simplest measure of agreement is the proportion of components that were classified the same by both judges, i.e.,  $\sum_i P_{ii} = P_{11} + P_{22}$ . However, Cohen suggested comparing the actual agreement,  $\sum_i P_{ii}$ , with the chance of agreement that would occur if the row and columns are independent, i.e.,  $\sum_i P_{i+} P_{+i}$ . The difference between the actual and chance agreements,  $\sum_i P_{ii} - \sum_i P_{i+} P_{+i}$ , is the percent agreement above which is due to chance. This difference can be standardized by dividing it by its maximum possible value, i.e.,  $100\% - \sum_i P_{i+} P_{+i}$ . The ratio of these is denoted by the Greek letter Kappa and is referred to as Cohen's Kappa.

Thus, Cohen's Kappa as a measure of agreement can be interpreted as the proportion of joint judgement in which there is agreement after chance agreement is excluded. The three basic assumptions for this agreement coefficient are: 1) the units are independent, 2) the categories of the nominal scale are independent and mutually exclusive, and 3) the judges operate independently.

For Kappa, no general agreement exists with respect to required scores. However, several studies have considered scores greater than 0.65 to be acceptable (e.g. Vessey, 1984; Jarvenpaa 1989). Landis and Koch (1977) have provided a more detailed guideline to interpret Kappa by associating different values of this index to the degree of agreement beyond chance. They suggest the following guideline:

1. Excellent Agreement: Kappa = .76 - 1.00
2. Fair To Moderate Agreement: Kappa = .40 - .75
3. Poor Agreement: Kappa = .39 or less

A second overall measure of both the reliability of the classification scheme and the validity of the items was developed by Moore and Benbasat (1991). The method required analysis of how many items were placed by the panel of judges for each round within the target construct. In other words, because each item was included in the pool explicitly to measure a particular underlying construct, a measurement was taken of the overall frequency with which the judges placed items within the intended theoretical construct. The higher the percentage of items placed in the target construct, the higher the degree of inter-judge agreement across the panel which must have occurred.

Scales based on categories which have a high degree of correct placement of items within them can be considered to have a high degree of construct validity, with a high potential for good reliability scores. This procedure is more a qualitative analysis than a rigorous quantitative procedure. There are no established guidelines for determining good levels of placement, but the matrix can be used to highlight any potential problem areas. The following exemplifies how this measure works.

Consider the simple case of four theoretical constructs with ten items developed for each construct. With a panel of three judges, a theoretical total of 30 placements could be made within each construct. Thereby, a theoretical versus actual matrix of item placements could be created as shown in the table below (including an ACTUAL "N/A: Not Applicable" column where judges could place items which they felt fit none of the categories). (See Table A on next page.)

The item placement ratio (the "Hit Ratio") is an indicator of how many items were placed in the intended, or target, category by the judges. Examination of the diagonal of the matrix shows that with a theoretical maximum of 120 target placements (four constructs at 30 placements per construct), a total of 102 "hits" were achieved, for an overall "hit ratio" of 85%. More important, an examination of each row shows how the items created to tap the particular constructs are actually being classified. For example, row C shows that all 30-item placements were within the target construct, but that in row B, only 60% (18/30) were within the target. In the latter case, 8 of the placements were made in construct A, which might indicate the items underlying these placements are not differentiated enough from the items created for construct A. This finding would lead one to have confidence in a scale based on row C, but be hesitant about accepting any scale based on row B. An examination of off-diagonal entries indicates how complex any construct might be. Actual constructs based on columns with a high number of entries in the off-diagonal might be considered too ambiguous, so any consistent pattern of item misclassification should be examined.

### Examples

#### Research on Post-industrial Manufacturing

Nahm (2000) studied the relationship between external environment, internal environment ("managerial beliefs and attitudes" and "organizational structure"), and manufacturing practices through a large-scale survey. As part of this research, questionnaire items were developed to measure the constructs "external environment," "managerial beliefs and attitudes," and "organizational structure" (items for "manufacturing practices" were adopted from Koufteros, Vonderembse, & Doll, 1998). The development of the instrument was carried out in two steps. The first step was item generation. The purpose of this step was to create pools of items for each of the constructs that fit the construct definitions.

