Political Analysis, 8:3

Item Similarity in Scale Analysis

Marco R. Steenbergen
University of North Carolina at Chapel Hill

A statistic—the similarity coefficient—is developed for assessing the property that a set of scale items measures one and only one construct. This statistic is rooted in an explicit measurement model and is flexible enough to be used in exploratory scale analyses, even in small samples. Methods for analyzing similarity coefficients are described and illustrated in analyses of Stimson's (1991) policy mood data and Markus' (1990) popular individualism items. The Appendix discusses the statistical properties of similarity coefficients.

1 Introduction

SUMMATED RATING SCALES play an important role in political analysis, offering researchers well-known advantages of greater reliability and precision in measurement (DeVellis 1991; Spector 1992). However, the validity of such scales hinges on an important attribute: the extent to which the items in a scale measure one and only one construct. This attribute, which is often referred to as unidimensionality, prevents confounding between scales (Anderson and Gerbing 1982, 1988; Burt 1976; Cook and Campbell 1979).

Political scientists can use a number of statistical tools to test for unidimensionality, of which confirmatory factor analysis is perhaps best known (see Anderson and Gerbing 1982, 1988; DeVellis 1991; Gerbing and Anderson 1988; Hunter and Gerbing 1982; Spector 1992). These methods, however, have important limitations. First, they usually perform best in large samples, which makes it difficult to apply them to the small pilot studies that are frequently used to assess unidimensionality. Second and more important, extant methods like confirmatory factor analysis require a good theoretical sense of how a set of items hangs together. In the absence of such knowledge, or when original theoretical expectations are not supported by the data, it is often difficult to determine structure among the items.

This paper presents an alternative tool for the determination of unidimensionality in summated rating scales—*similarity coefficients*. This tool has two attractions. First, it allows for an exploratory search for structure among items that is rooted in an explicit measurement model. Second, the tool can be used in very general circumstances, including conditions of sparse data. As such, similarity coefficients are a powerful instrument for conducting scale analyses. This is particularly true for pilot research, in which small samples and vaguely

Author's note: A previous version of this paper was prepared for the 52nd annual meeting of the Midwest Political Science Association, Chicago, 1994. I would like to express my gratitude to Stanley Feldman, George Rabinowitz, and Elaine Sieff for their comments on later drafts of this paper.

Copyright 2000 by the Society for Political Methodology

formulated ideas about item structure are common. However, the appeal of similarity coefficients goes beyond the concerns of pilot research because these coefficients can also be applied fruitfully in large sample analyses of items, especially in performing specification searches. This paper presents examples of both types of application of similarity coefficients.

The contribution of this paper is threefold. First, it introduces to political analysis a methodology that has been fruitfully employed by other areas of social science for over 25 years. Second, the paper adds to this methodology by introducing a new type of similarity coefficient and by describing various methods for analyzing similarity coefficients. Finally, this paper makes available user-friendly software for conducting analyses of unidimensionality through similarity coefficients.

The remainder of this paper is organized as follows. I begin with a brief review of the prerequisites for unidimensionality as rooted in classical test theory. Next, I motivate similarity coefficients by describing conventional methods of scale analysis and their limitations. I then proceed by discussing two types of similarity coefficient as well as methods for their analysis. This is followed by two examples of the use of similarity coefficients in political research. The paper concludes with several practical recommendations. The Appendix provides details about the robustness of similarity coefficients.

2 Prerequisites for Unidimensionality

To understand similarity coefficients we need to understand the precise meaning of unidimensionality. Scale items are unidimensional if they satisfy two conditions (Anderson and Gerbing 1982, 1988; Hunter and Gerbing 1982). First, the items should hang together, in the sense that they measure a common construct or dimension—a property called *internal consistency*. Second, no item should tap more than one construct—a property called *external consistency*.

2.1 *Consistency of Scale Items*

Internal consistency implies that the correlation between two scale items can be attributed entirely to their association with a common construct. That is,

$$\rho_{x_i x_j} = \rho_{x_i \xi} \rho_{x_j \xi}$$

(for all i, j), where x_i and x_j are two scale items, ξ is the construct that these items measure, and ρ denotes the correlation (Anderson and Gerbing 1982, 1988; Gerbing and Anderson 1988; Hunter and Gerbing 1982).

External consistency implies that the correlation between items from different scales is proportional to the correlation between the constructs that these items measure. That is,

$$\rho_{x_i y_j} = \rho_{x_i \xi} \rho_{\xi \eta} \rho_{y_j \eta}$$

(for all i, j), where x_i is a measure of construct ξ and y_j is a measure of construct η. This result suggests that no correlation should exist between the items from different scales other than that caused by the correlation between their underlying constructs (Anderson and Gebring 1982, 1988; Gerbing and Anderson 1988; Hunter and Gerbing 1982; Tucker 1940).

2.2 *Consistency and Psychometric Theory*

The conditions of internal and external consistency are closely tied to the classical test model in psychometrics.[1] This model attributes item scores to two sources: a true score and measurement error. The true score is the quantity of interest to the researcher; it is an error-free measure of a construct. Items are formulated to obtain insight about this true score. Items are not pure measures of the construct, however, because they are contaminated by measurement error, i.e., idiosyncratic features of each item that taint responses. Thus, the responses to an item are driven by the true score as well as by measurement error. In any single item it is generally impossible to disentangle these two components. By making certain assumptions, however, classical test theory permits inferences about the true score on the basis of a series of items (or replications of the same item). Moreover, with the aid of these assumptions it is possible to draw conclusions about the correlations between items (see, e.g., Lord and Novick 1968), and this is important for assessing the consistency of the items.

The most fundamental assumptions in classical test theory concern measurement error. Specifically, the classical test model assumes that measurement error is not systematic and is idiosyncratic, i.e., specific to an item. Because measurement errors are assumed to be nonsystematic, the expectation is that they average out to 0. And because measurement errors are assumed to be idiosyncratic, they should be independent across items[2] and unrelated to the true score (see, e.g., Lord and Novick 1968).

To formalize these ideas I use the notation from the previous section. Let x_i denote a particular item and let ξ denote the underlying construct (true score). Further, let δ_i denote random measurement error. Finally, assume without loss of generality that x_i, ξ, and δ_i are standardized. Then the classical test model can be formulated as

$$x_i = \lambda_i \xi + \delta_i$$

where $\lambda_i = \rho_{x_i \xi}$ is the square root of the reliability of x_i. Consistent with the notion that measurement error is idiosyncratic and nonsystematic, the model assumes that (1) $E[\delta_i] = 0$, (2) $E[\xi \delta_i] = 0$, and (3) $E[\delta_i \delta_j] = 0$ (for $i \neq j$).

With these assumptions it is possible to draw inferences about true scores from an individual's mean response across a series of items (specifically, $\xi = E[x_i]/E[\lambda_i]$). More important for our purposes, it can be easily demonstrated that classical test items are internally and externally consistent. Specifically, for two classical test items x_i and x_j of the same construct we have $\rho_{x_i x_j} = \lambda_i \lambda_j = \rho_{x_i \xi} \rho_{x_j \xi}$. Moreover, for two classical test items x_i and y_j from two constructs, we have $\rho_{x_i y_j} = \lambda_i \lambda_j E[\xi \eta] = \rho_{x_i \xi} \rho_{\xi \eta} \rho_{y_i \eta}$. A proof of these results follows from assumptions (2) and (3) and is given by Hunter (1973).

These results are important because they provide a handle on assessing unidimensionality in scales. As we have seen, unidimensionality requires that items are internally and

[1] Useful introductions to classical test theory can be found in Crocker and Algina (1986), McDonald (1999), and Suen (1990). More detailed treatments of the theory are given by Guilford (1954), Gulliksen (1987), and Lord and Novick (1968).

[2] This independence assumption is violated when there are methods effects that affect items in similar ways and thus cause measurement errors across items to be correlated. Such methods effects can be accommodated within the classical test framework by introducing methods factors (see Andrews 1984; Scherpenzeel and Saris 1997). The presence of these methods factors can be detected through similarity coefficients as long as substantive factors and methods factors are not confounded (as would be the case, for example, when all items are measured in identical ways). In this case, similarity coefficients tend to be attenuated by methods effects.

externally consistent. We now have established that classical test items meet these consistency requirements. Thus, if we know that a set of items behaves in accordance with the classical test model, then we also know that these items are unidimensional. Consequently, it becomes possible to evaluate the unidimensionality of a scale by checking whether the items in that scale are classical test items. This is the logic that drives much of covariance structure analysis of scale items. It is also the logic that I follow in developing similarity coefficients in this paper.

