A formal proof of a paradox associated with Cohen's kappa

Warrens, M.J.

Journal of Classification 27:322-332 (2010)
DOI: 10.1007/s00357-010-9060-x

A Formal Proof of a Paradox Associated with Cohen's Kappa

Matthijs J. Warrens

Leiden University, The Netherlands

Abstract: Suppose two judges each classify a group of objects into one of several nominal categories. It has been observed in the literature that, for fixed observed agreement between the judges, Cohen's kappa penalizes judges with similar marginals compared to judges who produce different marginals. This paper presents a formal proof of this phenomenon.

Keywords: Inter-rater reliability; Nominal agreement; Rearrangement inequality; Marginal homogeneity; Marginal asymmetry.

1. Introduction

The kappa statistic introduced by Cohen (1960) can be used as a descriptive measure for summarizing agreement between two judges across a number of objects (individuals, things) (Brennan and Prediger 1981; Zwick 1988; Warrens 2010). Compared to the observed proportion of agreement, the advantage of kappa is its correction for the amount of agreement that can be expected to occur by chance alone (Cohen 1960; Brennan and Prediger 1981; Kraemer, Periyakoil and Noda 2004). Cohen's kappa has been primarily used as a measure of agreement or reliability (Hubert 1977; Kraemer 1979; Brennan and Prediger 1981; Zwick 1988; Byrt, Bishop and Carlin 1993; Vach 2005), but has also been proposed as a measure of validity (Wackerly and Robinson 1983; Thompson and Walter 1988). Bakeman, Quera, McArthur and Robinson (1997) pointed out that there is no one value of kappa that can be regarded as universally acceptable. The popularity of kappa has led to the development of many extensions (Nelson and Pepe

Author's Address: Matthijs J. Warrens, Institute of Psychology, Unit Methodology and Statistics, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands, e-mail: warrens@fsw.leidenuniv.nl

Published online 8 October 2010

2000, p. 479; Kraemer et al. 2004), including, multi-rater kappas (Conger 1980; Lipsitz, Laird and Brennan 1994; De Mast 2007), weighted kappas (cf. Kraemer et al. 2004) and a fuzzy kappa (Dou, Ren, Wu, Ruan, Chen, Bloyet and Constans 2007).

Several authors have identified difficulties with kappa's interpretation (Brennan and Prediger 1981; Thompson and Walter 1988; Feinstein and Cicchetti 1990; Guggenmoos-Holzmann 1996; Lantz and Nebenzahl 1996; Nelson and Pepe 2000; Vach 2005; Gwet 2008). Because kappa takes the probabilities with which judges use rating categories into account, the statistic is known to be marginal dependent or prevalence dependent (Thompson and Walter 1988; Goodman 1991; Vach 2005; Von Eye and Von Eye 2008). A paradox associated with Cohen's kappa is that, for a fixed value of the proportion of observed agreement, tables with marginal asymmetry produce higher values of kappa than tables with homogeneous marginals. Judges that produce similar marginals are thus penalized compared to judges with different marginals. Because in a typical study of agreement (reliability) there is no criterion for the correctness of an assignment, and because there is no restriction on the distribution of the judgments over the categories for either judge (Cohen 1960; Kraemer 1979; Brennan and Prediger 1981, p. 692), one expects that judges that produce similar marginals obtain a higher agreement rate. The paradox was first observed in Brennan and Prediger (1981, p. 692), and is discussed in Zwick (1988, p. 377), Feinstein and Cicchetti (1990), Byrt et al. (1993, p. 424), Lantz and Nebenzahl (1996), Nelson and Pepe (2000) and Vach (2005, p. 656). The phenomenon was first called a paradox in Feinstein and Cicchetti (1990).

The paradox associated with kappa has been illustrated by several of the above authors with examples of agreement tables. This paper presents a formal proof of the paradox. The paradox has been primarily discussed for the 2×2 case (Feinstein and Cicchetti 1990; Cicchetti and Feinstein 1990; Lantz and Nebenzahl 1996). However, the Kappa Paradox Theorem presented in this paper formalizes the paradox for $n \times n$ agreement tables. It is proved that, for fixed observed agreement, an agreement table with balanced (uniform) marginals produces a higher value of kappa then the table with symmetric marginals. This phenomenon has been illustrated in Bakeman et al. (1997) and Von Eye and Von Eye (2008). Furthermore, it is proved that a table with asymmetric marginals produces a higher value of kappa then the table with balanced marginals. Thus, the notion that the value of Cohen's kappa is highest for balanced marginal distributions, is incorrect. Moreover, for fixed observed agreement, judges that produce similar marginals are penalized compared to judges with different marginals.

