ICIS 2010 Proceedings

International Conference on Information Systems (ICIS)

2010

MATRIX KAPPA: A PROPOSAL FOR A CARD SORT STATISTIC FOR IS SURVEY INSTRUMENT DEVELOPMENT

James S. Denford
Business Administration Department Royal Military College of Canada, jim.denford@rmc.ca

MATRIX KAPPA: A PROPOSAL FOR A CARD SORT STATISTIC FOR IS SURVEY INSTRUMENT DEVELOPMENT

Research-in-Progress

James S. Denford
Business Administration Department
Royal Military College of Canada
PO Box 17000 Stn Forces
Kingston, Ontario, Canada K7K 7B4
jim.denford@rmc.ca

Abstract

The card sort is a key scale development tool that is frequently used in IS survey instrument development. Cohen's Kappa is a recommended measure of inter-rater agreement in this process, however one of its underlying statistical assumptions is violated when it is used in open card sorts. To address this issue, Matrix Kappa is proposed as a complement to other card sort analysis techniques, reframing constructs in terms of item relationships and representing inter-rater agreement in terms of matrices. Matrix Kappa has the benefit of meeting Cohen's Kappa assumptions for open card sorts and can be used to differentiate both open and closed card sort results that Cohen's Kappa cannot.

Keywords: Statistical methods, Research methods/methodology, Survey research

Introduction

Moore and Benbasat's (1991) paper describing the development of an instrument designed to measure users' perceptions of adopting an IT innovation is a classic of IS survey instrument development and validation. One of the many innovations in the paper was the use of open card sorts, which was an extension of the instrument development method used by Davis (1989). The intent of this paper is to identify one small, but significant, weakness with this otherwise strong procedure and recommend a method and new statistic for improving upon it. The weak step in the open card sort analysis is for the researcher to assign *a posteriori* the categories created by judges to the *a priori* categories from the research model as this action violates an underlying assumption of the inter-rater agreement statistic. This is an important issue for MIS researchers as the use of open card sorts is a frequent step in the scale development process for IS survey instruments.

Background

The general principle of a card sort is to confirm the coverage of a domain with a set of constructs (Davis 1989). A participant is given a set of cards where on each card a single statement is written that represents a possible item. Multiple cards contain statements that reflect underlying constructs. The task is to sort the index cards into separate piles or categories based on the similarities and differences among the statements on each card, where each pile should reflect one underlying concept. The difference between a closed and an open card sort is that in the former the categories are given to the sorter and in the latter the sorter must define the categories based on the items and her or his perceptions of the underlying unifying constructs. The open card sort stage was devised to confirm that the meaning of the constructs were well understood, while the convergence and divergence of items within given categories of the closed card sort was used to demonstrate convergent and discriminant validity (Moore and Benbasat 1991). Cohen's Kappa (Cohen 1960) is recommended to determine inter-rater agreement of card sorts.

In this discussion, certain key terms must be identified and differentiated. Inter-rater reliability refers to the consistency in order of a rating while inter-rater agreement refers to the degree to which judges give exactly the same rating to a target (Burke, Finkelstein and Dusig 1999). Reliability is correlational in nature and refers to the proportion of variance among raters (Lawlis and Lu 1972). It is more concerned with equivalence of relative rankings between judges rather than equivalence of scores (LeBreton and Senter 2008). Agreement is concerned with the interchangeability of judges and the degree to which they assign the same ratings (James, Demaree and Wolf 1984). It is typically expressed in terms of a within-group rating dispersion (LeBreton and Senter 2008). There is no necessary relationship between the two measures and therefore they should be used precisely and not interchangeably (Tinsley and Weiss 1975). Nominal scales are categorical and do not imply order, therefore the distinction between reliability and agreement blurs as the concept of proportion of variance used in reliability is no longer usable and agreement is absolute – it either exists or it does not (Tinsley and Weiss 1975). As inter-rater reliability is nonsensical for nominal scales, the term inter-rater agreement is the appropriate one for card sort analyses.