As a final note, it is important to point out that classical test items come in various forms, depending on the additional assumptions that one makes about the items. The model formulated above is a model for *congeneric* items and makes relatively few assumptions about the items. That is, congeneric items can have different means, different loadings, and different amounts of measurement error. I also consider a more restrictive version of the classical test model in which the items are assumed to be equally reliable measures. I refer to such items as *semiparallel*.[3] The distinction between congeneric and semiparallel items is of little importance at this point, but it will play an important role in the development of different similarity coefficients.

3 Limitations of Conventional Scale Methodologies

The idea of relying on classical test theory in assessing unidimensionality is not new. Indeed, this idea lies at the heart of covariance structure analysis of scale items. This methodology begins with a particular specification of the classical test model, which is assumed to capture the structure of the scale items. This specification is then put to empirical test by assessing the fit of the model to the data. The appeal of this approach is that it provides a direct test of the classical test model.

If covariance structure analysis is a feasible approach to scale analysis, then why is it necessary to consider the alternative of similarity coefficients? The answer is that covariance structure analysis has important limitations. First, the method is sample-intensive, and when sample sizes become too small (e.g., $n < 100$), results may be incorrect.[4] The problem, of course, is that much of scale analysis takes place in pilot studies, and sample sizes in this type of research can be rather small.

Second, researchers may enter a scale analysis with poorly defined expectations about the structure of the items. Rather than testing models, these researchers may wish to *explore* how a set of items hang together and whether those items form a unidimensional scale. This exploratory focus is most likely to occur in pilot research, especially when a researcher is

[3]The term semiparallel is not used in classical test theory but is introduced here to indicate a variation on the well-known class of *parallel* items. Items are parallel when they have equal means, error variances, and loadings (see Lord and Novick 1968). The latter two requirements imply equal item reliabilities and are assumed to hold in semiparallel items. Unlike parallel items, however, semiparallel items may have different means. Thus, semiparallel items approximate truly parallel items in important ways but deviate from them because their means are not necessarily equal. Another way to conceptualize semiparallel items is that they contain parallel items as a special case (that arises when the item means happen to be equal).

[4]Of course, much depends on the specifics of the data. If items have powerful loadings on their underlying constructs, this is likely to be detected in a covariance structure analysis even if the sample size is small (see Saris and Satorra 1993). In other cases, however, serious problems in estimation may occur. Among the problems discussed in the literature are nonconvergence and improper solutions (Anderson and Gerbing 1984; Boomsma 1982, 1985), as well as bias in the estimated factor loadings and standard errors (Anderson and Gerbing 1984; Benson and Fleishman 1994; Boomsma 1982, 1985; Dolan 1994; Wolins 1995). Most of these results have been obtained through Monte Carlo simulations of the behavior of normal-theory maximum-likelihood estimators, but there is evidence that other estimators may also behave poorly in small samples (Benson and Fleishman 1994; Dolan 1994; see also Jöreskog 1990).

analyzing items from closely related constructs, in which case it may be difficult to specify a priori how the items relate to each construct. A confirmatory approach like covariance structure analysis may not be a natural choice in this situation.

Finally, covariance structure analysis of items may be insufficient when such an analysis shows that a researcher's original expectations about item structure are incorrect. In this case, the researcher will have to engage in a search for an alternative specification of the model, at least if she wants to understand how the items are related now that the original hypothesis about item structure has been rejected. Standard tools for performing such specification searches include modification indices and expected changes in model parameters, but these tools are limited in what they can accomplish (see Bollen 1989). Most importantly, modification indices and expected changes in parameters are limited to identifying nested models. That is, they suggest models that contain the original model and elaborate on it. If one wishes to explore nonnested models, however, standard approaches to model modification will not be very helpful and other exploratory tools are needed.

None of these observations imply that researchers should stop relying on covariance structure analysis for assessing unidimensionality. Rather, the point is that researchers often need an exploratory tool to use in addition to, and in conjunction with, covariance structure analysis. This paper introduces and develops such a tool: similarity coefficients.

Before doing so, however, I should address why extant exploratory methods of scale analysis are not satisfactory. Reliability analysis and exploratory factor analysis have been the workhorses of scale analysts for many years, and these methods are exploratory in character. Why, then, do we need another exploratory tool? The reason is that these other methods are insufficient for assessing unidimensionality.

Reliability analysis has the advantage that it is rooted in classical test theory (see, e.g., Crocker and Algina 1986; Suen 1990). However, this methodology is restricted to assessing internal consistency because only items from a single scale are considered. As we have seen, internal consistency is only half the story. Hence, it is not possible to adequately address unidimensionality solely with the help of reliability statistics such as the ubiquitous Cronbach's alpha.

Exploratory factor analysis can handle the exploration of items from multiple scales, but here the problem is that the methodology is not rooted in a measurement model such as the classical test model (Saris and Hartman 1990; Saris and Van Den Putte 1989). Therefore, this approach provides a very indirect test of unidimensionality and can even be misleading.

This point is worth elaborating. Factors in an exploratory factor analysis are statistical constructs, not psychometric constructs. They are the weighted sum of *all* items in an analysis, not just those items that are assumed to measure a particular construct. Factors, then, cannot be equated with constructs. But constructs and their relationship with items are critical for internal and external consistency, and hence exploratory factor analysis is a problematic tool for assessing unidimensionality (Gerbing and Anderson 1982). Indeed, Gorsuch (1983, p. 356), who is generally favorably disposed toward exploratory factor analysis of scale items, complains that "a common error is to fail to consider the interaction of the selection of initial variables and the definition of the factor(s)." Thus, if only a subset of items is supposed to measure a particular construct, factor analysis of all items and interpretation of factors in terms of all of the items may be misleading. Alas, this is the conventional approach to exploratory factor analysis in scale analysis.

In conclusion, despite an impressive "tool chest" for scale analysis that is available to researchers, there is a need for methods that are rooted in an explicit measurement model

and yet can be employed in exploratory work, even if this work involves relatively small samples. Such methods exist, but they have found relatively little application, especially in political analysis. For more than 25 years, researchers in other fields have employed similarity coefficients to aid in scale analysis. It is time to introduce this methodology to political analysis and to develop it further. This is what the remainder of this paper seeks to accomplish.

4 Similarity Coefficients

Similarity coefficients are based on an important statistical consequence of unidimensionality, namely, that classical test items should be correlated in similar ways with other items, both from the same scale and from other scales. Specifically, the correlation patterns of such items should be identical up to a factor that depends only on the ratio of the square roots of the reliabilities of the items. For example, let x_i and x_j be two classical test items for a common construct and, without loss of generality, let x_k be another classical test item for that construct. It is known that classical test items are internally consistent, so that $\rho_{x_i x_k} = \rho_{x_i \xi} \rho_{x_k \xi}$ and $\rho_{x_j x_k} = \rho_{x_j \xi} \rho_{x_k \xi}$. It is then easily verified that $\rho_{x_i x_k} = (\rho_{x_i \xi} / \rho_{x_j \xi}) \rho_{x_j x_k}$, which can also be written $\rho_{x_i x_k} = (\sqrt{\lambda_i^2} / \sqrt{\lambda_j^2}) \rho_{x_j x_k}$, where λ_i^2 and λ_j^2 are the reliabilities of x_i and x_j, respectively.[5] One sees, then, that the correlational patterns of x_i and x_j are very similar, one pattern being an exact linear transformation of the other.

This correlational similarity can be used to construct a statistical measure of unidimensionality. This measure quantifies the similarity of two sets of correlations involving different items. The greater the similarity, the better the items fit the classical test model, and the more valid the conclusion is that they form a unidimensional scale. The question is how one should quantify item similarity. I present two procedures. The first of these procedures was developed by Hunter (1973) and assumes congeneric items. The second procedure is a novel contribution of this paper and consists of simplifying Hunter's coefficient by making the assumption that the items have equal reliabilities.

4.1 *Hunter's Similarity Coefficient*

Hunter (1973) developed a similarity coefficient that is the square root of the index of proportionality found in cluster analysis (Tryon 1939). This similarity coefficient, φ, has been applied mostly in the fields of consumer research (Anderson and Gerbing 1982; Gerbing and Anderson 1988; Hunter and Gerbing 1982) and psychology (Hunter et al. 1982). The major appeal of the coefficient is that it makes only mild assumptions about the items, namely, that they are congeneric items.