Once the item pools were created, items for the various constructs were re-evaluated to eliminate redundant or ambiguous items. This was done through structured interviews with practitioners from five different manufacturing firms. The focus was to check the relevance of

Table A

		Actual					Total	% Hits
		A	B	C	D	N/A		
Theoretical	A	26	2	1	0	1	30	87
	B	8	18	4	0	0	30	60
	C	0	0	30	0	0	30	100
	D	0	1	0	28	1	30	93

Note: Item placements, 120; hits, 102; overall hit ratio, 85%.

each construct's definitions and clarity of wordings of sample questionnaire items. The result was the following number of items in each pool:

External Environment

Level of Market Heterogeneity	10
Degree of Market Segmentation	8
Length of Product Life Cycle	10
Driving Force for Manufacturing Technology	10
Number of Customer Requirements	10
Sub-Total	48

Managerial Beliefs and Attitudes

Goals for Investing in Facilities and Equipment	10
Organization of Work	10
Scope in Decision Making	11
Management Mechanism	10
Focus of Managerial Tasks	10
Focus of Supplier Relationships Management	10
Sub-Total	61

Organizational Structure

Locus of Decision Making	12
Nature of Formalization	10
Number of Layers in Hierarchy	8
Level of Horizontal Integration	12
Level of Communication	12
Sub-Total	54
<u>Total</u>	163

Items placed in a common pool were subjected to two sorting rounds by two independent judges per round. The procedure had practitioners from the industry act as judges and sort the items from the first stage into separate constructs, based on similarities and differences among items. Based on the placements made by the judges the items could then be examined and inappropriately worded or ambiguous item could be eliminated or reworded. Two

goals for this stage were: to attempt to identify any ambiguous items, and to pre-assess the construct validity of the various scales being developed.

First, judges sorted the questionnaire items into construct categories. Each item was printed on a 3 x 5 - inch index card. The cards were shuffled into random order for presentation to the judges. Each judge sorted the cards into categories. A "not available" category definition was included to ensure that the judges did not force any item into a particular category. During the two sorting rounds, two different pairs of judges were utilized. Each set of judges included a manufacturing engineer/manager or a top management executive to ensure that the perceptions of the target population would be included in the analysis. Prior to sorting the cards, the judges were briefed with a standard set of instructions that were previously tested with a separate judge to ensure comprehensiveness and comprehensibility. Judges were allowed to ask as many questions as necessary to ensure they understood the procedure.

To assess the reliability of the sorting conducted by the judges, two different measures were made. First, for each pair of judges in each sorting step, their level of agreement in categorizing items was measured using Cohen's Kappa.

First sorting round.

In the first round, the inter-judge raw agreement scores averaged 0.74 (Table 1), the initial overall placement ratio of items within the target constructs was 84 % (Table 2), and the Kappa scores averaged 0.73. A summary of the first round inter-judge agreement indices is shown in Table 3. Following the guidelines of Landis and Koch (1977) for interpreting the Kappa coefficient, the value of 0.73 indicates a moderate, but almost excellent level of agreement beyond chance for the judges in the first round. This value is slightly lower than the value for raw agreement, which is 0.74 (Table 1). The level of item placement ratios averaged 84%. For instance, the lowest item placement ratio value was 67% for the "level of communication" construct, indicating a low degree of construct validity. On the other hand, several constructs ("driving

force for manufacturing technology” and “number of layers in hierarchy”) obtained a 100% item placement ratio, indicating a high degree of construct validity.

In order to improve the Cohen’s Kappa measure of agreement, an examination of the off-diagonal entries in the placement matrix (Table 2) was conducted. Any ambiguous items (fitting in more than one category) or too indeterminate items (fitting in no category) were either deleted or reworded. Overall, 29 items were deleted, and 25 items were reworded.

Second sorting round.