Vach (2005, p. 659) points out that the paradox is a direct consequence of the definition of kappa and its aim to adjust the observed (raw)

agreement with respect to the expected amount of agreement under chance conditions. It is the aim of the kappa statistic to judge the same proportion of observed agreement differently in the light of the marginal distributions, which determine the expected amount of chance agreement. That this may lead to difficulties with kappa's interpretation is perhaps not a serious drawback of the measure (Vach 2005).

The paper is organized as follows. The Rearrangement Inequality, which is used in the proof of the Kappa Paradox Theorem, is discussed in the next section. The Kappa Paradox Theorem is then presented in Section 3. For the results in this paper it suffices to introduce kappa as a descriptive measure for summarizing agreement beyond chance. Alternatively, Kraemer (1979), Kraemer et al. (2004) and De Mast (2007) discuss kappa as a sample estimate of a parameter of a population (in the context of a statistical inference procedure). A second application of the Rearrangement Inequality is presented in Section 4. Section 5 contains a discussion.

2. The Rearrangement Inequality

For the definition of marginal symmetry and asymmetry of an agreement table in Section 3, we need the following definition for two tuples of the same length.

Definition. Two n-tuples $(a_1, ..., a_n)$ and $(b_1, ..., b_n)$ are said to be

- *similarly arranged* if there exists a permutation $(\sigma_1, ..., \sigma_n)$ of $1, ..., n$ such that the permuted tuples $(a_{\sigma_1}, ..., a_{\sigma_n})$ and $(b_{\sigma_1}, ..., b_{\sigma_n})$ are both increasing, that is, $a_{\sigma_1} \leq ... \leq a_{\sigma_n}$ and $b_{\sigma_1} \leq ... \leq b_{\sigma_n}$.

- *oppositely arranged* if there exists a permutation $(\sigma_1, ..., \sigma_n)$ of $1, ..., n$ such that of the permuted tuples $(a_{\sigma_1}, ..., a_{\sigma_n})$ and $(b_{\sigma_1}, ..., b_{\sigma_n})$ one is increasing and the other is decreasing, for example, $a_{\sigma_1} \leq ... \leq a_{\sigma_n}$ and $b_{\sigma_1} \geq ... \geq b_{\sigma_n}$.

The following result called the Rearrangement Inequality can be found in, for example, Hardy, Littlewood and Pólya (1988, p. 261).

Rearrangement Inequality. *Let $(a_1, ..., a_n)$ and $(b_1, ..., b_n)$ be two n-tuples of real numbers and $(x_1, ..., x_n)$ a permutation of $(b_1, ..., b_n)$. If $(a_1, ..., a_n)$ and $(b_1, ..., b_n)$ are similarly arranged, then*

$$\sum_{i=1}^{n} a_i b_i \geq \sum_{i=1}^{n} a_i x_i. \tag{1}$$

If $(a_1, ..., a_n)$ *and* $(b_1, ..., b_n)$ *are oppositely arranged, then*

$$\sum_{i=1}^{n} a_i b_i \leq \sum_{i=1}^{n} a_i x_i. \tag{2}$$

An application of the Rearrangement Inequality is presented in Section 4. The Rearrangement Inequality is also used in the proof of Theorem 1. We need Theorem 1 in the proof of the Kappa Paradox Theorem considered in Section 3.

