Psychological Bulletin
1996, Vol. 119, No. 1, 159–165

0033-2909/96/$3.00

Measuring Ordinal Association in Situations That Contain Tied Scores

Richard Gonzalez
University of Washington

Thomas O. Nelson
University of Maryland at College Park

The construal of ties is critical for assessing the association between two variables. Ties should be excluded when the investigator's data-collection procedure forces ties to occur (e.g., a J-place rating scale is used to rate K items, with J < K; a criterion variable contains fewer than K possible outcomes per item). Four measures arising from excluding or including ties on 2 ordinal variables are Goodman & Kruskal's G, Somers's d_{yx}, Kim's $d_{y.x}$, and Wilson's e. In contrast to measures having variance-accounted-for interpretations, probabilistic interpretations developed here can be applied meaningfully both to ordinal-scaled variables and to stronger scales. Recommendations are offered for which measure to use in various situations.

Hypotheses in psychology about the association between two variables are typically ordinal in nature (e.g., one variable is hypothesized to increase monotonically as another variable increases or decreases), and many scales in psychology might best be treated as ordinal (for an elaboration, see Surber, 1984; Townsend & Ashby, 1984). However, the choice of a measure for the degree of ordinal association becomes problematic when ties occur in either (or both) of the variables (Kendall, 1955).

This article evaluates a family of four measures of ordinal association in situations that contain tied scores on one or the other (or both) of the two variables, presents new probabilistic interpretations for those measures, and offers a new algorithm for determining which measure to use in each situation. The occurrence of tied scores is frequent in diverse areas of psychology including cognitive, developmental, clinical, and social psychology, as well as other areas outside of psychology (e.g., business, marketing, meteorology, sociology, and law). The present evaluation, probabilistic interpretations, and algorithm can be applied in all of those situations.

A Specific Instantiation: Accuracy of Predictions

In explaining the nature of our approach and the role of ties for ordinal association in general, we refer to a particular situation in which the investigator desires to assess the accuracy of the ordinal aspects of people's predictions. We have selected this example because it provides a good illustration of how different kinds of ties can arise and should be treated, and because it builds on our earlier, related work (e.g., Nelson, 1984). The conclusions about the treatment of ties in this example, and the corresponding choice and interpretation of a measure, can easily be extended to any situation in which the degree of ordinal association between two variables is assessed. Thus, our algorithm is intended to be applied broadly, and other examples of potential applications are mentioned at the end of the article.

In the assessment of predictive accuracy, one variable is designated as the *predictor variable* and is comprised of some number of ordered levels; another variable is designated as the *criterion variable* and is comprised of some number of ordered levels; the two variables do not necessarily have the same number of levels. The ordering for the predictor variable can be derived from a person's paired comparisons or rankings (e.g., Nelson & Narens, 1980), from rating scales (e.g., Nelson, 1984), from numerical probability judgments[1] (e.g., Yates, 1990), or from nonnumerical probability judgments (e.g., Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986). Consider a judge attempting to predict his or her performance on each of several binary outcomes (e.g., performance at recalling the answers to general-information questions). The central issue is the degree to which the ordering on the predictor variable matches the ordering on the criterion variable, but frequently this is difficult to establish because of the presence of ties.

We focus exclusively on measures of ordinal association and, therefore, do not review issues surrounding tests of different models of independence (see Clogg & Shihadeh, 1994, for a review), nor do we review indices based on the log-linear model

Richard Gonzalez, Psychology Department, University of Washington; Thomas O. Nelson, Psychology Department, University of Maryland at College Park.

Research was supported by Grant SES9110572 from the National Science Foundation and by Grant R01-MH32205 and a career development award (K05-MH1075) from the National Institute of Mental Health. Some portions of the article were written at the University of Washington when we were both on that faculty, and other portions were written at the University of Wuerzburg in Germany when Thomas Nelson held an Alexander von Humboldt senior scientist award. We reported some of the findings at the European Mathematical Psychology meeting in Brussels, Belgium, in July 1992.

We thank A. Graf, J. Leonesio, L. Narens, and J. Treadwell for helpful comments on an earlier draft.