Again, two judges were involved in the second sorting round, which included the reworded items developed after the first sorting round. In the second round the inter-judge raw agreement scores averaged 0.80 (Table 4), the initial overall placement ratio of items within the targets constructs was 88 % (Table 5), and the Kappa scores averaged 0.78. A summary of the second round inter-judge agreement indices is shown in the second column of Table 3. The value for Kappa coefficient of 0.78 is higher than the value obtained in the first round, and indicates an excellent fit, based on the guidelines of Landis and Koch (1977) for interpreting the Kappa coefficient. The level of item placement ratios averaged 88%. The lowest item placement ratio value was that of 69% for the “scope in decision making” construct, indicating a low degree of construct validity. Again several constructs (“goals for investing in facilities and equipment,” “focus of supplier relationships management,” and “nature of formalization”) obtained a 100% item placement ratio, indicating a high degree of construct validity.

In order to further improve potential reliability and construct validity, an examination of the off-diagonal entries in the placement matrix (Table 5) was conducted. Again, any ambiguous items (fitting in more than one category) or too indeterminate items (fitting in no category) were either deleted or reworded. Overall, 29 items were further deleted, and 15 items were reworded.

At this point, we stopped the Q-sort method at round two, for the raw agreement score of 0.80, Cohen’s Kappa of 0.78, and the average placement ratio of 88% were considered as an excellent level of inter-judge agreement, indicating high level of reliability and construct validity.

#### Research on Quality Management

The second example is from Solis-Galvan (1998) who studied the relationship between quality management and manufacturing competitive capabilities. The basic steps for the instrument development were the same as with Nahm (2000). The first step was item generation. The following number of items were generated in each pool:

<u>Management Based Quality Practices</u>	
Top Management Support	11
Strategic Quality Planning	5
Sub-Total	16
<u>Employee Based Quality Practices</u>	
Employee Training	5
Employee Involvement	5
Employee Empowerment	5
Employee Recognition	4
Sub-Total	19
<u>Information Based Quality Practices</u>	
Quality Information Availability	5
Quality Information Usage	4
Benchmarking	4
Sub-Total	13
<u>Customer Based Quality Practices</u>	
Customer Orientation	9
Customer Closeness	6
Sub-Total	15
<u>Product/Process Based Quality Practices</u>	
Product Design	8
SPC Usage	3
Process Design	3
Sub-Total	14
<u>Supplier Based Quality Practices</u>	
Supplier Quality	7
<u>Total</u>	84

The second step was item testing. The basic procedure was to have quality managers, plant managers, and top management executives judge and sort the items from the first stage into separate quality management categories, based on similarities and differences among items. A group of potential judges were identified from the local Chapter of the *American Society for Quality*. All of the potential participants were representatives of the population targeted for this study, and considered as knowledgeable in the quality field, and with the required experience and position to assess the impact of TQM practices in their organization. From this group, a random sample of six judges was selected to participate during this stage.

First sorting round.

Two judges were involved in the first round, which included items developed for the quality management constructs (84 items). In the first round, the inter-judge raw agreement scores averaged 0.71 (60 agreements among 84

items), the initial overall placement ratio of items within the target constructs was 84 %, and the Kappa scores averaged 0.64 (Table 6).

Following the guidelines of Landis and Koch for interpreting the Kappa coefficient, the value of 0.64 indicates a moderate level of agreement beyond chance for the judges in the first round. This value is slightly lower than the value for raw agreement, which is 0.71. The level of item placement ratios averaged 84%. For instance, the lowest item placement ratio value was 55% for the "customer orientation" construct, indicating a low degree of construct validity. On the other hand, several constructs ("employee recognition," "quality information availability," "quality information usage," "usage of statistical quality control," and "benchmarking") obtained a 100% item placement ratio, indicating a high degree of construct validity.

In order to improve the Cohen's Kappa measure of agreement, an examination of the items that had a mismatch between the theoretical category and the actual placements by the judges was conducted. The analysis led to the rewording of ambiguous or indeterminate items. One additional item was included in the "product design" construct after reviewing its definition. Feedback obtained from both judges lead to the creation of two additional items for the "customer closeness" construct and one additional item for the "supplier quality" construct.

#### Second sorting round.

Two judges were involved in the second sorting round, which included the reworded and new items developed after the first sort round. In the second round the inter-judge raw agreement scores averaged 0.70 (62 agreements among 88 items).