The earliest agreement indices used proportion of agreement as an indicator, but these statistics were deficient in that they did not adjust for chance agreement (Kozlowski and Hattrup 1992). As it does not adjust for chance, pure proportion of agreement will tend to overestimate the true absolute agreement between judges (Tinsley and Weiss 1975). Cohen's Kappa is a coefficient of agreement for nominal scales between two judges that adjusts for chance agreement. Assumptions of Cohen's Kappa include that: (1) units are independent; (2) categories are independent, mutually exclusive and collectively exhaustive; and (3) judges operate independently (Cohen, 1960). Cohen's Kappa is expressed as

$$\kappa = \frac{p_o - p_e}{1 - p_e} \qquad (1)$$

where p_o represents the proportion of agreement between judges and p_e represents the proportion of expected chance agreement. An alternate description of Cohen's Kappa is a ratio of disagreement between two observers, where distance is measured by aggregating binary agreement (one) or disagreement (zero) (Light 1971). A value of .70 is considered a minimum level of agreement required to justify newly developed measures (LeBreton and Senter 2008).

Card Sort Application

In a card sort, the interrelationships between the items as theoretically specified by the researcher form the benchmark for comparison of each judge. The fundamental difference between open and closed card sorts is the *a priori* assignment of categories in the latter, or rather the lack of it in the former. Judges may not develop the exact category intended by the researcher and hence there may be challenges for the researcher in determining the mapping between categories. Essentially, there are *n* sets of pairs of measures incorporating the researcher and *n* judges, where overall agreement is determined by the average of these assessments. It should be noted that the practice of comparing all judges using Cohen's Kappa as an aggregate measure should be discouraged, as it is a paired inter-rater agreement index only. For multiple judge comparisons, Fleiss' Kappa (1971), Conger's (1980) "Fleiss exact" κ_m[1], Berry and Mielke's (1988) $\mathcal{R}$ or a multiple-rater extension of Cohen's Kappa (Janson and Olsson, 2001) should be considered.

In either a hit-count measures or a Cohen's Kappa analysis, the calculation of p_o requires a judgment call by the researcher whether the emergent construct label is close enough to the intended construct label to constitute a hit – there is the potential for a substantial degree of subjectivity in this process. As the categories are unspecified, they cannot be independent, mutually exclusive or collectively exhaustive; hence Cohen's Kappa cannot be used as one of its assumptions has been violated. Rigorous validation may require statistical substantiation of category assignments, but Cohen's Kappa is inappropriate when applied to *a posteriori* allocation of judges' categorizations, leaving the researcher with a quandary of how to objectively substantiate these judgments.

The revalidation of an existing validated multiple construct scale – the STRategic Orientation of the Existing Portfolio of Information Systems (STROEPIS – Chan, Huff, Barclay and Copeland 1997) as depicted in Figure 1 – is used to illustrate this issue. The conceptualization of STROEPIS used in this paper includes seven dimensions: IS Support for Aggressiveness, IS Support for Analysis, IS Support for Internal Defensiveness, IS Support for External Defensiveness, IS Support for Futurity, IS Support for Proactiveness and IS Support for Riskiness. For the sake of brevity, all further use of the construct name in this paper drops, but still assumes, the 'IS Support for' component. Items used in this illustration are listed in Appendix A in order to help conceptualize occasions of misloading.

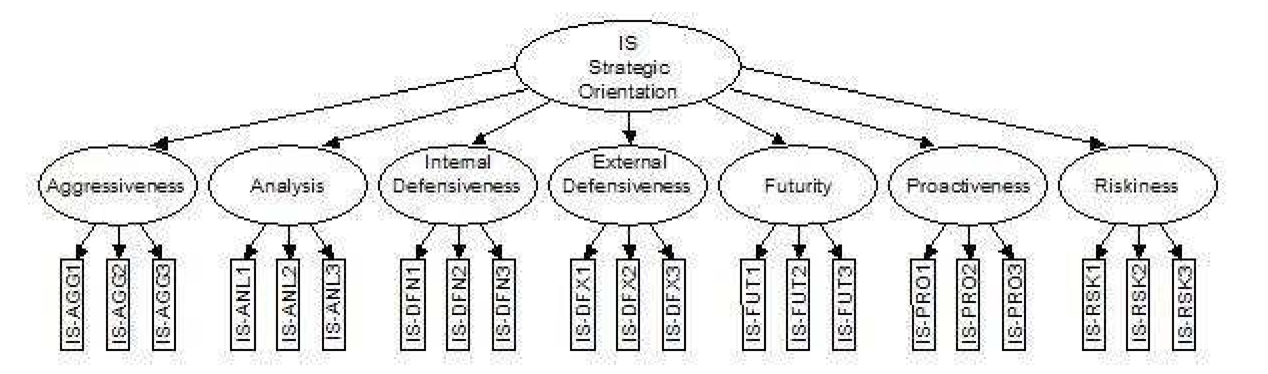


Figure 1. The STRategic Orientation of the Existing Portfolio of Information Systems (STROEPIS)