4.1.1 Constructing the Coefficient

We can develop Hunter's similarity coefficient in two ways. Hunter (1973) relied on scalar notation. Assume that a total of P items is analyzed in the scale analysis. These items may come from a single scale, or they may be drawn from multiple scales that are intended to measure different constructs. Then the similarity between two items, x_i and x_j, can be

[5] An identical result would have held had x_k been a measure of a different construct.

formulated as

$$\varphi_{x_i x_j} = \frac{\sum_{p=1}^{P} \rho_{x_i z_p} \rho_{x_j z_p}}{\sqrt{\sum_{p=1}^{P} \rho_{x_i z_p}^2} \sqrt{\sum_{p=1}^{P} \rho_{x_j z_p}^2}}$$

Here z_p $(p = 1, \ldots, P)$ refers to all of the items that are considered in the analysis, including x_i and x_j. It is important to note that Hunter (1973; Hunter and Gerbing 1982) defined the correlation of an item with itself (i.e., $\rho_{x_i x_i}$) as the item reliability or communality. I return to this definition later in this section.

An alternative way to develop φ, which is computationally more attractive, is to rely on matrix notation. To this effect consider a correlation matrix **R**. This matrix brings together the correlations between all P items in the analysis. Each column in the matrix defines the correlations of a particular item with all other items, as well as the reliability or communality of the item. These reliabilities or communalities are found on the diagonal of the correlation matrix. Thus, **R** can be partitioned into as many vectors as there are items in the analysis. Each vector specifies the correlational pattern for a particular item.

Let $\mathbf{r_i'} = (\rho_{i1}\, \rho_{i2} \ldots \rho_{iP})$ and $\mathbf{r_j'} = (\rho_{j1}\, \rho_{j2} \ldots \rho_{jP})$ define two of these vectors, corresponding to two items, i and j. (For the sake of simplicity, I refrain from identifying items by the more cumbersome notation of x_i and y_j in the remainder of this paper.) These items may belong to the same scale, or they may belong to different scales. Then Hunter's similarity coefficient between i and j can be defined as

$$\varphi_{ij} = \mathbf{r_i'} \mathbf{r_j} [\|\mathbf{r_i}\| \|\mathbf{r_j}\|]^{-1} \tag{1}$$

Here $\mathbf{r_i'} \mathbf{r_j}$ is the inner product of the correlation vectors of items i and j. This yields the same result as the numerator of Hunter's formula presented earlier. Further, $\|\mathbf{r_i}\| = \sqrt{\mathbf{r_i'} \mathbf{r_i}}$ is the norm of vector $\mathbf{r_i}$. This term is equivalent to the first term in the denominator of Hunter's formula. The second term in the denominator of that formula is represented by $\|\mathbf{r_j}\|$, the norm of vector $\mathbf{r_j}$.

4.1.2 Interpretation

It can be demonstrated that φ is bounded between -1 and 1. At these extremes two items possess correlational patterns that are identical up to a linear transformation. Such items possess correlational patterns that are themselves perfectly correlated.[6] As such, the items perfectly satisfy the requirements of internal and external consistency and hence constitute a unidimensional scale.[7] It is easily verified that these extreme values arise when the items are *congeneric* (see Hunter 1973).

A value of 0 on φ implies that the correlational patterns for two items are dissimilar. This suggests that the two items are not internally and externally consistent and should not be placed together in the same scale. Such items cannot be considered classical test items for the same construct.

[6] Hunter's similarity coefficient may be interpreted as a correlation between vectors of correlations only at the extremes. That is, when $|\varphi_{ij}| = 1$ this indicates perfect correlation between the correlation vectors $\mathbf{r_i}$ and $\mathbf{r_j}$. However, when $|\varphi_{ij}| < 1$ a correlational interpretation is no longer appropriate.

[7] Negative values for φ imply unidimensionality only if two items are scaled in opposite directions. If the items are scaled in the same direction, negative values for φ indicate severe problems with unidimensionality in a scale.

In general, one would want φ to be high for two items included in the same scale. In addition, one would want φ to be higher for two items in the same scale than for two items from different scales (unless the underlying constructs are perfectly correlated). If these patterns hold, there is evidence that a unidimensional scale can be formed.

4.1.3 Limitations

The computation of φ requires knowledge of item reliabilities or communalities, which we usually do not have. Hunter (1973, p. 60) recognized this problem and suggested to "put on the diagonal [of **R**] whatever seems appropriate and ask if the similarity is 'high.'" The justification for this trial-and-error approach is that the diagonal elements of **R** generally play only a minor role in determining the numerical value of φ.

Hunter's recommendation leaves researchers in somewhat of a quandary, however. While the diagonal elements of the correlation matrix are not very influential on the similarity coefficients, some choices of these elements are clearly better than others, in that they are closer to the true reliabilities or communalities. However, since the true reliabilities or communalities are generally unknown,[8] one cannot be sure that a particular specification of diagonal elements is better than some other specification. Thus, an element of ambiguity is introduced into Hunter's similarity coefficient. This problem is sufficiently severe to ask whether alternatives exist.

4.2 *A Semiparallel Similarity Coefficient*

An alternative approach is possible if one makes additional assumptions about scale items. Specifically, if it is assumed that two items are equally reliable, the requirements of internal and external consistency imply that the correlational patterns for these items are completely identical. In this case, similarity of the correlational patterns no longer depends on item reliabilities, so it is unnecessary to use reliability information in the computation of similarity coefficients. Consequently, the diagonal elements of **R** can be ignored altogether. This leads to a new type of similarity coefficient, which I dub the *semiparallel similarity coefficient* because it achieves its maximum value for semiparallel items.

4.2.1 Constructing the Coefficient

To construct the semiparallel version of Hunter's similarity coefficient the correlation matrix needs to be transformed by removing the diagonal elements. This results in the following correlation vectors for two items, i and j: $\mathbf{q}'_{\mathbf{i}} = (\rho_{ij}\mathbf{r}^{*\prime}_{\mathbf{i}})$ and $\mathbf{q}'_{\mathbf{j}} = (\rho_{ji}\mathbf{r}^{*\prime}_{\mathbf{j}})$. Here $\rho_{ij} = \rho_{ji}$ is the correlation between i and j. Further, $\mathbf{r}^{*\prime}_{\mathbf{i}}$ is the vector of correlations of item i with all other items in the analysis, except j (and, of course, i itself). The vector $\mathbf{r}^{*\prime}_{\mathbf{j}}$ is defined in a similar way for item j.

[8] The problem of estimating item reliabilities can be alleviated to some extent by relying on the methodology proposed by Andrews (1984) and Scherpenzeel and Saris (1997). These authors present procedures for inferring reliability information for new items from item characteristics such as response modality and item placement. Much is known about the influence of these characteristics on reliability, so that educated guesses about the diagonal elements of the correlation matrix can be made. The procedures sketched by these authors should be used more commonly in political analysis. However, it is important to note that these procedures lead to better guesses of item reliabilities, not to exact information about those reliabilities.

We can now compute a parallel similarity coefficient by applying

$$\psi_{ij} = \mathbf{q_i'q_j}[\|\mathbf{q_i}\|\|\mathbf{q_j}\|]^{-1} \tag{2}$$

where ψ_{ij} is the sp. sim. coeff. between items i and j, and the terms on the right-hand side of (c) the different terms have the same meaning as in Eq. (1). Like φ, this semiparallel similarity coefficient is bounded between -1 and 1. However, these extremes arise for *semiparallel* items, not necessarily for congeneric items (see the Appendix for a proof).

4.2.2 Limitations

At first sight, it may appear problematic that the ψ-values of congeneric items may not reach 1. After all, such items satisfy the consistency requirements discussed earlier. However, ψ is sensitive to violations of the assumption that items are equally reliable, and such violations often occur in congeneric items. Fortunately, this sensitivity is weak, as the Appendix shows. It takes major violations of the assumption that items have equal reliabilities to bring ψ down so far that one might erroneously conclude that two consistent items are too dissimilar to be placed together in a scale.

4.3 *Comparison*

The choice between Hunter's similarity coefficient and the semiparallel similarity coefficient comes down to the following question: Is a researcher better off risking an erroneous choice of item reliabilities or risking an incorrect assumption of equal item reliabilities? It is impossible to answer this question in general because the nature of scale items varies greatly. However, a series of Monte Carlo simulations reveals that the semiparallel similarity coefficient usually compares favorably with Hunter's similarity coefficient as calculated with different guesses of item reliabilities. Only when these reliability guesses are close to the truth will Hunter's similarity coefficient outperform the semiparallel similarity coefficient. It is important to note that this result holds even when the assumption of equal item reliabilities is severely violated. Details about these simulation results can be found at this journal's web site.

In light of these results, it appears that ψ is a good substitute for φ as a diagnostic of unidimensionality. It has the advantage of simplicity: one does not have to speculate about item reliabilities or make arbitrary decisions about those reliabilities. This advantage often compares favorably to the burden of making an additional assumption about the data, namely, that items are equally reliable.

4.4 *Implementing Similarity Coefficients*

Implementation of Hunter's similarity coefficient is straightforward and can be accomplished through standard matrix operations. The implementation of the semiparallel similarity coefficient is more complicated because the removal of diagonal elements of the correlation matrix requires extensive rearrangement of the elements of this matrix. Once the correlation matrix has been rearranged, computation of the parallel similarity coefficients again involves standard matrix operations.