Correspondence concerning this article should be addressed to Richard Gonzalez, Psychology Department (NI-25), University of Washington, Seattle, Washington 98195. Electronic mail may be sent to gonzo@u.washington.edu or to tnelson@glue.md.edu.

[1] In the special case of numerical probability judgments, we deal with *resolution* (aka discrimination), which is the degree to which the overall performance being predicted differs across the various levels of ordered predictions. Some researchers also assess *calibration*, an index of how closely numerical probability estimates for events match the relative frequency of those events. Lichtenstein and Fischhoff (1977) concluded that resolution, in comparison with calibration, "is a more fundamental aspect of probabilistic functioning" (p. 181).

(e.g., Goodman, 1991). As explained below, the measures of ordinal association reviewed here are based on the notion of a qualitative comparison between a pair of items (or subjects), and we propose a straightforward probabilistic interpretation for each measure: the probability that if the value for Item A (or Person A) is greater than the value for Item B (or Person B) on the predictor variable, then the value of A is greater than the value of B on the criterion variable. This contrasts with the goal of model testing where the interest is in finding the best-fitting model according to a goodness-of-fit criterion.

Dyad as Unit of Analysis

The primitive concept for the present formulation is qualitative (see Narens & Luce, 1993, p. 128), in particular a qualitative comparison of two items, which is referred to as a *dyad* (Costner, 1965; Goodman & Kruskal, 1954; Kruskal, 1958; Nelson, 1984). A *dyad* is comprised of the joint outcome of one item versus another item on the predictor variable and on the criterion variable. For instance, one dyad is the joint outcome of (1) Item A being predicted to do better than Item B, and (2) Item A doing better than Item B on the criterion variable. The dyad is appropriate as the primitive concept for an analysis of ordinal association because it is the smallest possible unit involving a comparison of one item with any other item.

When the ordering between the two items on the predictor variable is consistent with the ordering of the same two items on the criterion variable (i.e., $A_p > B_p$ and $A_c > B_c$, where subscripts denote predictor [p] and criterion [c]), the dyad is a *concordance,* and the total number of concordances is designated as C. Of course, the judge might have rated Item A more likely to be recalled than Item B, yet subsequent recall was correct for Item B and not Item A (i.e., $A_p > B_p$ and $A_c < B_c$). This is a *discordance,* and the total number of discordances is designated as D.

Ties may occur on one or both variables. The dyad consisting of Item A and Item B may be tied on the predictor variable (i.e., $A_p = B_p$) but ordered on the criterion variable (e.g., $A_c > B_c$). Therefore, this dyad is "not tied on the criterion variable," and the total number of these dyads is designated as T_p. Conversely, a dyad may be tied on the criterion variable (i.e., $A_c = B_c$) but ordered on the predictor variable (e.g., $A_p > B_p$). This dyad is "not tied on the predictor variable," and the total number of these dyads is designated as T_c. Finally, a dyad may be tied on both the predictor and the criterion variables, and the total number of those dyads is designated as T_{pc}.

Equation 1 shows that the total number of dyads (the combination of K items taken two at a time) is a sum of the five kinds of dyads:

$$K(K-1)/2 = C + D + T_p + T_c + T_{pc}. \quad (1)$$

The decomposition into the five kinds of dyads is shown in the Appendix.

Table 1 shows hypothetical data for one person. The predictor variable in this example, predicted likelihood of recall, has three possible values: low, medium, and high. The criterion variable in this example, recall, has two possible values: correct or wrong. The eight items produce a total of 28 dyads. We now decompose these 28 dyads using the notation i & j to refer to the dyad between Item i and Item j. Ten of the 28 dyads are concordances: 2 & 3, 2 & 6, 3 & 5, 3 & 7, 4 & 5, 4 & 7, 5 & 6, 5 & 8, 6 & 7, and 7 & 8; only 1 dyad is a discordance: 1 & 2. The remaining 17 dyads have a tie on one or both variables. Four dyads are not tied on the criterion: 1 & 5, 1 & 7, 2 & 4, and 2 & 8; 10 dyads are not tied on the predictor: 1 & 3, 1 & 4, 1 & 6, 1 & 8, 2 & 5, 2 & 7, 3 & 4, 3 & 8, 4 & 6, and 6 & 8; and 3 dyads are tied on both the predictor and the criterion: 3 & 6, 4 & 8, and 5 & 7.