A possible solution is to change the level of analysis from the construct, which is undefined, to the item, which is well-defined. When reflective items load well on their constructs, they should also be highly correlated with each other and not correlated with other items (Straub, Gefen and Boudreau 2004). Using the Aggressiveness and Analysis constructs from STROEPIS as an example, Figure 2a shows three items loading correctly on each of the two constructs, while Figure 2b shows one of the Analysis items inappropriately loading on the Aggressiveness construct. The relationships between items exist whether or not the construct has been explicitly defined, so in the absence of a defined construct, this relationship can be used to infer the existence of the unobserved construct through the linkages between items.

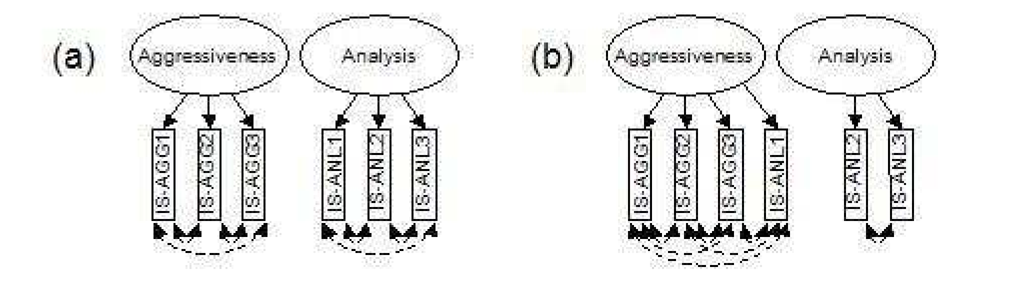
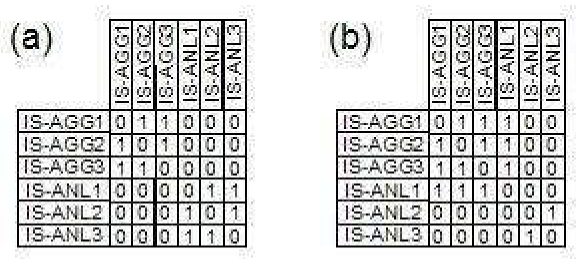


Figure 2. Graphical Depiction of Constructs and Items

Light (1971) advocated representing agreement in binary form, but at the construct level. Using this conceptualization at the measure level, items can be considered nodes related to an overall construct and a comparison can be made between the paths between nodes. For example, the three items for Aggressiveness should be related to each other and not to Analysis, therefore each item should have two paths, as illustrated in Figure 3a. Lack of a path and additional paths both indicate items not mapping on the construct. In Figure 3b, the additional paths in Aggressiveness and the lack of expected paths in Analysis illustrate the misloading of an item.



(a)

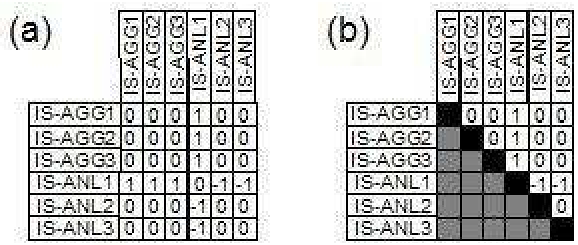
	IS-AGG1	IS-AGG2	IS-AGG3	IS-ANL1	IS-ANL2	IS-ANL3
IS-AGG1	0	1	1	0	0	0
IS-AGG2	1	0	1	0	0	0
IS-AGG3	1	1	0	0	0	0
IS-ANL1	0	0	0	0	1	1
IS-ANL2	0	0	0	1	0	1
IS-ANL3	0	0	0	1	1	0

(b)