The tasks of rearranging the correlation matrix and performing the required matrix operations have been automated in my computer program UniSim (Steenbergen 1999). UniSim computes both Hunter's similarity coefficient and the semiparallel similarity coefficient on the basis of a correlation matrix specified by the user. In addition, the program computes

the descriptive measures that are introduced later in this paper. The program UniSim is available as freeware from this journal's web site.

4.5 *Conclusion*

In this section, two similarity coefficients—φ and ψ—were described. What makes these coefficients so attractive is that they are rooted in the classical test model, which satisfies the requirements of internal and external consistency. This model is not directly tested, however, as is done in covariance structure analysis. Rather, similarity coefficients consider the statistical implications of classical test theory for the correlations of scale items with other items. Computation of these correlations does not require estimators whose properties are defined asymptotically, requiring large samples, as is the case with covariance structure analysis.[9] On the contrary, unbiased estimators of Pearson product–moment correlations can be obtained in samples as small as 15 cases (see Lehmann 1983). This makes similarity coefficients an attractive option for small-*n* pilot research.

More important, similarity coefficients allow researchers to *explore* the structure of scale items. It is surely possible to begin an analysis of similarity coefficients with well-conceived ideas about the structure linking the items. However, it is also possible to start the analysis agnostically and to let the similarity coefficients reveal how well items hang together and along what dimensions. This exploratory use of similarity coefficients can be very useful and makes these coefficients an attractive choice in scale analysis and a useful complement to covariance structure analysis.

5 Analyzing Similarity Coefficients

Similarity coefficients are useful because they provide summaries of the internal and external consistency properties of scale items. Unfortunately, these summaries are not particularly parsimonious: there are as many nonredundant φ's and ψ's as there are nonredundant correlations. Moreover, the summaries are not defined at the correct level: they are computed for pairs of items, whereas scale analysis typically focuses on all items in a scale. Thus, similarity coefficients by themselves provide both too much information (too many coefficients) and too little (there is no diagnostic for an entire scale). To resolve the first problem, it is useful to represent graphically similarity coefficients through multidimensional scaling (MDS), since graphs often reveal patterns that would go unnoticed when inspecting a large matrix of similarity coefficients.[10] To solve the second problem (as well as the first), it is useful to condense similarity coefficients into several summary statistics.

5.1 *Graphical Methods*

MDS of similarity coefficients yields a graphical image from which the consistency properties of items can be readily inferred. Because similarity coefficients contain metric

[9] An exception to this rule is the estimation of polychoric and polyserial correlations, which require maximum-likelihood estimation. In general, larger samples are required to estimate these types of correlations than to estimate Pearson product–moment correlations (Jöreskog and Sörbom 1988). Even so, the computation of similarity coefficients on the basis of polychoric and polyserial correlations is less sample-intensive than covariance structure analysis, which, in the case of polychoric and polyserial correlations, requires the use of the very sample intensive weighted least-squares estimation procedure (Jöreskog 1990).

[10] It is also possible to obtain a graphic display by applying hierarchical cluster analysis to the similarity coefficients. The display (the dendrogram) that is obtained in this way usually does not lead to dramatically different interpretations of the behavior of scale items. For this reason, I do not discuss cluster analysis further in this paper.

information, Torgerson's (1958) classical scaling model can be used to project them graphically. To use this model, the similarity coefficients have to be transformed into dissimilarities that are mapped into Euclidian distances, according to

$$\delta_{ij} = d_{ij} = \left[\sum_k (x_{ik} - x_{jk})^2\right]^{1/2}$$

Here $\delta_{ij} = 1 - \varphi_{ij}$ (or $\delta_{ij} = 1 - \psi_{ij}$) are dissimilarities, d_{ij} are distances, and there are k dimensions in Euclidian space. The objective is to minimize k while still maintaining an adequate fit between distances and dissimilarities (as measured, for example, by Shepard's stress coefficient).

If the dissimilarities between all item pairs were perfect, the Euclidian space would either be a point (if $\varphi_{ij} = \psi_{ij} = 1$ for all $j \leq i$) or, in the case of reversed items, a line. As item similarities become more heterogeneous, the dimensionality of the Euclidian space may become more complex but items should still be tightly clustered together in this space. To the extent that this is no longer the case, it is safe to conclude that the complete set of items does not form a unidimensional scale. At this point it becomes important to identify anomalous items and to determine whether results improve after eliminating those items.

It is important to point out that the main objective of MDS is to aid in the identification of clusters of items. In principle it is possible to interpret the dimensions of the space in which the items are projected, but this is not of primary interest, nor is it required in the analysis of scale items. I should also note that it is not always necessary to perform MDS. This method is helpful only in those instances in which a clear pattern does not emerge from the matrix of similarity coefficients.

5.2 *Descriptive Statistics*

If one's objective is to relate similarity coefficients to entire scales, they must be summarized. I propose two summary statistics for this purpose. The first statistic is the item-wise similarity, which gives the average similarity between an item and a scale. The second statistic is the scale-wise similarity, which captures the similarity between two scales. Both statistics are available through my computer program UniSim.

To derive the item-wise similarity, let i denote a particular item and B denote a scale, where i may or may not be a part of B. Without loss of generality, assume further that a φ_{ij} -matrix has been computed. Then the item-wise similarity between i and B can be computed as the average

$$\varphi_{iB} = \frac{1}{P_B}\sum_{j\in B}\varphi_{ij} + D_i\left\{\frac{1}{P_B(P_B-1)}\sum_{j\in B}\varphi_{ij} - \frac{1}{P_B - 1}\right\}$$

where P_B is the number of items in B, and $D_i = \begin{cases} 1 & \text{if} \quad i \in B \\ 0 & \text{if} \quad i \notin B \end{cases}$.[11]

[11]The last term in this expression is a correction factor that is applicable in situations in which i is an element of scale B. In this case, one of the φ_{ij}-terms is the similarity of i to itself. This similarity is uninteresting because it is, by definition, 1. The correction factor removes the similarities of items to themselves and also corrects the number of items by which the sum of similarity coefficients is divided, reducing that number by 1.

To obtain scale-wise similarities between two scales, A and B, it is sufficient to average the item-wise similarities with B of all of the items that make up A:

$$\varphi_{AB} = \frac{1}{P_A} \sum_{i \in A} \varphi_{iB}$$

where P_A is the number of items in A. If $A = B$, the scale-wise similarity measures the average similarity of all of the items that make up a scale. I call this the *internal* scale-wise similarity. If $A \neq B$, then the scale-wise similarity measures the average similarity between all of the items that make up one scale and all of the items that make up another scale. I call this the *crossover* scale-wise similarity.

In evaluating scale-wise similarities the following rules of thumb can be applied:

1. In general, an internal scale-wise similarity of at least .80 suggests good consistency properties of the scale items (see Anderson and Gerbing 1982).
2. To ensure discriminant validity, internal scale-wise similarities should be higher than crossover scale-wise similarities. (If the crossover similarities are higher, this suggests that scale items have more in common with the items from another scale than among themselves.)
3. Crossover scale-wise similarities should be higher for related scales than for unrelated scales, to ensure convergent validity.

6 Illustrations

To illustrate the use of similarity coefficients I present two examples. The first example pertains to the policy mood concept developed by Stimson (1991). The methodological significance of this example is that the number of observations entering the analysis is very small ($n = 19$), making scale analysis challenging. The second example pertains to the items that Markus (1990) proposed for measuring popular individualism ($n = 383$). This example is methodologically interesting because it demonstrates how similarity coefficients can be used in detecting alternative specifications of item structures in the covariance structure analysis of scale items. The main focus in these illustrations is on summarizing the data patterns that are revealed by similarity coefficients and showing the advantages that these coefficients offer. A more detailed analysis of the two examples can be found at this journal's web site.

6.1 *Policy Mood*

Stimson's (1991) work on policy mood, defined as the mass public's aggregate ideological disposition, has attracted a great deal of scholarly attention in the past decade. Recently, this work has also begun to stir controversy. One of the more important issues of contestation is whether policy mood, along with the measures of policy mood that Stimson (1991) proposes, is unidimensional or multidimensional, as Best (1999) has argued.

The determination of the dimensionality of Stimson's (1991) policy mood items is not trivial. Stimson (1991) considers nine policy domains—education, health, urban affairs, race, welfare, crime, military spending, the size of the government, and the environment—with a sample size of only 19 years.[12] Confirmatory factor analysis with such a small sample

[12]This sample size is based on the policy mood time series reported by Stimson (1991), which runs through 1989. The time series has been updated, but for my purposes the original data are a better choice because the smaller

is out of the question. Even an exploratory factor analysis can be problematic, however, because this method is known to overfit in small samples by extracting too many factors (see Aleamoni 1973). A factor analysis of the policy mood items indeed suggests the existence of multiple factors, presenting a weak case for the claim that policy mood is unidimensional. A much stronger case for unidimensionality can be made on the basis of similarity coefficients.