Theorem 1. *Let* $(a_1, ..., a_n)$ *and* $(b_1, ..., b_n)$ *be two* n*-tuples of real numbers that satisfy*

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i = 1, \tag{3}$$

and let $(x_1, ..., x_n)$ *be a permutation of* $(b_1, ..., b_n)$*. If* $(a_1, ..., a_n)$ *and* $(b_1, ..., b_n)$ *are similarly arranged, then*

$$\sum_{i=1}^{n} a_i b_i \geq \frac{1}{n}. \tag{4}$$

If $(a_1, ..., a_n)$ *and* $(b_1, ..., b_n)$ *are oppositely arranged, then*

$$\sum_{i=1}^{n} a_i b_i \leq \frac{1}{n}. \tag{5}$$

Proof: We first consider the proof of (4). Consider the n variants of (1) such that each product $a_i b_j$ for $i, j = 1, ..., n$ on the right-hand side occurs exactly once. Adding these n variants and dividing the result by n, we obtain

$$\sum_{i=1}^{n} a_i b_i \geq \frac{1}{n} \left(\sum_{i=1}^{n} a_i \right) \left(\sum_{i=1}^{n} b_i \right). \tag{6}$$

Inequality (4) then follows from using (3) in (6). Inequality (5) follows from considering n variants of inequality (2).

■

3. The Kappa Paradox Theorem

Suppose two judges classify each of m $(m > 0)$ objects into one of n $(n \geq 2)$ categories. The classifications can be displayed in an $n \times n$ agreement table **P** with entries p_{ij}, where p_{ij} is the proportion of objects placed in category i by the first judge and in category j by the second judge.

The observed (raw) and expected proportions of agreement are given by, respectively,

$$p_o = \sum_{i=1}^{n} p_{ii} \quad \text{and} \quad p_e = \sum_{i=1}^{n} p_{i+}p_{+i},$$

where

$$p_{i+} = \sum_{j=1}^{n} p_{ij} \quad \text{and} \quad p_{+j} = \sum_{i=1}^{n} p_{ij},$$

are the marginal probabilities of $\mathbf{P}$. Suppose the data are a product of chance concerning two different frequency distributions, one for each nominal variable (judge). Quantity p_e is the value of p_o under statistical independence. p_e can be obtained by considering all permutations of the observations of one of the nominal variables, while preserving the order of the observations of the other variable. For each permutation the value of p_o can be determined. The arithmetic mean of these values is $\sum_{i=1}^{n} p_{i+}p_{+i}$.

As a measure for nominal agreement, Cohen (1960) proposed the kappa coefficient:

$$\kappa = \frac{p_o - p_e}{1 - p_e}.$$

The Kappa Paradox Theorem below is concerned with symmetric and asymmetric marginal probabilities. The following definition concerns the marginals of $\mathbf{P}$. The definition of strong marginal symmetry is merely presented to distinguish strong and weak marginal symmetry.

Definition. Consider the marginals $(p_{1+}, ..., p_{n+})$ and $(p_{+1}, ..., p_{+n})$ of $\mathbf{P}$. Table $\mathbf{P}$ is

- *strongly marginal symmetric* if $p_{i+} = p_{+i}$ for all i.
- *weakly marginal symmetric* if $(p_{1+}, ..., p_{n+})$ and $(p_{+1}, ..., p_{+n})$ are similarly arranged.
- *marginal asymmetric* if $(p_{1+}, ..., p_{n+})$ and $(p_{+1}, ..., p_{+n})$ are oppositely arranged.

Marginals $(p_{1+}, ..., p_{n+})$ and $(p_{+1}, ..., p_{+n})$ are *balanced* if $p_{i+} = 1/n$, respectively $p_{+i} = 1/n$, for all i.

Next, we present the Kappa Paradox Theorem for $n \times n$ tables. Note that the three agreement tables $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$ in the Kappa Paradox Theorem can have completely different marginal distributions.

Kappa Paradox Theorem. *Let $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$ be three agreement tables with the same proportion of observed agreement p_o, and let κ_1, κ_2 and κ_3, denote the values of kappa of the three tables. Furthermore,*

- *let* $\mathbf{P}_1$ *be weakly marginal symmetric.*
- *let either the row or column marginals (or both) of* $\mathbf{P}_2$ *be balanced.*
- *let* $\mathbf{P}_3$ *be marginal asymmetric.*

Then $\kappa_1 \leq \kappa_2 \leq \kappa_3$.