	IS-AGG1	IS-AGG2	IS-AGG3	IS-ANL1	IS-ANL2	IS-ANL3
IS-AGG1	0	1	1	1	0	0
IS-AGG2	1	0	1	1	0	0
IS-AGG3	1	1	0	1	0	0
IS-ANL1	1	1	1	0	0	0
IS-ANL2	0	0	0	0	0	1
IS-ANL3	0	0	0	0	1	0

Figure 3. Matrix Depiction of Item Relationships

Misloading is a term that can be used only if there is an *a priori* expectation of how the items should load by the researcher. This can be stated as the items were specifically selected by the researcher to load on particular constructs that he or she has developed (Moore and Benbasat 1991). It does not, however, equate to judges being 'incorrect' in their item selection, as the intent of the card sort is to validate the instrument using their independent perceptions of the items and constructs. Using the same example, Figures 2a and 3a can be used to represent the intended item loadings by the researchers and Figures 2b and 3b can be used to represent the item loadings of a judge, where one item is misplaced. The difference between the matrices can be used to capture the degree of disagreement between the two item lists, where zeros represent agreement, minus ones represent expected links that are missing and plus ones represent unexpected links that are present, as illustrated in Figure 4a. The relevant level of agreement can be found in the upper corner above the diagonal, as illustrated in Figure 4b. Here, there are 15 possible agreements, where an agreement is defined in terms of either concurrence on whether there should be the existence or absence of a theorized relationship expected by the researcher. Of the 15 possible agreements there are three unexpected relationships and two absent ones, leaving ten actual agreements. This matrix representation can aid the researcher in visually and intuitively identifying potential issues with scale development.



(a)

	IS-AGG1	IS-AGG2	IS-AGG3	IS-ANL1	IS-ANL2	IS-ANL3
IS-AGG1	0	0	0	1	0	0
IS-AGG2	0	0	0	1	0	0
IS-AGG3	0	0	0	1	0	0
IS-ANL1	1	1	1	0	-1	-1
IS-ANL2	0	0	0	-1	0	0
IS-ANL3	0	0	0	-1	0	0

(b)

	IS-AGG1	IS-AGG2	IS-AGG3	IS-ANL1	IS-ANL2	IS-ANL3
IS-AGG1	■	0	0	1	0	0
IS-AGG2	▨	■	0	1	0	0
IS-AGG3	▨	▨	■	1	0	0
IS-ANL1	▨	▨	▨	■	-1	-1
IS-ANL2	▨	▨	▨	▨	■	0
IS-ANL3	▨	▨	▨	▨	▨	■

Figure 4. Matrix Depiction of Inter-Rater Agreement

In addition to the visual cues, matrix representation aids statistical analysis of the card sort. Hit-counts have already been seen to be ineffective as they do not adjust for chance agreement. In the matrix-item form, however, all three of Cohen's (1960) assumptions are met as now the paths are independent, mutually exclusive and collectively exhaustive, therefore Kappa can be calculated. To differentiate from Cohen's Kappa, this new statistic is termed 'Matrix Kappa'.

First, the hit-count can be used to calculate a matrix-size adjusted $p_{om.}$ For the example of two constructs each with three items, there are ten of 15 possible agreements, for a raw probability of 0.667. It should be noted that as constructs and items are added, the matrix size expands exponentially, hence an exponential deflator is required to ensure proportionality of Matrix Kappa with Cohen's Kappa. For example, the addition of a third three-item construct would increase the number of possible agreements to 36, a 140% increase, while increasing the number of constructs and items only by 50%. To adjust for the size of the matrix, it is suggested to take into consideration the number of items per construct (I_j) and the number of constructs (n), adjusted for degrees of freedom, to determine an adjustment constant (c)

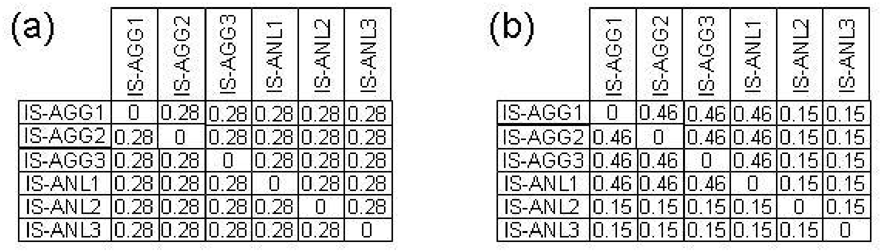
$$c_o = \frac{n\,(n-1)}{\sum_{j=1}^{n} (I_j - 1)} \tag{2}$$

where

$$p_{om} = p_o^{\,co} \tag{3}$$

In the case of the example with two constructs, each with three items, the constant would be 0.5 and the p_{om} would be 0.816, compared to a Cohen's Kappa p_o of 0.833.