Table 1
Examples of Item Prediction and Their Correspondence to Criterion Performance

Item	Predictions	Criterion
1	High	Wrong
2	Medium	Correct
3	Low	Wrong
4	Medium	Wrong
5	High	Correct
6	Low	Wrong
7	High	Correct
8	Medium	Wrong

Note. An example of item prediction is predicted likelihood of recall. An example of criterion performance is correct versus wrong recall.

A Family of Four Measures

When ties do not occur, every dyad is either a concordance or a discordance, and the relative preponderance of concordances to discordances is summarized by

$$(C - D)/[K(K-1)/2], \quad (2)$$

which has the difference between the number of concordances and discordances in the numerator and the total number of dyads in the denominator. Equation 2 is equivalent to the index proposed by Kendall (1955), which he called tau. Under the special case when ties do not occur (i.e., T_c, T_p, and T_{pc} all equal 0), tau and the four measures reviewed below are all equivalent.

Kendall (1955) claimed that tau should not be used when ties occur. Although Kendall offered several alternatives to tau, none of them were on strong statistical footing or gained acceptance (Kruskal, 1958). The first accepted extension of Kendall's tau was proposed by Goodman and Kruskal (1954) in a measure they called gamma (see Kruskal, 1958, for a historical review). Gamma is defined as

$$G = (C - D)/(C + D), \quad (3)$$

where G is the sample estimate of gamma. Gamma handles the problem of ties by eliminating from consideration all dyads that contain ties. Freeman (1986) pointed out that "some writers have criticized gamma for the 'undesirable property' of restricting its calculations only to untied pairs. This property, however is neither a flaw nor a weakness" (p. 63). Indeed, the question of how to handle ties has been a central problem, and below we present an algorithm for deciding when tied dyads are relevant or irrelevant.

We do not consider in this review the rank correlation known as Spearman's rho. Spearman's rho, like tau, has difficulties with the way it handles ties (see Kendall, 1955); for example, for the situation shown in Table 1, Spearman's rho could not attain a value of +1.0, even if the three corrects and the five wrongs on the criterion variable were rearranged to maximize predictive accuracy.

Sociologists proposed three other measures of ordinal association that are related to gamma, so as to explore different definitions of monotonicity (Freeman, 1986; Wilson, 1974), to formulate asymmetric measures of ordinal association (Kim, 1971; Somers, 1962), and to use the property of proportion reduction in error (PRE) interpretation, which is a desirable property for measures of association and prediction (Freeman, 1986; Kim, 1971; Somers, 1962, 1968). These three other measures count ties as inaccuracies in the sense that relevant ties are included in the denominator.

For comparing these three indices, consider the general quantity

$$(C - D)/(C + D + T_i), \qquad (4)$$

where T_i is the number of ties considered to be relevant. Gamma defines $T_i = 0$. The measures proposed by Kim (1971), Somers (1962), and Wilson (1974), which are summarized in Table 2, differ in which tied dyads are consider relevant. Because the measures under consideration have the same numerator but differ in their denominators, their relative magnitudes are known a priori. Gamma is always greater than or equal to Kim's, Somers's, and Wilson's measures; Wilson's e is always less than or equal to gamma and to Kim's and Somers's measures (Freeman, 1986). To illustrate, the data in Table 1 yield the following: Goodman and Kruskal's (1954) G = .82, Kim's $d_{y.x}$ = .60, Somers's d_{yx} = .43, and Wilson's e = .36.

Probabilistic Interpretations

The interpretation of predictive accuracy can be formalized in terms of the following conditional probability developed by Nelson (1984, p. 112) for interpreting gamma: the likelihood that Item A is greater than Item B on the criterion variable, given that Item A is greater than Item B on the predictor variable, that is, $Pr(A_c > B_c | A_p > B_p)$. [Note: For situations in which either there are no ties or all ties are ignored, the conditional probability $Pr(A_c > B_c | A_p > B_p)$ and the opposite conditional probability $Pr(A_p > B_p | A_c > B_c)$ yield identical values.] Discussions of how this probabilistic interpretation can be a foundation for predictive accuracy occur in Nelson (1984), and examples of its use occur in Nelson, Leonesio, Landwehr, and Narens (1986) and in Nelson and Dunlosky (1991).