Proof: Since kappa is a decreasing function of p_e (Byrt et al. 1993, p. 429; Warrens, 2008a, p. 496), it must be shown that the p_e of κ_1 is never smaller than the p_e of κ_2, which in turn must never be smaller than the p_e of κ_3. If the row marginals are balanced, we have

$$\sum_{i=1}^{n} \frac{p_{+i}}{n} = \frac{1}{n}, \quad \text{because} \quad \sum_{i=1}^{n} p_{+i} = 1.$$

The same property holds for balanced column marginals. By Theorem 1, $p_e \geq 1/n$ if the agreement table is weakly marginal symmetric, and $p_e \leq 1/n$ if the agreement table is marginal asymmetric. This completes the proof.
■

An illustration of the Kappa Paradox Theorem is presented in Table 1. Table 1 contains three hypothetical 4×4 agreement tables with categories A, B, C and D. In case 1, the table is weakly marginal symmetric. In case 2, both the row and column marginals of the table are balanced. In case 3, the table is marginal asymmetric. In all three cases the proportion of observed agreement $p_o = .65$. Furthermore, the three agreement tables in Table 1 have completely different marginal distributions.

The table with balanced marginals produces a higher value of kappa then the table with symmetric marginals. Furthermore, the table with asymmetric marginals produces a higher value of kappa then the table with balanced marginals. For fixed observed agreement, judges that produce similar marginals are thus penalized compared to judges with different marginals.

4. Another Theorem

In the Kappa Paradox Theorem, the three agreement tables $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$ may have completely different marginal distributions. Using the Rearrangement Inequality, we may derive an additional result for tables that have the same marginals $(p_{1+}, ..., p_{n+})$ and $(p_{+1}, ..., p_{+n})$, but that differ in how the marginals are arranged over the categories. For a fixed value of the proportion of observed agreement and given marginals $(p_{1+}, ..., p_{n+})$ and $(p_{+1}, ..., p_{+n})$, the lowest (highest) value of kappa is obtained if $(p_{1+}, ..., p_{n+})$ and $(p_{+1}, ..., p_{+n})$ are similarly (oppositely) arranged.

Table 1: Values of kappa for three hypothetical cases. In all three cases the proportion of observed agreement $p_o = .65$, but the row and column marginals are different.

		Judge 2				
	Judge 1	A	B	C	D	Total
Case 1:	A	.30	.15	.05		.50
Weak marginal	B	.05	.20			.25
symmetry	C	.05		.10		.15
	D	.05			.05	.10
$\kappa = .47$						
	Total	.45	.35	.15	.05	1.00
Case 2:	A	.20			.05	.25
Balanced	B	.05	.15		.05	.25
marginals	C			.20	.05	.25
	D		.10	.05	.10	.25
$\kappa = .53$						
	Total	.25	.25	.25	.25	1.00
Case 3:	A	.10		.10	.20	.40
Marginal	B		.25		.05	.30
asymmetry	C			.20		.20
	D				.10	.10
$\kappa = .56$						
	Total	.10	.25	.30	.35	1.00

Theorem 2 is a straightforward application of the Rearrangement Inequality. The result follows from using similar arguments as in the proof of the Kappa Paradox Theorem.

Theorem 2. *Consider several agreement tables that have the same proportion of observed agreement and the same marginals* $(p_{1+}, ..., p_{n+})$ *and* $(p_{+1}, ..., p_{+n})$*, but that differ in how the marginals are arranged over the categories. Then*

- *the table that is weakly marginal symmetric has the lowest value of kappa.*
- *the table that is marginal asymmetric has the highest value of kappa.*
- *the values of kappa for the other tables are between these two values.*

An illustration of Theorem 2 is presented in Table 2. Table 2 contains three hypothetical 4×4 agreement tables with categories A, B, C and D. In case 1, the table is weakly marginal symmetric. In case 3, the table is

Table 2: Values of kappa for three hypothetical cases. For all three cases the proportion of observed agreement $p_o = .55$. The values of the row and column marginals are the same in all three cases, but the column marginals are distributed differently over the categories.

		Judge 2				
	Judge 1	A	B	C	D	Total
Case 1:	A	.30	.05		.05	.40
Weak marginal	B	.05	.15	.10		.30
symmetry	C	.10	.05	.05		.20
	D	.05			.05	.10
$\kappa = .34$						
	Total	.50	.25	.15	.10	1.00
Case 2:	A	.10	.20	.10		.40
No symmetry	B		.25		.05	.30
No asymmetry	C		.05	.15		.20
	D	.05			.05	.10
$\kappa = .38$						
	Total	.15	.50	.25	.10	1.00
Case 3:	A	.10		.05	.25	.40
Marginal	B		.15		.15	.30
asymmetry	C			.20		.20
	D				.10	.10
$\kappa = .45$						
	Total	.10	.15	.25	.50	1.00

marginal asymmetric. In case 2, the table is neither symmetric nor asymmetric. In all three cases the proportion of observed agreement $p_o = .55$. The table with asymmetric marginals produces the highest value of kappa, whereas the table that is weakly marginal symmetric has the lowest value of kappa.