Next, the adjustment for chance can be determined. The calculation of p_{em} is based on the joint probabilities of the marginal proportions described by Cohen (1960), but adjusted for matrix representation. The marginal probabilities can be represented as another matrix, with the diagonal set to zero, as represented in Figures 5a and 5b for the theoretical item distribution and the judges distribution.



(a)

	IS-AGG1	IS-AGG2	IS-AGG3	IS-ANL1	IS-ANL2	IS-ANL3
IS-AGG1	0	0.28	0.28	0.28	0.28	0.28
IS-AGG2	0.28	0	0.28	0.28	0.28	0.28
IS-AGG3	0.28	0.28	0	0.28	0.28	0.28
IS-ANL1	0.28	0.28	0.28	0	0.28	0.28
IS-ANL2	0.28	0.28	0.28	0.28	0	0.28
IS-ANL3	0.28	0.28	0.28	0.28	0.28	0

(b)

	IS-AGG1	IS-AGG2	IS-AGG3	IS-ANL1	IS-ANL2	IS-ANL3
IS-AGG1	0	0.46	0.46	0.46	0.15	0.15
IS-AGG2	0.46	0	0.46	0.46	0.15	0.15
IS-AGG3	0.46	0.46	0	0.46	0.15	0.15
IS-ANL1	0.46	0.46	0.46	0	0.15	0.15
IS-ANL2	0.15	0.15	0.15	0.15	0	0.15
IS-ANL3	0.15	0.15	0.15	0.15	0.15	0

Figure 5. Matrix Depiction of Marginal Probabilities

The value for p_{em} is calculated as the sum of products between the two marginal probability matrices, adjusted for both the diagonal zeros and the relevant upper corner above the diagonals. This is the matrix analogue to the calculation of joint probabilities in the original Cohen's Kappa formulation, which can be expressed mathematically as

$$c_e = \left(2 \sum_{i=1}^{n} \sum_{j=1}^{m} A_{ij} \right) \left(2 \sum_{i=1}^{n} \sum_{j=1}^{m} B_{ij} \right) \tag{4}$$

where

$$p_{em} = c_e \sum_{i=1}^{n} \sum_{j=1}^{m} A_{ij}B_{ij} \tag{5}$$

In the case of the example, the constant would be 2.69 and p_{em} would be 0.403.

As the theoretically derived constructs are known, it is possible to compare the original Cohen's Kappa values with those of Matrix Kappa. Using the two-construct, six-item example, κ is 0.667, p_o is 0.833 and p_e is 0.5, while κ_m is 0.692, p_{om} is 0.816 and p_{em} is 0.403. This demonstrates that, for this example, both Kappas would find that the measure was below the .70 minimum level of agreement required for new scale development (LeBreton and Senter, 2008). The next section provides an empirical example of this congruence, but also highlights where the measures may differ.