6.1.1 Principal-Component Analysis

Principal-component analysis has been the method of choice in the analysis of policy mood to date because of the small number of time points for which survey data are available (Best 1999; Stimson 1991). Unfortunately, this type of analysis has produced ambiguous evidence concerning the dimensionality of policy mood. For instance, focusing on Stimson's (1991, p. 79) results, he extracts two principal components, each of which accounts for a sizable portion of the variance. Stimson's (1991) evaluation of these results is that the second component is noisy and that only the first component is meaningful. Thus, he concludes that the policy mood items form a unidimensional measure. However, a different conclusion is also possible. Since different items load on different components, one could develop substantively different interpretations for the two components. In this case, the second component is not discarded and one would conclude that policy mood has two dimensions. Indeed, using a different set of policy mood indicators, Best (1999) has argued in favor of this two-dimensional conceptualization.

6.1.2 Similarity Coefficients

Is the unidimensional or two-dimensional interpretation more appropriate? The principal-component analysis reported by Stimson (1991) does not provide a clear answer. However, in my judgment, the problem is not so much that the second principal component is not easily interpreted, as Stimson (1991) claims, but rather that it may be an artifact. At least, this is the conclusion that emerges from an analysis of the similarity coefficients of the policy mood items.

Table 1 shows the item-wise and scale-wise semiparallel similarity coefficients of the policy mood items based on the Pearson product–moment correlations between those items. The pattern that emerges from this table is very clear (in fact, the clarity is such that MDS of the similarity coefficients is unnecessary). First, consider the second column in the table, which reports similarity coefficients for all of the items. There is very little evidence in this column that the policy mood items span more than one dimension. With the exception of one item, the item-wise similarity coefficients are about .80 or higher, the conventional cutoff for acceptable item similarities (Anderson and Gerbing 1982). Moreover, the scale-wise similarity is a respectable .819.

The exceptional item is crime, which has an item-wise similarity coefficient of only .569, more than five standard deviations below the scale-wise similarity. This suggests that crime may not be measuring the same construct as the remaining policy mood items. However, it would be premature to infer a separate dimension from just this one anomalous item.

number of observations in these data more clearly reveals the advantages of similarity coefficients. Each item is based on a comparison of the percentage of liberal and conservative responses to different survey items (for details see Stimson 1991).

Table 1 Semiparallel similarity coefficients for policy mood items[a,b]

Item	*Including crime*	*Excluding crime*
Crime	.569	
Education	.742	.890
Environment	.867	.951
Health	.870	.940
Military	.880	.960
Race	.882	.957
Size of government	.810	.882
Urban affairs	.882	.952
Welfare	.869	.953
Scale-wise	.819	.936

[a]Similarity coefficients are based on Pearson product–moment correlations.
[b]Table entries are item-wise similarity coefficients, unless indicated otherwise.

If crime is an anomalous item, one would expect the similarity coefficients for the remaining policy mood items to improve dramatically after the removal of crime, at least to the extent that these items span a single dimension. Considering the last column in Table 1, this expectation is supported handsomely by the data. After removing the crime issue, the item-wise similarity coefficients for the remaining policy mood items lie well above .80 and, in most cases, exceed .90. As a result, the scale-wise similarity coefficient is now a very impressive .936.

These results leave little doubt about the dimensionality of the policy mood items. There is no evidence of a two-dimensional structure. Rather, it appears that policy mood is unidimensional and that this unidimensionality is distorted only because of the crime item.

6.1.3 Conclusion

The methodological implication of this example is that the principal-component analysis may exaggerate the number of dimensions underlying the policy mood items. This is a well-known problem with principal-component analysis in small samples (Aleamoni 1973). Moreover, it is a problem that does not disappear with the removal of the crime item. Even in this case, a principal-component analysis continues to suggest two distinct components (details are available at this journal's web site). The similarity coefficients suggest that this two-component solution is unnecessary and overfits the data.[13] Thus, the present evidence supports Stimson's (1991) conclusions, although, contrary to his conclusions, it also points out that it may be best to remove crime from the policy mood scale.

[13]It is useful to comment also on the shortcomings of reliability analysis. The problem here is that the problematic nature of the crime item remains hidden. For example, Cronbach's α for a policy mood scale including crime is .91, while it is .92 for a scale that excludes crime. Based on this marginal change in reliability, one would never suspect that the crime item behaves anomalously. Part of the reason why Cronbach's α changes only slightly is that the improved interitem correlation that is brought about by the removal of crime is offset by the smaller number of items in the scale.

6.2 *Popular Individualism*

In the 1989 American National Election Studies pilot survey, Markus (1990; see also Feldman 1999) made an important attempt at measuring popular individualism. He identified four dimensions of individualism—personal autonomy, self-reliance, limited government, and laissez-faire capitalism—and developed forced-choice items to measure each. In total, 12 items were proposed for measuring support for abstract individualistic principles; the question wording for these items appears in Table 2.

The obvious question about the Markus (1990) individualism items is whether they measure what they are supposed to measure. While Markus (1990) concludes that his items are unidimensional measures for the various aspects of individualism, a careful consideration of similarity coefficients lends little support to this conclusion. Instead these coefficients suggest a very different interpretation of the individualism items.

6.2.1 Factor Analysis

Markus' (1990, Table 4) evidence for unidimensionality comes from a principal-axis factor analysis of the 12 items. This analysis suggests that there are four principal factors, which are only moderately correlated. Moreover, with only three exceptions, the items load on the factors they are supposed to measure. The exceptional items are V7503, which did not correlate with any of the other items, as well as V7501 and V7504.

Impressive as these results seem, Markus' (1990) conclusions cannot be accepted at face value. First, his analysis suffers from a statistical flaw, in that he relies on Pearson product–moment correlations to analyze a set of dichotomous items (see Feldman 1999). This can lead to biased estimates of the relationships between the items (Hollis and Muthén 1987; Jöreskog 1990; Jöreskog and Sörbom 1988) and hence to bias in the factor analysis.

Second, there is evidence that some of the items have sizable loadings on multiple factors. While it is problematic to interpret factors in an exploratory factor analysis in terms of constructs, the evidence may indicate that the items measure multiple aspects of individualism. This calls into question the unidimensionality of scales formed with those items.

In sum, there is good reason to take a closer look at Markus' (1990) individualism items. In the analysis below, I use similarity coefficients to do this. After showing that the items lack unidimensionality, I show what alternative item configuration is suggested by the similarity coefficients.

6.2.2 Similarity Coefficients

Table 3 displays the semiparallel internal and crossover scale-wise similarity coefficients when Markus' (1990) conceptualization of the individualism items is accepted. These similarity coefficients are based on tetrachoric correlations between the items, which are better estimators of the true correlations (Hollis and Muthén 1987; Jöreskog 1990; Jöreskog and Sörbom 1988). As can be seen from Table 3, the internal scale-wise similarity coefficients are for the most part respectable (i.e., above .80), with the exception of the autonomy items. On the basis of the diagonal elements in Table 3 one would be inclined to agree with Markus (1990) that the items for limited government, self-reliance, and capitalism form homogeneous subsets, suggesting that it is possible to form unidimensional scales from them. One would also be inclined to conclude that the autonomy items lack unidimensionality, which may be due to the inclusion of V7503.

Table 2 Question wording for the ANES popular individualism items

		Classification	
Label	*Item*[a]	*Markus*[b]	*Alternative*[c]
V7364	ONE, government regulation of big business and corporations is necessary to protect the public or, TWO, **government regulation does more harm than good?**	C	OSL
V7365	ONE, the government should try to ensure that all Americans have such things as jobs, health care, and housing or, TWO, **the government should not get involved in this?**	LG	OSL
V7366	ONE, is it better to fit in with the people around you or, TWO, **is it better to conduct yourself according to your own standards even if that makes you stand out?**	A	NSL
V7367	ONE, **people should take care of themselves and their families and let others do the same** or, TWO, people should care less about their own success and more about the needs of society?	SR	OSL
V7368	ONE, **when raising children it is more important to teach them to be independent-minded and think for themselves** or, TWO, it is more important to teach them obedience and respect for authorities?	A	NSL
V7369	ONE, **most poor people are poor because they don't work hard enough** or, TWO, they are poor because of circumstances beyond their control?	SR	—
V7501	ONE, we need a strong government to handle today's complex economic problems or, TWO, **the free market can handle these problems without government being involved?**	C	OSL
V7502	ONE, **the less government the better** or, TWO, there are more things that government should be doing?	LG	OSL
V7503	ONE, there is too little respect for traditional authorities, such as religious leaders and government officials, or, TWO, **there is too much restriction and regulation of personal opinion and behavior?**	A	—
V7504	ONE, it is more important to be a cooperative person who works well with others or, TWO, **it is more important to be a self-reliant person able to take care of oneself?**	SR	NSL
V7505	ONE, **society is better off when businesses are free to make as much profit as they can** or, TWO, businesses should be prohibited from earning excessive profits?	C	OSL
V7506	ONE, **the main reason that government has gotten bigger is because it has gotten involved in things that people should do for themselves** or, TWO, government has gotten bigger because the problems we face have gotten bigger?	LG	OSL

[a]Response alternatives in boldface type indicate the individualistic response to the items.
[b]A, autonomy; SR, self-reliance; LG, limited government; C, laissez-faire capitalism.
[c]OSL, old-style (classical) liberalism; NSL, new-style liberalism.