5. Discussion

This paper presents a formal proof of a paradox associated with Cohen's kappa, namely that, for fixed observed agreement between the judges, Cohen's kappa penalizes judges with similar marginals compared to judges who produce different marginals. Vach (2005) and Von Eye and Von Eye (2008) emphasize that kappa should not simply be interpreted as a measure of agreement, but that kappa expresses the degree to which observed agreement exceeds the agreement that was expected by chance. The paradox is a direct consequence of the definition of kappa and its aim to adjust the observed agreement with respect to the expected amount of agreement under chance conditions (Vach 2005, p. 659).

Various authors have proposed agreement measures that posses different properties compared to Cohen's kappa. For 2×2 tables (Martín Andrés and Femia-Marzo 2008; Warrens 2008a,c,d,e, 2009), alternatives to kappa are discussed in Cicchetti and Feinstein (1990) and Lantz and Nebenzahl (1996). For $n \times n$ tables, alternatives to kappa are discussed in Brennan and Prediger (1981), Aickin (1990), Martín Andrés and Femia Marzo (2004) and Gwet (2008). Vach (2005) argues to keep using Cohen's kappa.

It should be noted that the results in this paper are also relevant to the field of cluster analysis. Warrens (2008b) showed that in the special case of 2×2 tables, Cohen's (1960) kappa is equivalent to the Hubert-Arabie (1985) adjusted Rand index. The latter measure is the preferred statistic for comparing partitions from two different clustering algorithms (Steinley 2004). Warrens (2008a) derives what association coefficients for 2×2 tables become kappa after correction for chance. Some bounds of the 2×2 kappa are presented in Warrens (2008a, 2008e).

References

AICKIN, M. (1990), "Maximum Likelihood Estimation of Agreement in the Constant Predictive Model, and Its Relation to Cohen's Kappa," *Biometrics, 26,* 293–302.

BAKEMAN, R., QUERA, V., MCARTHUR, D., and ROBINSON, B.F. (1997), "Detecting Sequential Patterns and Determining Their Reliability with Fallible Observers," *Psychological Methods, 2,* 357–370.

BRENNAN, R.L., and PREDIGER, D.J. (1981), "Coefficient Kappa: Some Uses, Misuses, and Alternatives," *Educational and Psychological Measurement, 41,* 687–699.

BYRT, T., BISHOP, J., and CARLIN, J.B. (1993), "Bias, Prevalence and Kappa," *Journal of Clinical Epidemiology, 46,* 423–429.

CICCHETTI, D.V., and FEINSTEIN, A.R. (1990), "High Agreement but Low Kappa: II. Resolving the Paradoxes," *Journal of Clinical Epidemiology, 43,* 551–558.

COHEN, J. (1960), "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement, 20,* 213–220.

CONGER, A.J. (1980), "Integration and Generalization of Kappas for Multiple Raters," *Psychological Bulletin, 88,* 322–328.

DE MAST, J. (2007), "Agreement and Kappa-Type Indices," *The American Statistician, 61,* 149–153.

DOU, W., REN, Y., WU, Q., RUAN, S., CHEN, Y., BLOYET, D., and CONSTANS, J.-M. (2007), "Fuzzy Kappa for the Agreement Measure of Fuzzy Classifications," *Neurocomputing, 70,* 726–734.

GUGGENMOOS-HOLZMANN, I. (1996), "The Meaning of Kappa: Probabilistic Concepts of Reliability and Validity Revisited," *Journal of Clinical Epidemiology, 49,* 775–783.

GWET, K.L. (2008), "Computing Inter-rater Reliability and Its Variance in the Presence of High Agreement," *British Journal of Mathematical and Statistical Psychology, 61,* 29–48.

FEINSTEIN, A.R., and CICCHETTI, D. V. (1990), "High Agreement but Low Kappa: I. The Problems of Two Paradoxes," *Journal of Clinical Epidemiology, 43,* 543–549.