Empirical Illustration

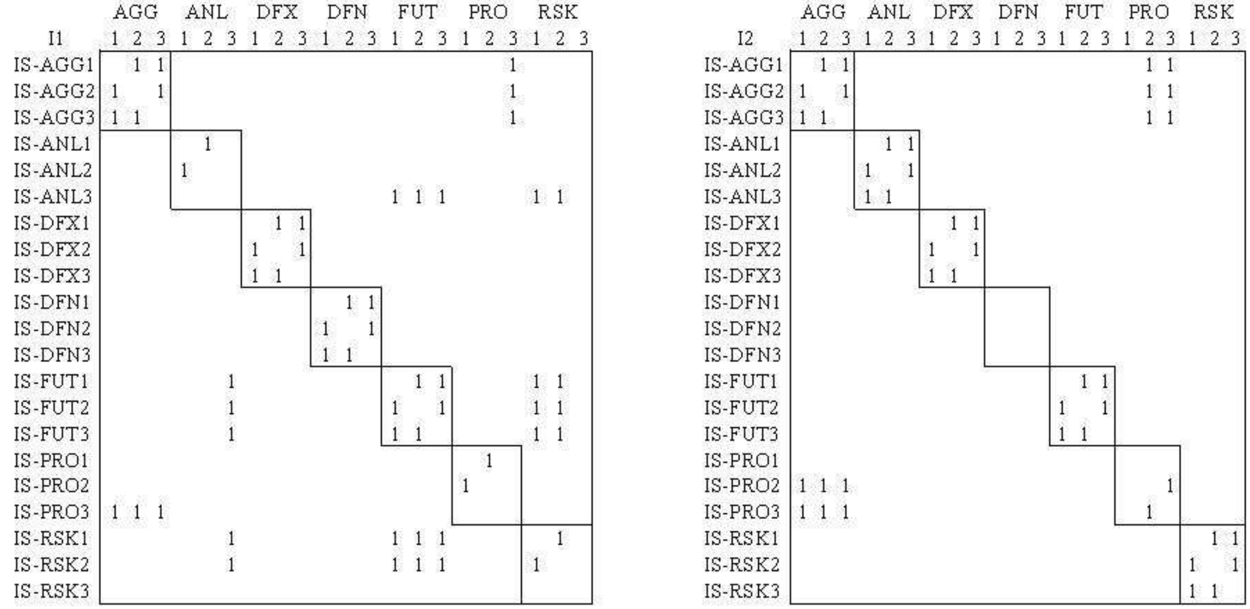
An important question in judging the usefulness of a new method is how closely it approximates the results of the existing procedure. A test was conducted using four judges sorting STROEPIS items in a closed card sort, with results identified in Table 1. A closed card sort was used for the test as it allows the use of Cohen's Kappa as a references since the statistical assumptions of both Cohen's Kappa and Matrix Kappa are met. The average difference between the Kappas was 0.026 and the root of the average sum of squares differences between the two Kappas was .056. While this absolute difference between the Kappa values would be a strong measure, there were insufficient judges for each sort task to meet the assumptions of parametric statistics to comment on significance.

Table 1. Card Sort Results

	Cohen's Kappa				Matrix Kappa			
	Kappa	Po	Pe	Rank	Kappa	Pom	Pem	Rank
I1	0.778	0.810	0.143	4	0.676	0.728	0.161	4
I2	0.781	0.810	0.129	3	0.813	0.844	0.164	3
I3	1.000	1.000	0.143	1	1.000	1.000	0.169	1
I4	0.944	0.952	0.143	2	0.912	0.927	0.168	2

With the small sample of judges, an appropriate comparison measure would be the Spearman rank-order correlation coefficient. As the ranks are identical for each of the three strategy types, it can be seen that the value of P (rho) would be 1 for the set, indicating perfect correlation.

A contrasting question to answer regarding the use of a new procedure is if it makes a difference or not to findings or the process it supports. An illustration of the difference can be found by examining judges I1 and I2, whose Cohen's Kappas were nearly identical but whose Matrix Kappas were very different. Both judges had 17 out of 21 right items, yielding a P_o of 0.81. However, the P_{om} for I1 was 0.728 and for I2 was 0.844. Examining the matrices in Figure 6, while I1 and I2 had a similar number of missing paths (six and five respectively), I1 had many more extra paths defined than I2 (14 and six respectively). At the attribute level, I1 only agreed with the researcher on two constructs – Internal and External Defensiveness – having assigned several Analysis, Proactiveness and Riskiness items to Aggressiveness and Futurity attributes. In comparison, I2 agreed with the researcher on four attributes completely but fused Proactiveness and Aggressiveness items into a single attribute and did not identify a coherent Internal Defensiveness attribute. It may be argued that I1 had a worse outcome than I2 as a greater proportion of the instrument was supported, which was signaled by the lower Matrix Kappa and could be identified through the matrix representation.