We now re-express $Pr(A_c > B_c | A_p > B_p)$ by the following equation to facilitate generalization to the other three measures under consideration:

$$P_g = Pr(A_c > B_c \;\&\; A_p > B_p \,|\, A \;\&\; B \text{ not tied on p or c}), \qquad (5)$$

which indicates the probability of a concordance given that all ties between A and B on either the predictor variable or the criterion variable are excluded. P_g denotes the probabilistic interpretation for gamma.

When the number of concordances is equal to the number of discordances (e.g., random responding), there is no predictive accuracy for a given set of items, and therefore $P_g = 0.5$ (Nelson, 1984). When predictive accuracy is perfect, $P_g = 1.0$.

Relationship Between the Four Measures and Their Probabilistic Interpretations

The measure G is linearly related to the conditional probability P_g by the formula $P_g = 0.5 + 0.5G$ (for derivations, see Nelson, 1984, p. 116; Kruskal, 1958, p. 822; Costner, 1965, p. 347). This linear relationship provides a straightforward interpretation of G (e.g., G = 0.50 corresponds to $P_g = 0.75$, or a probability of 0.75 for a concordance).

The other three measures of ordinal association (i.e., Kim's, 1971, $d_{y.x}$; Somers's, 1962, d_{yx}; and Wilson's, 1974, e) are also linearly related to a correspondingly defined conditional probability (for derivations, see the Appendix). Let the subscripts k, s, and w denote Kim's, Somers's, and Wilson's measures, respectively. Then, because Kim's $d_{y.x}$ includes in the denomina-

Table 2
Measures of Ordinal Association as a Function of Ties on the Predictor and the Criterion Variables

Criterion variable	Predictor variable: Exclude ties	Predictor variable: Include ties
Exclude ties	Goodman & Kruskal's (1954) G $(C - D)/(C + D)$	Kim's (1971) $d_{y.x}$ $(C - D)/(C + D + T_p)$
Include ties	Somers's (1962, 1968) d_{yx} $(C - D)/(C + D + T_c)$	Wilson's (1974) e $(C - D)/(C + D + T_c + T_p)$

Note. All four measures have the same numerator (the difference between concordances and discordances) but differ in the number of ties, if any, that appear in the denominator. The denominator is the sum of concordances, discordances, and relevant ties. C = frequency of concordances; D = frequency of discordances; T_c = frequency of dyads tied on only the criterion variable; T_p = frequency of dyads tied on only the predictor variable.

tor the dyads that are tied only on the predictor variable, the corresponding conditional probability is

$$P_k = \Pr(A_c > B_c \;\&\; A_p > B_p \mid A \text{ and } B \text{ not tied on } c), \quad (6)$$

and Kim's $d_{y.x}$ is linearly related to P_k by the formula

$$P_k = 0.5 + 0.5\, d_{y.x} - 0.5 \Pr(T_p), \quad (7)$$

where $\Pr(T_p)$ is the proportion of dyads that are tied on only the predictor variable.

Similarly, Somers's (1962) d_{yx} corresponds to the conditional probability

$$P_s = \Pr(A_c > B_c \;\&\; A_p > B_p \mid A \text{ and } B \text{ not tied on } p) \quad (8)$$

and is linearly related to P_s by

$$P_s = 0.5 + 0.5\, d_{yx} - 0.5 \Pr(T_c). \quad (9)$$

Previous investigators have advocated the use of Kim's (1971) $d_{y.x}$ and Somers's d_{yx} in situations when it is appropriate to penalize the judge for ties (Liberman & Tversky, 1993; Nelson, 1984, p. 112).