A summary of the second round inter-judge agreements indices is shown in the second column of Table 6. The value for Kappa coefficient of 0.68 was higher than the value obtained in the first round, but still indicated a moderate level of agreement beyond chance for the judges in the second round. The level of item placement ratios averaged 82%. The lowest item placement ratio was 50% for the "employee involvement" construct, indicating a low degree of construct validity. Several constructs ("employee recognition," "quality information availability," "quality information usage," "usage of statistical quality control," and "process design") obtained a 100% item placement ratio, indicating a high degree of construct validity.

Examination of the items that had a mismatch between the theoretical and actual placements resulted in rewording four items belonging to the constructs "customer orientation," "product design," and "supplier quality." One additional question suggested by two judges was added to the "SPC usage" construct.

#### Third sorting round.

Two judges participated in the third sorting round, which included the reworded items and the new item added to the "SPC usage" construct. In the third round the inter-judge raw agreement scores averaged 0.81 (72 agreements among 89 items), the overall placement ratio of items within the targets constructs was 89%, and the Kappa scores averaged 0.81 (Table 6).

The value for Kappa coefficient of 0.81 is significantly higher than the value obtained in the second round, and indicates an excellent level of agreement beyond chance. The level of item placement ratios averaged 89%. The lowest item placement ratio value was that of 75% for the "quality information usage" construct, indicating a moderate to good degree of construct validity. The constructs "employee training," "quality information availability," "usage of statistical quality control," "process design," "supplier quality," and "benchmarking" obtained a 100% item placement ratio. This placement of items within the target construct shows that a high degree of construct validity and potential reliability were achieved.

#### Conclusion

Until now, researchers engaged in survey research had no simple way of assessing the reliability and construct validity of their questionnaire items before the large scale survey except for conducting a pilot survey on a reduced scale. By applying the Q-sort method, one can now assess them in a timely and cost-efficient manner. The method is a qualitative rather than a quantitative method, and thus may not be compared to the usefulness and the statistical power of a pilot survey. Nevertheless, in light of the importance of a quality scale development in survey research, a method like the one described here, which is a cost-efficient and simple yet insightful method of assessing reliability and construct validity of measurement items, should be of great help to researchers.

#### References

- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, Spring*, 37-46.
- Jarvenpaa, S. (1989). The effect of task demands and graphical format on information processing strategies. *Management Science*, 35, 285-303.
- Koufteros, X. A., Vonderembse, M. A., & Doll, W. J. (1998). Developing measures of time-based manufacturing. *Journal of Operations Management*, 16, 21-41.
- Landis, J. R., & Koch, C. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.

Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, 2, 192-222.

Nahm, A. Y. (2000). Exploring the relationship between external environment, internal environment, and manufacturing practices: An industrial/post-industrial perspective. Unpublished doctoral dissertation, The University of Toledo, Ohio.

Solis-Galvan, L. E. (1998). An exploratory study of the relationship between quality management and manufacturing competitive capabilities. Unpublished doctoral dissertation, The University of Toledo, Ohio.

Vessey, I. (1984). An investigation of the psychological processes underlying the debugging of computer programs. Unpublished doctoral dissertation, The University of Queensland, Queensland, Australia.

## Appendix

Table 1

Inter-Judge Raw Agreement Scores: First Sorting Round

Constructs <sup>a</sup>	Judge 1																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	NA		
1	6	2			1														
2		6			1														
3			8	1															
4				11															
5	2			1	9														
6						6													
J							1	5	3	1	1								
u								4	6	3	1								
d									1		7								
g																			
e						1					8	1							
2									1			9							
12													8		2				
13													2	9		2			
14															8				
15													1	1		10			
16																3	2	5	1
NA							1	1											

Note. Total items placement, 163; number of agreements, 121; agreement ratio, 0.74.