One would not reach these conclusions so easily, however, when the internal scale-wise similarity coefficients are compared to the crossover scale-wise similarities. Of 12 possible comparisons, there are 5 instances in which a crossover scale-wise similarity exceeds an internal scale-wise similarity. In another four instances, an internal scale-wise

Table 3 Scale-wise similarity coefficients for popular individualism items[a–c]

	Item subset[d]			
Item subset	*A*[e]	*SR*	*LG*	*C*
A	.468	.616	.581	.551
SR	.616*	.805	.853	.784
LG	.581*	.853*	.962	.889*
C	.551*	.784	.889	.861

[a] $n = 383$.
[b] Similarity coefficients are based on tetrachoric correlations.
[c] Internal similarities appear on the diagonal in boxes.
[d] A, autonomy; SR, self-reliance; LG, limited government; C, capitalism.
[e] An asterisk indicates that the crossover scale-wise similarity coefficient exceeds the internal similarity coefficient in a column.

similarity exceeds a crossover similarity by a margin of less than .10. Even when the poorly behaved autonomy items are ignored, the internal and crossover scale-wise similarities for the remaining items are generally very close, indeed too close for comfort. This indicates that the different item subsets possess little discriminant power.

To obtain better insight into what is driving these results, it is useful to perform MDS on the similarity coefficients. The best-fitting configuration that emerges from this analysis is in three dimensions and is depicted in Fig. 1. Several patterns in this figure stand out. First, V7503 indeed emerges as an anomalous item, just like Markus (1990) concluded. The item-wise similarity of this item with the remaining autonomy items is extremely weak (.320), driving down the scale-wise similarity coefficient for the autonomy scale. Hence, it may be best to omit this item.

Contrary to Markus' (1990) analysis, however, one of the other items also stands out as problematic: V7369. While not as anomalous as V7503 (its item-wise similarity with the remaining self-reliance items is a respectable .829), this item does not clearly fall into one of the central clusters that are demarcated in Fig. 1. Instead, this item falls in between clusters, suggesting poor discriminant power. Therefore, I remove this item as well.

Second, the remaining items fall into two clusters, not four clusters as Markus (1990) suggested. This is an important difference with the principal-axis factor results discussed earlier. The emergence of two distinct clusters shows that there is poor discrimination among all four of the item subsets. More importantly, these results show that it is not necessary to use four dimensions to capture relationships between the individualism items. On the contrary, two dimensions probably suffice. Thus, the semiparallel similarity coefficients provide little support for Markus' (1990) assessment of the behavior of the individualism items.

Indeed, the conclusion that the items do not fit a four-dimensional structure is corroborated by a confirmatory factor analysis. A four-factor model, which excludes the two anomalous items (V7369 and V7503), fits the data only marginally: $\chi^2 = 42.42$, $df = 29$, $p = .052$.[14] Moreover, there are several large modification indices for the factor loadings, suggesting that several items load on multiple factors.

[14] The model was analyzed using LISREL 8.12, using tetrachoric correlations as inputs and a weighted least-squares (WLS) estimator.

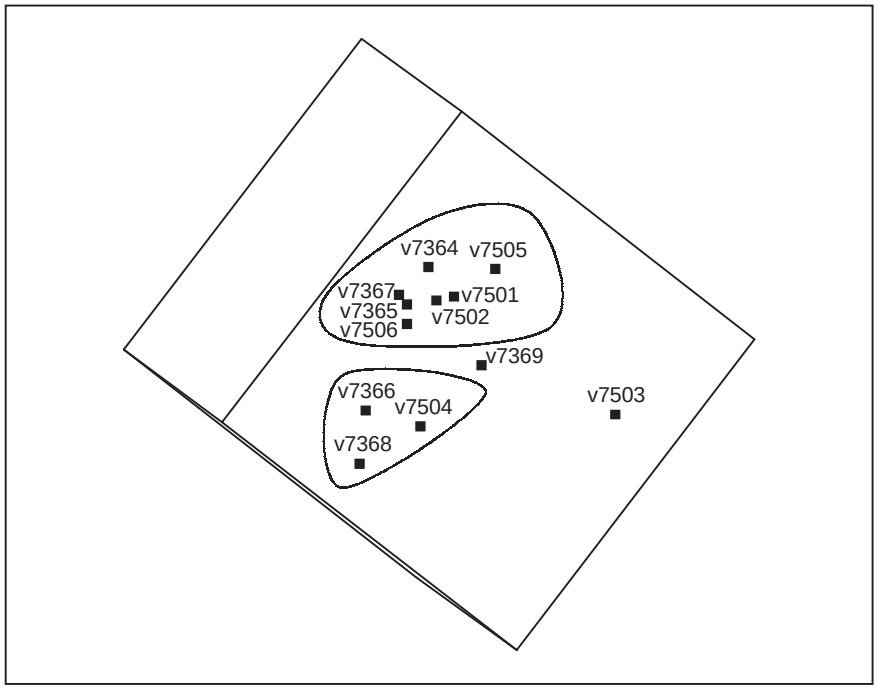

Fig 1 Three-dimensional similarity-based MDS stimulus configuration of individualism items. Euclidean distance model. Shepard's stress coefficient = .055. Item clusters are enclosed.

6.2.3 Alternative Item Structure

This poor fit of the four-factor model raises the question of how the items are truly structured. While the modification indices suggest model changes within the four-factor framework, the similarity coefficients suggest a much more dramatic change in understanding the dimensionality of the individualism items. Indeed, these coefficients suggest a more parsimonious two-dimensional structure, which is markedly different from the structure that Markus (1990) proposed and which can be readily interpreted in terms of varieties of the liberal creed.

To see this, consider Fig. 1 again. One of the two clusters that emerge in this figure consists of items V7364, V7365, V7367, V7501, V7502, V7505, and V7506. All of these items tap aspects of the classical liberal creed, or what one could call old-style liberalism or economic liberalism—limited government, the Protestant work ethic, entrepreneurship, and a belief in the free market. The second cluster in Fig. 1 contains V7366, V7368, and V7504. These items measure a newer liberal creed, which is libertarian in nature and emphasizes the importance of individuality and independence. The last column in Table 2 shows how the popular individualism items can be reclassified in terms of the constructs of old-style and new-style liberalism.

This two-dimensional structure appears to fit the data quite nicely. First, consider the scale-wise similarity coefficients in Table 4. This table shows handsome internal scale-wise similarities (especially for old-style liberalism). Moreover, the two item subsets discriminate well, with a crossover similarity coefficient that is considerably lower than the internal scale-wise similarities of each of the item subsets.

Table 4 Scale-wise similarity coefficients for individualism items conceived of in liberal ideological terms[a–c]

	Item subset[d]	
Item subset	*NSL*	*OSL*
NSL	.843	.720
OSL	.720	.909

[a] $n = 383$.
[b] Similarity coefficients are based on tetrachoric correlations.
[c] Internal similarities appear on the diagonal in boxes.
[d] NSL, new-style liberalism; OSL, old-style liberalism.

Next consider the results from a confirmatory factor analysis. A two-factor model with factors for old-style and new-style liberalism fits the data quite well: $\chi^2 = 39.91$, $df = 33$, $p = .190$; the comparative goodness of fit index is .98. Moreover, there are no significant modification indices on the factor loadings, suggesting that the items from each subset can be used to form unidimensional scales.

These results suggest that the individualism items can be thought of in terms of a two-dimensional structure, rather than a four-dimensional structure as Markus (1990) had suggested. This two-dimensional structure not only fits the data better, but is also more parsimonious. Most important, the structure is intuitively plausible because it nicely recovers the different types of liberal ideology that have been described in American politics (see, e.g., Asher 1992).