GOODMAN, L.A. (1991), "Measures, Models, and Graphical Displays in the Analysis of Cross-classified Data," *Journal of the American Statistical Association, 86,* 1085–1111.

HARDY, G.H., LITTLEWOOD, J. E., and PÓLYA, G. (1988), *Inequalities* (2nd ed.), Cambridge: Cambridge University Press.

HUBERT, L. (1977), "Kappa Revisited," *Psychological Bulletin, 84,* 289–297.

HUBERT, L.J., and ARABIE, P. (1985), "Comparing Partitions," *Journal of Classification, 2,* 193–218.

KRAEMER, H.C. (1979), "Ramifications of a Population Model for κ as a Coefficient of Reliability," *Psychometrika, 44,* 461–472.

KRAEMER, H.C., PERIYAKOIL, V.S., and NODA, A. (2004), "Tutorial in Biostatistics: Kappa Coefficients in Medical Research," *Statistics in Medicine, 21,* 2109–2129.

LANTZ, C.A., and NEBENZAHL, E. (1996), "Behavior and Interpretation of the κ Statistic: Resolution of the Paradoxes," *Journal of Clinical Epidemiology, 49,* 431–434.

LIPSITZ, S.R., LAIRD, N.M., and BRENNAN, T.A. (1994), "Simple Moment Estimates of the κ-Coefficient and Its Variance," *Applied Statistics, 43,* 309–323.

MARTÍN ANDRÉS, A. and FEMIA MARZO, P. (2004), "Delta: A New Measure of Agreement Between Two Raters," *British Journal of Mathematical and Statistical Psychology, 57,* 1–19.

MARTÍN ANDRÉS, A. and FEMIA MARZO, P. (2008), "Chance-corrected Measures of Reliability and Validity in 2×2 Tables," *Communications in Statistics, Theory and Methods, 37,* 760–772.

NELSON, J.C., and PEPE, M.S. (2000), "Statistical Description of Interrater Variability in Ordinal Ratings," *Statistical Methods in Medical Research, 9,* 475–496.

SIM, J., and WRIGHT, C.C. (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements, *Physical Therapy, 85,* 257–268.

STEINLEY, D. (2004), "Properties of the Hubert-Arabie Adjusted Rand Index," *Psychological Methods, 9*, 386–396.

THOMPSON, W.D., and WALTER, S.D. (1988), "A Reappraisal of the Kappa Coefficient," *Journal of Clinical Epidemiology, 41,* 949–958.

VACH, W. (2005), "The Dependence of Cohen's Kappa on the Prevalence Does not Matter," *Journal of Clinical Epidemiology, 58,* 655–661.

VON EYE, A., and VON EYE, M. (2008), "On the Marginal Dependency of Cohen's κ," *European Psychologist, 13,* 305–315.

WACKERLY, D.D., and ROBINSON, D.H. (1983), "A More Powerful Method for Testing Agreement Between a Judge and a Known Standard," *Psychometrika, 48,* 183–193.

WARRENS, M.J. (2008a), "On Similarity Coefficients for 2×2 Tables and Correction for Chance," *Psychometrika, 73,* 487–502.

WARRENS, M.J. (2008b), "On the Equivalence of Cohen's Kappa and the Hubert-Arabie Adjusted Rand Index," *Journal of Classification, 25,* 177–183.

WARRENS, M.J. (2008c), "On Association Coefficients for 2×2 Tables and Properties That Do Not Depend on the Marginal Distributions," *Psychometrika, 73,* 777–789.

WARRENS, M.J. (2008d), "On the Indeterminacy of Resemblance Measures for (Presence/Absence) Data," *Journal of Classification, 25,* 125–136.

WARRENS, M.J. (2008e), "Bounds of Resemblance Measures for Binary (Presence/Absence) Variables," *Journal of Classification, 25,* 195–208.

WARRENS, M.J. (2009), "k-Adic Similarity Coefficients for Binary (Presence/Absence) Data," *Journal of Classification, 26,* 227–245.

WARRENS, M.J. (2010), "Inequalities Between Kappa and Kappa-like Statistics for $k \times k$ Tables," *Psychometrika, 75,* 176–185.

ZWICK, R. (1988), "Another Look at Interrater Agreement," *Psychological Bulletin, 103,* 374–378.