- <sup>a</sup>Constructs:
1. Market heterogeneity
  2. Degree of market segmentation
  3. Length of product life cycle
  4. Driving force for manufacturing technology
  5. Number of customer requirements
  6. Goals for investing in facilities and equipment
  7. Organization of work
  8. Scope in decision-making
  9. Management mechanism
  10. Focus of managerial tasks
  11. Focus of supplier relationships management
  12. Locus of decision-making
  13. Nature of formalization
  14. Number of layers in hierarchy
  15. Level of horizontal integration
  16. Level of communication



Table 2

Items Placement Ratios: First Sorting Round

Constructs <sup>a</sup>	Actual Categories																T	%		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16			NA	
T h e o r e t i c a l	1	16	1		3													20	80%	
	2	1	14		1													16	88%	
	3			17	3														20	85%
	4				20														20	100%
	5				1	19													20	95%
	6						14	2			2	2							20	70%
	7							15	3	2									20	75%
	8							5	16									1	22	73%
	9								3	16								1	20	80%
	10								2	1	17								20	85%
	11								1		1	18							20	90%
	12												19	3	2				24	79%
	13												1	18		1			20	90%
	14														16				16	100%
	15												1	1		22			24	92%
	16													1	3	3	16	1	24	67%

Note. Total items placement, 326; number of hits, 273; overall hit ratio, 84%.

<sup>a</sup>The names of constructs are as listed in Table 1.

Table 3

Inter-Judge Agreements

Agreement Measure	Round 1	Round 2
Raw Agreement	0.74	0.80
Cohen's Kappa	0.73	0.78
Placement Ratio Summary		
External environment	80%	86%
Degree of market segmentation	88%	93%
Length of product life cycle	85%	81%
Driving force for manufacturing technology	100%	95%
Number of customer requirements	95%	85%
Goals for investing in facilities and equipment	70%	100%
Organization of work	75%	93%
Scope in decision making	73%	69%
Management mechanism	80%	86%
Focus of managerial tasks	85%	94%
Focus of supplier relationships management	90%	100%
Locus of decision making	79%	83%
Nature of formalization	90%	100%
Number of layers in hierarchy	100%	81%
Level of horizontal integration	92%	91%
Level of communication	67%	75%
Average	84%	88%

Table 4

Inter-Judge Raw Agreement Scores: Second Sorting Round

Constructs <sup>a</sup>	Judge 3																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	NA	
1	5	1			1													
2	1	6																
3		1	6	1														
4				10														
5				1	7												2	
J	6					7				1								
u	7						6	3	1									
d	8							3		1	1							
g	9							2	5									
e	10									7								
	11										10							
4	12											7					1	
	13											2	9				1	
	14											1		5				
	15												1				9	
	16													1			1	5
	NA																2	

Note. Total items placement, 134; number of agreements, 107; agreement ratio, 0.80.

<sup>a</sup>The names of constructs are as listed in Table 1.

Table 5

Items Placement Ratios: Second Sorting Round

Constructs <sup>a</sup>	Actual Categories																T	%		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16			NA	
T h e o r e t i c a l	1	12	1			1												14	86%	
	2	1	13															14	93%	
	3		1	13	2														16	81%
	4			1	19														20	95%
	5				1	17												2	20	85%
	6						14												14	100%
	7							13	1										14	93%
	8							3	11		1	1							16	69%
	9								2	12									14	86%
	10						1				15								16	94%
	11											20							20	100%
	12												15	2		1			18	83%
	13													18					18	100%
	14												1		13			2	16	81%
	15													2		20			22	91%
	16												2	1		1	12		16	75%

Note. Total items placement, 268; number of hits, 237; overall hit ratio, 88%.

<sup>a</sup>The names of constructs are as listed in Table 1.

Table 6

## Inter-Judge Agreements

Agreement Measure	Round 1	Round 2	Round 3
Raw Agreement	0.71	0.70	0.81
Cohen's Kappa	0.64	0.68	0.81
Placement Ratio Summary			
Top Management Involvement	86%	91%	91%
Strategic Quality Planning	80%	80%	90%
Employee Training	80%	100%	100%
Employee Involvement	80%	50%	80%
Employee Empowerment	80%	70%	80%
Employee Recognition	100%	80%	88%
Availability of Quality Information	100%	100%	100%
Usage of Quality Information	100%	100%	75%
Customer Orientation	55%	67%	83%
Closeness to Customers	91%	63%	75%
Product Design	75%	78%	83%
Statistical Process Control Usage	100%	100%	100%
Process Management	66%	100%	100%
Supplier Relationships	93%	88%	100%
Benchmarking	100%	88%	100%
Average	84%	82%	89%