6.3 *Conclusions*

These two examples clearly demonstrate the utility of similarity coefficients in scale analysis. The first example shows the contribution that similarity coefficients make to scale analysis when covariance structure analysis is not feasible. The second example shows that even when covariance structure modeling is feasible, similarity coefficients can greatly aid in the specification of alternative measurement models by showing patterns in the relationships among items.

The reason scale analysis via similarity coefficients is so informative is that these coefficients combine exploratory power with the strengths of a measurement model. There are not many scale analysis tools that combine these strengths, and this makes it particularly worthwhile to add similarity coefficients to the repertoire of methods for scale analysis in political science. Similarity coefficients are of course no panacea, but as the examples show, they can surely make the lives of scale analysts a lot easier.

7 Conclusions—Practical Considerations

In this paper, I have sketched a new methodology for scale analysis that is particularly useful for assessing whether a set of items measures one and only one construct. This methodology is not meant to replace existing methodologies such as covariance structure analysis. On the contrary, when possible I recommend the use of covariance structure analysis over the use

of similarity coefficients because it provides a direct test of unidimensionality. Covariance structure analysis is not always feasible, however, nor does it provide the exploratory options that scale analysts sometimes need. In light of these limitations, similarity coefficients can provide a much-needed research tool.

To take full advantage of this tool one should take some simple principles into consideration. First, similarity coefficients can be only as good as the correlations on which they are based. It is important to use correlation coefficients that are appropriate for the level of measurement of a set of scale items, since a failure to use a correct correlation coefficient can cause biased estimates of the true relationship between two items (Hollis and Muthén 1987; Jöreskog 1990; Jöreskog and Sörbom 1988).

Second, similarity coefficients can reveal only as much about the behavior of scale items as the researcher allows them to reveal. The similarity between two items may look very good when these items are considered in conjunction only with other items from the same scale. It is quite possible, however, that the similarity looks considerably worse when items from other scales are introduced. The patterns that similarity coefficients reveal thus depend on the way the analysis is set up.

There are few hard and fast rules about the proper structure of a scale analysis that is based on similarity coefficients. In general, however, I recommend the inclusion of items from other scales in the computation of similarity coefficients when those other scales play an important role in subsequent analyses or when those scales pertain to constructs that are similar to the construct whose measures are being evaluated. In addition, it is often useful to perform a sensitivity analysis of the similarity coefficients by computing these coefficients with and without the inclusion of other scales.

If these simple principles are followed, meaningful similarity coefficients can be obtained. These coefficients can greatly aid the analysis and development of scales and can help ensure that political constructs are properly measured without confounding them. After all, proper measurement is the goal of any scale analysis.

Appendix

In computing the parallel similarity coefficient one has to assume that items are semiparallel. While this is a common assumption in scale analysis, it is rather stringent. This raises the question of how robust the semiparallel similarity coefficient is against violations of the parallel test assumption.

Consider two classical test items, i and j, with correlation vectors $\mathbf{q}_\mathbf{i}' = [\rho_{ij}\ \mathbf{r}_\mathbf{i}^{*\prime}]$ and $\mathbf{q}_\mathbf{j}' = [\rho_{ji}\ \mathbf{r}_\mathbf{j}^{*\prime}]$, as discussed in the context of Eq. (2). The classical test model implies that

$$\begin{aligned} \rho_{ij} &= \rho_{ji} = \lambda_i \lambda_j \\ \mathbf{r}_\mathbf{i}^* &= \lambda_i \boldsymbol{\Lambda} \boldsymbol{\tau} \\ \mathbf{r}_\mathbf{j}^* &= \lambda_j \boldsymbol{\Lambda} \boldsymbol{\tau} \end{aligned}$$

where $\boldsymbol{\Lambda}$ is a matrix of factor loadings of the items in $\mathbf{r}_\mathbf{i}^*$ and $\mathbf{r}_\mathbf{j}^*$ on their various constructs, and $\boldsymbol{\tau}$ is a vector of correlations of all constructs with the construct that i and j measure.

An assumption in the computation of ψ_{ij} is that $\lambda_i^2 = \lambda_j^2$. What happens if this assumption is violated? Assume, without loss of generality, that i is the more reliable item. It is possible to express the reliability of j as a fraction of the reliability of i:

$$\lambda_j^2 = k\lambda_i^2$$

where $0 < k \leq 1$. Evaluation of Eq. (2) then gives

$$\psi_{ij} = \left[k\lambda_i^4 + \sqrt{k}\lambda_i^2\boldsymbol{\tau}'\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\boldsymbol{\tau}\right] \times \left[\sqrt{k\lambda_i^4 + \lambda_i^2\boldsymbol{\tau}'\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\boldsymbol{\tau}}\sqrt{k\lambda_i^4 + k\lambda_i^2\boldsymbol{\tau}'\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\boldsymbol{\tau}}\right]^{-1}$$

This result can be simplified further by setting $\boldsymbol{\tau}'\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\boldsymbol{\tau} = m\lambda_i^2$, with $m \geq 0$. Now it can be shown that

$$\psi_{ij} = \frac{\sqrt{k} + m}{\sqrt{k + m}\sqrt{1 + m}}$$

The first implication that can be derived from this expression is that $\psi_{ij} = 1$ when $m = 0$. This implies that the semiparallel similarity coefficient between two items is, by definition, 1, if the only correlation on which the coefficient is based is the correlation between these items. Thus, semiparallel similarity coefficients are of interest only when at least three items are considered in the analysis.

The second implication that can be derived shows how ψ_{ij} behaves when two items are indeed equally reliable. In this case, $k = 1$ and it follows immediately that $\psi_{ij} = 1$, regardless of the value of m. This proves that ψ_{ij} reaches its extreme value for two semiparallel items.

Finally, we can consider the behavior of ψ_{ij} when $k < 1$ and $m > 0$. For this analysis it is useful to plot the value of ψ_{ij} for different combinations of k and m, as done in Fig. A1.

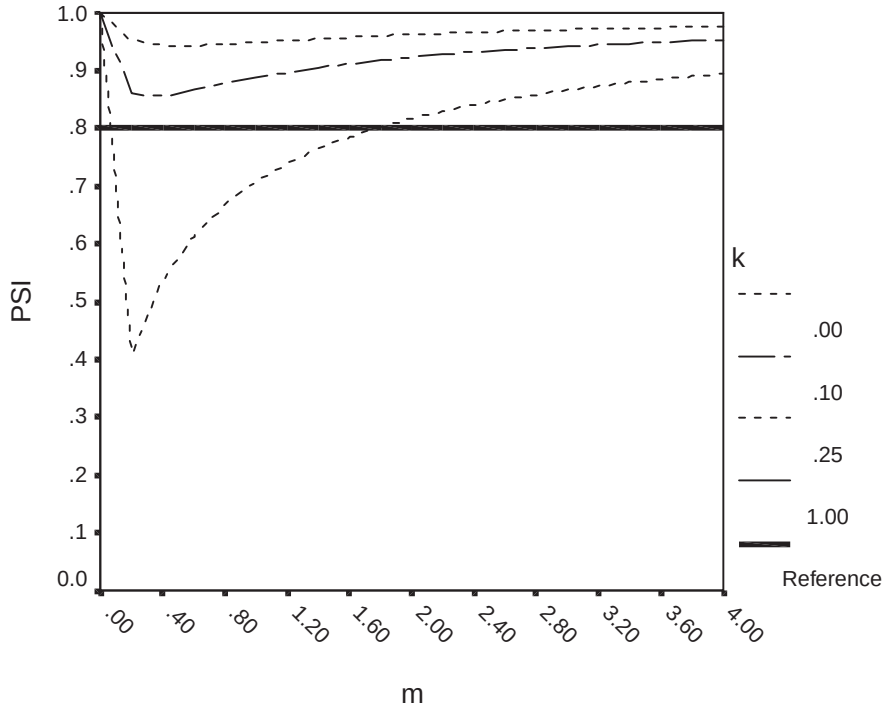

Fig A1 Sensitivity analysis of ψ. k is the ratio of reliabilities between two items. m is a multiple of the reliability of one of the items and is a function of the sum of the cross-products of the item with other items. The bold reference line indicates the conventional cutoff for an acceptable item reliability (Anderson and Gerbing 1982).

This figure shows the behavior of ψ_{ij} in the cases that j is only a quarter as reliable as i ($k = .25$), that j is only one-tenth as reliable as i ($k = .10$), and that j has a reliability that approaches 0 ($k \rightarrow 0$). Further, the size of $\boldsymbol{\tau}'\boldsymbol{\Lambda}'\boldsymbol{\Lambda}\boldsymbol{\tau}$ is set to vary from 0 ($m = 0$) to four times the reliability of i ($m = 4$). For purposes of benchmarking, Fig. A1 also depicts ψ_{ij} in the case that i and j are equally reliable ($k = 1$) and it depicts the minimally acceptable value of ψ_{ij} [i.e., .80 (see Anderson and Gerbing 1982)].