	AGG			ANL			DFX			DFN			FUT			PRO			RSK		
I1	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
IS-AGG1		1	1															1			
IS-AGG2	1		1															1			
IS-AGG3	1	1																1			
IS-ANL1					1																
IS-ANL2				1																	
IS-ANL3													1	1	1				1	1	
IS-DFX1								1	1												
IS-DFX2							1		1												
IS-DFX3							1	1													
IS-DFN1											1	1									
IS-DFN2										1		1									
IS-DFN3										1	1										
IS-FUT1						1								1	1				1	1	
IS-FUT2						1							1		1				1	1	
IS-FUT3						1							1	1					1	1	
IS-PRO1																	1				
IS-PRO2																1					
IS-PRO3	1	1	1																		
IS-RSK1						1							1	1	1					1	
IS-RSK2						1							1	1	1				1		
IS-RSK3																					

	AGG			ANL			DFX			DFN			FUT			PRO			RSK		
I2	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
IS-AGG1		1	1														1	1			
IS-AGG2	1		1														1	1			
IS-AGG3	1	1															1	1			
IS-ANL1					1	1															
IS-ANL2				1		1															
IS-ANL3				1	1																
IS-DFX1								1	1												
IS-DFX2							1		1												
IS-DFX3							1	1													
IS-DFN1																					
IS-DFN2																					
IS-DFN3																					
IS-FUT1														1	1						
IS-FUT2													1		1						
IS-FUT3													1	1							
IS-PRO1																					
IS-PRO2	1	1	1															1			
IS-PRO3	1	1	1														1				
IS-RSK1																				1	1
IS-RSK2																			1		1
IS-RSK3																			1	1	

Figure 6. Matrix Comparison Between Judges I1 and I2

Discussion and Conclusions

The Matrix Kappa statistic appears to be more sensitive to misassignment of items, particularly where concepts intended as being individual are grouped into homogeneous groups. The inability to discriminate between constructs could then be determined using this statistic, but not necessarily Cohen's Kappa. If this argument were satisfactorily demonstrated, then the Matrix Kappa statistic could be posited as being a complement to Cohen's Kappa in discriminant validity assessment for closed card sorts and as being a replacement for Cohen's Kappa for open card sorts. This recommendation would require additional research using different data and scales prior to adoption by the research community.

Specifically, further validation of the Matrix Kappa method is required prior to any suggestion of wide-spread adoption. The current work-to-date merely points to potential benefits of matrix representation as a graphical tool and Matrix Kappa as a statistic. Future work will include validation using constructed, random and actual data sets. Additional study will also include combinatorial analysis (as the number of outcomes is large but finite) and/or a Monte Carlo simulation to test the characteristics and limits of the method. Finally, an examination of previous studies using Cohen's Kappa for scale development to identify if Matrix Kappa would result in a different outcome would also be appropriate.

As a methodological contribution, matrix representation and Matrix Kappa are proposed as complements to the card sort analysis techniques recommended by Moore and Benbasat (1991). Matrix Kappa has the benefit of meeting Cohen's Kappa assumptions for open card sorts and can be used to discriminate different levels of card sort success that Cohen's Kappa cannot. Additionally, while Matrix Kappa addresses statistical issues particular for open card sorts, it can be used for both open and closed sorts.

As seen in the brief test case of the Analysis scale in STROEPIS, the underlying matrix analysis allows for more rigorous open card sorts in scale development. This has the potential to decrease the need for multiple pilot tests at later stages of instrument testing or the requirement to return to scale development for additional card sorts after a pilot test as the presentation of the matrix can flag consistent issues with individual items that a single statistic may not. By reframing the card sort in terms of nodes-and-paths, a useful statistical tool and an intuitive graphical one have been proposed for researchers to identify problems with item groupings early in the scale development process.

Acknowledgements

I would like to acknowledge the funding support of the Royal Military College of Canada and The Monieson Centre at Queen's University, without which this research could not have been undertaken. I would also like to thank for their comments on earlier versions of this paper Yolande Chan, Brent Gallupe, Kathryn Brohman, Alain Pinsonneault, Keith Rogers, Tracy Jenkin and Terrence Beckman and for their editorial input on this paper the Associate Editor and three anonymous reviewers.