Finally, Wilson's (1974) e corresponds to the conditional probability

$$P_w = \Pr(A_c > B_c \;\&\; A_p > B_p \mid A \text{ and } B \text{ not tied on both p and c}) \quad (10)$$

and is linearly related to P_w by

$$P_w = 0.5 + 0.5\, e - 0.5[\Pr(T_p) + \Pr(T_c)]. \quad (11)$$

Algorithm for Deciding Which Ordinal Measure to Use

Consider the case when the judge states two items are identical on the predictor variable, and the judge *intends* the tie. By *intend,* we mean only that the procedure did not force the judge to respond with a tie, but nevertheless he or she did respond with a tie. Contrast this with another case in which the judge would like to place two items in different categories, but the response scale does not allow the judge an opportunity to express that difference. Thus, the procedure forces the judge to place in the same category the two items believed to be different. This occurs whenever the number of items exceeds the number of allowable points on the response scale. An example is shown in Table 1, where a 3-point scale is used to predict performance on each of the eight items; even if the judge believed all eight items to be different from one another, the belief could not be shown on a 3-point scale. It is always the case that when a J-point scale is used to evaluate K items with $J < K$, then the judge is not allowed to indicate that all items are untied, and therefore at least some ties are necessarily *forced by the procedure.*

A corresponding analysis of ties can be made on the criterion variable. A tie might occur because two items are, in fact, identical on the criterion variable. However, ties might also occur when potentially more categories underlie the criterion variable than are distinguished by the measuring instrument (e.g., whatever underlies the criterion variable is continuous, but the measurement is discrete, as in the case of "memory strength" underlying recall performance; see Table 1) or because the measuring instrument is insensitive at detecting the difference on whatever underlies criterion performance. These ties are also forced by the procedure.

When the experimenter knows that the ties were intended, there is no ambiguity as to how to interpret ties. If a judge gave two items the same prediction (and could have given them different predictions), but the two items differ on the criterion variable, then that dyad is an incorrect prediction. Such intended ties are *unambiguous* with respect to interpretation. When all ties on a particular variable can be interpreted as intended, then the incorrect ties should count against the judge's score of predictive accuracy.

However, when some of the ties might be forced by the procedure, then the experimenter has an epistemological problem, because any particular observed tie could have been either intended or forced by the procedure. Thus, the tie is *ambiguous,* and the judge's predictive accuracy score should not be penalized for such ties.

In summary, ties that are forced by the procedure are construed as ambiguous because a finer scale might not have yielded those ties. Only when an investigator can be sure that the judge intended the ties and when the ties on the criterion variable are not forced by the procedure can the tied dyads be legitimately construed as unambiguous and interpretable.

Table 3 shows the recommended measures that arise from the combinations of different patterns of ties on the predictor and criterion variables. The proposed algorithm is that *ties on a given variable that are ambiguous* (in the meaning described above) *should be excluded from the denominator because those dyads are uninterpretable. Only when all of the ties are unambiguous should the dyads containing those ties be included in the denominator.* That is, the judge is penalized for incorrect ties when those ties are interpretable, but the judge is not penalized for incorrect ties when they are ambiguous and might be due to only a procedure that forced the ties to occur. This recommendation is in accord with Costner's (1965) suggestion that "such tied pairs . . . are discarded as pairs for which order on at least one variable is indeterminate" (p. 346).

Accordingly, when both variables contain some ambiguous ties, Goodman and Kruskal's (1954) G should be used. When ambiguous ties occur, the designation of a concordance or discordance cannot be made for all the available data (although perhaps it could be made with more sensitive data-collection procedures). The measure G uses only the dyads that are interpretable with respect to strict orderings, namely, concordances and discordances.

When all of the ties on both variables are unambiguous, then the ties on the predictor variable and ties on the criterion variable should be included because the judge's predictions are unambiguously incorrect. Then Wilson's (1974) e should be used.

The arguments made above are easily extended: Exclude ties from any variable that contains ambiguous ties, and include ties from any variable that contains only unambiguous ties. This leads to Kim's (1971) $d_{y.x}$ when ties on the predictor variable (but not on the criterion variable) are unambiguous and to Somers's (1962) d_{yx} when ties on the criterion variable (but not on the predictor variable) are unambiguous.