As the analysis in Fig. A1 shows, ψ_{ij} is very robust against violations of the parallelism assumption. Except in the extreme case that k approaches 0, ψ_{ij} tends to be high even when m is extremely small. For example, even in the case that j has only 10% of the reliability of i, none of the ψ_{ij} values ever fall below .80, and most lie above .90. When the reliability differential between i and j becomes even smaller, it is virtually impossible to find any values of ψ_{ij} that lie below .90.

These results are important. Typically, the unidimensionality of a pair of items is called into question when ψ_{ij} falls short of .80 (the bold reference line in Fig. A1). Figure A1 shows that it is unlikely to observe values for ψ_{ij} at this level merely because the parallelism assumption is violated. Thus, when ψ_{ij} is found to be very low this can be considered legitimate evidence that a problem with unidimensionality exists.

References

Aleamoni, Lawrence M. 1973. "Effects of Size of Sample on Eigenvalues, Observed Communalities, and Factor Loadings." *Journal of Applied Psychology* 58:266–269.

Anderson, James C., and David W. Gerbing. 1982. "Some Methods for Respecifying Measurement Models to Obtain Unidimensional Construct Measurement." *Journal of Marketing Research* 19:453–460.

Anderson, James C., and David W. Gerbing. 1984. "The Effect of Sampling Error on Convergence, Improper Solutions and Goodness-of-Fit Indices for Maximum Likelihood Confirmatory Factor Analysis." *Psychometrika* 49:155–173.

Anderson, James C., and David W. Gerbing. 1988. "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach." *Psychological Bulletin* 103:411–423.

Andrews, Frank M. 1984. "Construct Validity and Error Components of Survey Measures: A Structural Modeling Approach." *Public Opinion Quarterly* 48:409–442.

Asher, Herbert B. 1992. *Presidential Elections and American Politics: Voters, Candidates, and Campaigns Since 1952*, 5th ed. Pacific Grove, CA: Brooks Cole.

Benson, Jeri, and John A. Fleishman. 1994. "The Robustness of Maximum Likelihood and Distribution-Free Estimators to Non-Normality in Confirmatory Factor Analysis." *Quality & Quantity* 28:117–136.

Best, Samuel J. 1999. "The Sampling Problem in Measuring Policy Mood: An Alternative Solution." *Journal of Politics* 61:721–740.

Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.

Boomsma, Anne. 1982. "The Robustness of LISREL against Small Sample Sizes in Factor Analysis Models." In *Systems Under Indirect Observation: Causality, Structure, Prediction*, ed. Karl G. Jöreskog and Herman Wold. Amsterdam: North-Holland.

Boomsma, Anne. 1985. "Nonconvergence, Improper Solutions and Starting Values in LISREL Maximum Likelihood Estimation." *Psychometrika* 50:229–242.

Burt, Ronald S. 1976. "Interpretational Confounding of Unobserved Variables in Structural Equation Models." *Sociological Methods & Research* 5:3–49.

Cook, Thomas D., and Donald T. Campbell. 1979. *Quasi-Experimental Design & Analysis: Issues for Field Settings*. Boston: Houghton Mifflin.

Crocker, Linda, and James Algina. 1986. *Introduction to Classical and Modern Test Theory*. Fort Worth, TX: Harcourt Brace Jovanovich.

DeVellis, Robert F. 1991. *Scale Development: Theory and Applications*. Newbury Park, CA: Sage.

Dolan, Conor V. 1994. "Factor Analysis of Variables with 2, 3, 5 and 7 Response Categories: A Comparison of Categorical Variables Estimators Using Simulated Data." *British Journal of Mathematical and Statistical Psychology* 47:309–326.

Feldman, Stanley. 1999. "Economic Values and Inequality." In *Measurement of Political Attitudes*, eds. John P. Robinson, Phillip R. Shaver, and Lawrence S. Wrightsman. San Diego, CA: Academic Press.

Fornell, Claes, and Roland T. Rust. 1989. "Incorporating Prior Theory in Covariance Structure Analysis: A Bayesian Approach." *Psychometrika* 54:249–259.

Gerbing, David W., and James C. Anderson. 1988. "An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment." *Journal of Marketing Research* 25:186–192.

Gorsuch, Richard L. 1983. *Factor Analysis*. Hillsdale, NJ: Erlbaum.

Guilford, J. Paul. 1954. *Psychometric Methods*, 2nd ed. New York: McGraw–Hill.

Gulliksen, Harold. 1987. *Theory of Mental Tests*. Hillsdale, NJ: Erlbaum.

Hollis, Michael, and Bengt O. Muthén. 1987. "Structural Covariance Models with Categorical Data: An Illustration Involving the Measurement of Political Attitudes and Belief Systems." Paper presented at the annual meeting of the American Political Science Association, Chicago.

Hunter, John E. 1973. "Methods for Reordering the Correlation Matrix to Facilitate Visual Inspection and Preliminary Cluster Analysis." *Journal of Educational Measurement* 10:51–61.

Hunter, John E., and David W. Gerbing. 1982. "Unidimensional Measurement, Second Order Factor Analysis, and Causal Models." *Research in Organizational Behavior* 4:267-320.

Hunter, John E., David W. Gerbing, and Franklin J. Boster. 1982. "Machiavellian Beliefs and Personality: Construct Invalidity of the Machiavellianism Dimension." *Journal of Personality and Social Psychology* 43:1293–1305.

Jöreskog, Karl G. 1974. "Analyzing Psychological Data by Structural Analysis of Covariance Matrices." In *Contemporary Developments in Mathematical Psychology, Volume II: Measurement, Psychophysics, and Neural Information Processing*, eds. David H. Krantz, Richard C. Atkinson, R. Duncan Luce, and Patrick Suppes. San Francisco: W. H. Freeman.

Jöreskog, Karl G. 1990. "New Developments in LISREL: Analysis of Ordinal Variables Using Polychoric Correlations and Weighted Least Squares." *Quality & Quantity* 24:387–404.

Jöreskog, Karl G., and Dag Sörbom. 1988. *PRELIS: A Program for Multivariate Data Screening and Data Summarization*. Scientific Software.

Lehmann, Erich L. 1983. *Theory of Point Estimation*. New York: Wiley.

Lord, Frederic M., and Melvin R. Novick. 1968. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison–Wesley.

Markus, Gregory. 1990. "*Measuring Popular Individualism*," Mimeo.

McDonald, Roderick P. 1999. *Test Theory: A Unified Treatment*. Mahwah, NJ: Erlbaum.

Saris, Willem E., and Harm Hartman. 1990. "Common Factors Can Always Be Found but Can They Also Be Rejected?" *Quality & Quantity* 24:471–490.

Saris, Willem E., and Albert Satorra. 1993. "Power Evaluations in Structural Equation Models." In *Testing Structural Equation Models*, eds. Kenneth A. Bollen, and J. Scott Long. Newbury Park, CA: Sage.

Saris, Willem E., and Bas Van Den Putte. 1989. "True Score or Factor Models: A Secondary Analysis of the ALBUS-Test-Retest Data." *Sociological Methods and Research* 17:123–157.

Scherpenzeel, Annette C., and Willem E. Saris. 1997. "The Validity and Reliability of Survey Questions: A Meta-Analysis of MTMM Studies." *Sociological Methods & Research* 25:341–383.

Spector, Paul E. 1992. *Summated Rating Scale Construction: An Introduction*, Sage University Paper Series on Quantitative Applications in the Social Sciences 07-082. Newbury Park, CA: Sage.

Steenbergen, Marco R. 1999. *UniSim: A Program for Assessing Scale Unidimensionality*. Chapel Hill: University of North Carolina.

Stimson, James A. 1991. *Public Opinion in America: Moods, Cycles and Swings*. Boulder, CO: Westview.

Suen, Hoi K. 1990. *Principles of Test Theories*. Hillsdale, NJ: Erlbaum.

Torgerson, Warren S. 1958. *Theory and Methods of Scaling*. Malabar, FL: Krieger.

Tryon, Robert C. 1939. *Cluster Analysis: Correlation Profile and Orthometric Analysis for the Isolation of Unities in Mind and Personality*. Ann Arbor, MI: Edwards Brother.

Tucker, Ledyard R. 1940. "The Role of Correlated Factors in Factor Analysis." *Psychometrika* 5:141–152.

Wolins, Leroy. 1995. "A Monte Carlo Study of Constrained Factor Analysis Using Maximum Likelihood and Unweighted Least Squares." *Educational and Psychological Measurement* 55:545–557.