Appendix A – STROEPIS Items

Item	Description
IS-AGG1	Our IS help us be (or become) one of the top companies in our market(s).
IS-AGG2	Our IS help us stay ahead of (or catch up with) the competition.
IS-AGG3	Our IS helps us try to be a market leader.
IS-ANL1	Our IS allow us to be number-oriented and analytical in our operations.
IS-ANL2	Our IS provide us with detailed, factual information to support our decision making.
IS-ANL3	Our IS help us develop comprehensive situational analyses to aid decision making.
IS-DFX1	Our IS enable us to establish close relationships with our suppliers.
IS-DFX2	Our IS enable us to establish close relationships with our customers.
IS-DFX3	Our IS enables us to integrate forwards with customers and backwards with suppliers.
IS-DFN1	Our IS help us search for new methods for reducing costs.
IS-DFN2	Our IS are focused on helping us to improve operating efficiency.
IS-DFN3	Our IS enable cost control through performance monitoring.
IS-FUT1	Our IS provide performance metrics that emphasize our long-term business effectiveness.
IS-FUT2	Our IS provide information supporting capital budget allocation decisions reflecting long-term considerations.
IS-FUT3	Our IS provide data that is oriented to short-term decision making.
IS-PRO1	Our IS enables us to be pioneers in new markets.
IS-PRO2	Our IS enable us to be the first ones to introduce various products and/or services in the market.
IS-PRO3	Our IS make it easier for us to adopt innovations earlier than competitors.
IS-RSK1	Our IS support our tendency to be risk averse in our decision making.

Item	Description
IS-RSK2	Our IS provide data to support conservative decision making.
IS-RSK3	Our IS help us to avoid risky projects.

References

Berry, K.J., and Mielke Jr, P.W. 1988. "A Generalization of Cohen's Kappa Agreement Measure to Internal Measurement and Multiple Raters," *Educational and Psychological Measurement* (48:4), December, pp. 921-933.

Burke, M.J., Finkelstein, L.M., and Dusig, M.S. 1999. "On Average Deviation Indices for Estimating Interrater Agreement," *Organizational Research Methods* (2:1), January, pp. 49-68.

Chan, Y.E., Huff, S.L., Barclay, D.W., and Copeland, D.G. 1997. "Business Strategy Orientation, Information Systems Strategic Orientation, and Strategic Alignment," *Information Systems Research* (8:2), June, pp. 125-150.

Cohen, J. 1960. "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement* (20:1), April, pp. 37-46.

Conger, A. J. 1980. "Integration and Generalization of Kappas for Multiple Raters," *Psychological Bulletin* (88:2), September, pp. 322-328.

Davis, F.D. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* (13:3), September, pp. 319-339.

Fleiss, J. L. 1971. "Measuring nominal scale agreement among many raters," *Psychological Bulletin* (76:5), November, pp. 378–382.

James, L.R., Demaree R.G., and Wolf, G. 1984. "Estimating Within-Group Interrater Reliability With and Without Response Bias," *Journal of Applied Psychology* (69:1), February, pp. 85-98.

Janson, H., and Olsson, U. 2001. "A Measurement of Agreement for Interval or Nominal Multivariate Observations," *Educational and Psychological Measurement* (61:2), April, pp. 277-289.

Kozlowski, S.W.J., and Hattrup, K. 1992. "A Disagreement About Within-Group Agreement: Disentanglement Issues of Consistency Versus Consensus," *Journal of Applied Psychology* (77:2), April, pp. 161-167.

Lawlis, G.F., and Lu, E. 1972. "Judgment of Counseling Process: Reliability, Agreement, and Error," *Psychological Bulletin* (78:1), July, pp. 17-20.

LeBreton, J.M., and Senter, J.L. 2008. "Answers to 20 Questions About Interrater Reliability and Interrater Agreement," *Organizational Research Methods* (11:4), October, pp. 815-852.

Light, R.J. 1971. "Measures of Response Agreement for Qualitative Data: Some Generalizations and Alternatives," *Psychological Bulletin* (76:5), November, pp. 365-377.

Moore, G.C., and Benbasat, I. 1991. "Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation," *Information Systems Research* (2:3), September, pp. 192-221.

Straub, D. W., Gefen, D., and Boudreau M. 2004. "Validation Guidelines for IS Positivist Research," *Communications of the Association for Information Systems* (13:Article 24), pp. 380-427.

Tinsley, H.E.A., and Weiss, D.J. 1975. "Interrater Reliability and Agreement of Subjective Judgments," *Journal of Counseling Psychology* (22:4), July, pp. 358-376.