Table 3
Recommended Measures of Predictive Accuracy (and their Probabilistic Interpretations) for Various Combinations of Ties on the Predictor and the Criterion Variables

Ties on the criterion variable	Ties on the predictor variable: Forced by procedure (exclude ties)	Ties on the predictor variable: Not forced by procedure (include ties)
Forced by procedure (exclude ties)	Goodman & Kruskal's (1954) G $Pr(A_c > B_c \& A_p > B_p \mid A \& B$ not tied on p or c)	Kim's (1971) $d_{y.x}$ $Pr(A_c > B_c \& A_p > B_p \mid A \& B$ not tied on c)
Not forced by procedure (include ties)	Somers's (1962, 1968) d_{yx} $Pr(A_c > B_c \& A_p > B_p \mid A \& B$ not tied on p)	Wilson's (1974) e $Pr(A_c > B_c \& A_p > B_p \mid A \& B$ not tied on both p and c)

Note. The upper entry in each cell is the measure of predictive accuracy, and the lower entry is the probabilistic interpretation for that measure. The relation between each measure, m, and its probabilistic interpretation, P_m, is linear by the equation $P_m = 0.5\,[1 - Pr(T_i) + m]$, where $Pr(T_i)$ is the proportion of dyads in the denominator of m that contains ties (see text). Pr = probability; p = predictor; c = criterion.

Perfect Ordinal Association

Whenever one of the variables contains fewer potential levels of measured fineness than the other, ties can occur that are forced by the procedure. Having different numbers of measurement categories on the two variables is a sufficient reason to exclude dyads that are tied on the variable with fewer categories but ordered on the variable with more categories (cf. Somers, 1962). Such tied dyads are ambiguous because they are forced by the procedure, and a finer categorization scheme might have broken the ties. This has obvious implications for situations where one of the two variables is a predictor variable and the other is a criterion variable: Whenever the predictor variable has fewer categories than the criterion variable, the investigator should choose between the two measures in column 1 of Table 3, with the choice being contingent on the interpretability of ties on the criterion variable (see Algorithm for Deciding Which Ordinal Measure to Use). Whenever the criterion variable has fewer categories than the predictor variable (e.g., Table 1), the investigator should choose between the two measures in row 1 of Table 3, with the choice being contingent on the interpretability of ties on the predictor variable (see same section as cited above).

Ideally, the interpretability of ties should be considered prior to the data collection. If the investigator does not want to elicit ambiguous ties from the person and does not want ambiguous ties to occur on either variable, then a different data-collection procedure may be needed. For example, a person could do a ranking task with the understanding that ties are allowed and will be interpreted as intended (e.g., Liberman & Tversky, 1993), or the investigator could use finer-grained performance tasks to reduce or eliminate ambiguous ties. Additional discussion of how problems of interpretability can be solved during data collection rather than during data analysis occurs in Blalock (1974) and in Nelson and Narens (1980).

Summary and Recommendations

The four measures reviewed above have a straightforward probabilistic interpretation and do not involve other potentially problematic formulations such as variance accounted for (e.g., Birnbaum, 1973; Surber, 1984). The proposed algorithm is that ties should be included only when they are unambiguously intended by the person and when the observed performance could have yielded no ties among the items, and that ties should be excluded whenever they are forced by the data-collection procedure. The data-collection procedure forces ties whenever there are not enough measurement categories on either (or both) of the two variables being correlated to assign every item a unique value on each variable. For instance, in the special case where the two variables consist of a predictor variable and a criterion variable, the data-collection procedure forces ties whenever there are not enough measurement categories for the judge to assign each to-be-judged item to a value on the predictor variable that is different from the value assigned to every other to-be-judged item, or whenever the investigator does not have a fine enough measurement of criterion performance to allow each item to have a unique level of criterion performance. Such situations are frequent (e.g., Table 1).

Easy computational formulas for all five kinds of dyads appear in articles by Kim (1971) and by Wilson (1974), and the measure G is available in the BMDP and SPSS statistical packages and in a stand-alone BASIC program (Nelson, 1986). Tests of significance for a single G or a single Somers's (1962) d_{yx} are given in a book by Siegal and Castellan (1988). Although the statistical analyses are conducted on a measure of association, the statistical conclusions extend directly to the probabilistic interpretation because these four measures of association are related to their corresponding probabilistic interpretations by a linear transformation (for rationale, see Townsend & Ashby, 1984).[2]

Research domains in which an investigator might apply the present algorithm and probabilistic interpretations to evaluate hypotheses about the degree of ordinal association are ubiquitous in cognitive psychology (e.g., Does subsequent memory performance increase as people's ratings of what they know increases?), developmental psychology (e.g., Does the fre-

[2] Statistical conclusions for the three measures of ordinal association $d_{y.x}$, d_{yx}, and e (in contrast to those for G) may differ when the measures are converted to their probability scores if people have different proportions of relevant ties. Even though the transformation between the measures $d_{y.x}$, d_{yx}, e, and their respective probabilistic interpretations is linear within subjects, each person's intercept is a function of the proportion of relevant tied dyads; if an investigator wants to compare the probabilistic interpretation across conditions for those three measures, the proportion of relevant ties can be used as a blocking variable. This is not a problem for G because all ties are excluded from its computation.

quency of some particular kind of response increase with age?), personality and individual-difference psychology (e.g., Do people who have more of this trait also have more of that trait?), clinical psychology (e.g., Does more of this treatment yield a greater likelihood of a cure for the problem?), industrial psychology (e.g., Do people who have more of this leadership trait also obtain greater productivity from their workers?), social psychology (e.g., Are people who have more of this attitude more likely to vote in an election?), sport psychology (e.g., Do players who practice more of this also perform better during a contest?), animal psychology (e.g., Do animals who eat more of this also show more of that behavior?), and educational psychology (e.g., Do students who receive more of this treatment when studying also obtain higher test scores?).

References

Birnbaum, M. H. (1973). The devil rides again: Correlation as an index of fit. *Psychological Bulletin, 79,* 239–242.

Blalock, H. M. (1974). Beyond ordinal measurement: Weak tests of stronger theories. In H. M. Blalock (Ed.), *Measurement in the social sciences* (pp. 424–455). Chicago: Aldine.

Clogg, C. C., & Shihadeh, E. S. (1994). *Statistical models for ordinal variables.* Thousand Oaks, CA: Sage.

Costner, H. L. (1965). Criteria for measures of association. *American Sociological Review, 30,* 341–353.

Freeman, L. C. (1986). Order-based statistics and monotonicity: A family of ordinal measures of association. *Journal of Mathematical Sociology, 12,* 49–69.

Goodman, L. A. (1991). Models, measures, and graphical displays in the analysis of contingency tables. *Journal of the American Statistical Association, 86,* 1085–1111.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association, 49,* 732–764.

Kendall, M. G. (1955). *Rank correlations methods* (2nd ed.). London: Charles Griffin.

Kim, J. O. (1971). Predictive measures of ordinal association. *American Journal of Sociology, 76,* 891–907.

Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American Statistical Association, 53,* 814–861.

Liberman, V., & Tversky, A. (1993). On the evaluation of probability judgments: Calibration, resolution, and monotonicity. *Psychological Bulletin, 114,* 162–173.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 20,* 159–183.

Luce, R. D. (1956). Semiorders and a theory of utility discrimination. *Econometrika, 24,* 178–191.

MacLane, S., & Birkhoff, G. (1988). *Algebra* (3rd ed.). New York: Chelsea.

Narens, L., & Luce, R. D. (1993). Further comments on the "nonrevolution" arising from axiomatic measurement theory. *Psychological Science, 4,* 127–130.

Nelson, T. O. (1984). A comparison of current measures of accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95,* 109–133.

Nelson, T. O. (1986). BASIC programs for computation of the Goodman–Kruskal gamma coefficient. *Bulletin of the Psychonomic Society, 24,* 281–283.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science, 2,* 267–270.

Nelson, T. O., Leonesio, R. J., Landwehr, R. S., & Narens, L. (1986). A comparison of three predictors of an individual's memory performance: The individual's feeling of knowing versus the normative feeling of knowing versus base-rate item difficulty. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 279–287.

Nelson, T. O., & Narens, L. (1980). A new technique for investigating the feeling of knowing. *Acta Psychologica, 46,* 69–80.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Somers, R. H. (1962). A new asymmetric measure of association for ordinal variables. *American Sociological Review, 27,* 799–811.

Somers, R. H. (1968). On the measurement of association. *American Sociological Review, 33,* 291–292.

Surber, C. F. (1984). Issues in using quantitative rating scales in developmental research. *Psychological Bulletin, 95,* 226–246.

Townsend, J. T., & Ashby, F. G. (1984). Measurement scales and statistics: The misconception misconceived. *Psychological Bulletin, 96,* 394–401.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General, 115,* 348–365.

Wilson, T. P. (1974). Measures of association for bivariate ordinal hypotheses. In H. M. Blalock (Ed.), *Measurement in the social sciences* (pp. 327–342). Chicago: Aldine.

Yates, J. F. (1990). *Judgment and decision making.* Englewood Cliffs, NJ: Prentice-Hall.

Appendix

Composition of Dyads and the Relation Between the Four Measures of Ordinal Association and Their Conditional Probabilities

We illustrate the decomposition of K objects into dyads. The objects represent the unit of analysis (e.g., items, subjects). For K objects, there are $K(K - 1)/2$ unique dyads. Assuming the trichotomy law holds (MacLane & Birkhoff, 1988), one of the following relations must be true for each pair of objects on the predictor variable: $A > B$, $A = B$, or $A < B$, and one of these relations must also be true for each pair of objects on the criterion variable. For example, two items are either ordered or judged as equivalent (e.g., indifference) on the predictor variable, and the same two items are either ordered or judged as equivalent on the criterion. These relations can be either qualitative or quantitative. In this article, we do not discuss the issues underlying indifference but refer the reader to Luce's (1956) treatment of semiorder.

The $K(K - 1)/2$ unique dyads can be decomposed into Table A1, a 3 × 3 table of frequencies (the cell entries are frequencies).

The frequency of concordances (C), discordances (D), ties only on the predictor variable (T_p), ties only on the criterion variable (T_c), and ties on both predictor variable and criterion variable (T_{pc}) are given by the sums of the relevant entries in Table A1. Specifically,

$$C = a + i$$
$$D = c + g$$
$$T_p = b + h$$
$$T_c = d + f$$
$$T_{pc} = e.$$

Note that all four measures reviewed in this article ignore the cell e, the frequency T_{pc} (tied on both variables).

The four measures of ordinal association reviewed in this article have the same numerator $(C - D)$ but different denominators. In general, the denominator can be expressed as $C + D + T_i$, where T_i is the number of relevant ties. For example, when computing the measure G, $T_i = 0$ because the measure G does not include any dyads with ties. When computing Wilson's (1974) e, $T_i = T_p + T_c$. Thus, a general expression for the four measures of ordinal association considered here can be written as $m = (C - D)/(C + D + T_i)$, and a general expression for the probability of a concordance for measure m is $C/(C + D + T_i)$.

Table A1
A 3 × 3 Table of Frequencies

Criterion	Prediction $A_p > B_p$	$A_p = B_p$	$A_p < B_p$
$A_c > B_c$	*a*	*b*	*c*
$A_c = B_c$	*d*	*e*	*f*
$A_c < B_c$	*g*	*h*	*i*

Note. c = criterion; p = prediction.

The measure m is linearly related to P_m. From the definition of m, we derive

$$\begin{aligned} m &= (C - D)/(C + D + T_i) \\ &= C/(C + D + T_i) - D/(C + D + T_i) \\ &= P_m - D/(C + D + T_i) \\ &= P_m - [1 - (C + T_i)/(C + D + T_i)] \\ &= 2P_m - 1 + \Pr(T_i) \end{aligned}$$

The term $\Pr(T_i)$ is the proportion of ties considered relevant by the particular measure.

Rearranging terms yields a general relationship between P_m and the measure of ordinal association m: $P_m = 0.5 + 0.5\,m - 0.5\Pr(T_i)$.

Received July 7, 1993
Revision received June 30, 1995
Accepted July 5, 1995 ■