MONTE CARLO SIMULATIONS OF SIMILARITY COEFFICIENTS

Marco R. Steenbergen
Department of Political Science
University of North Carolina, Chapel Hill
Chapel Hill, NC 27599-3265

Steenbergen (2000) discusses the use of similarity coefficients for assessing whether a set of scale items measure one and only one construct (i.e., unidimensionality). He proposes two types of similarity coefficient. The first similarity coefficient was developed by Hunter (1973) and assumes congeneric test items. To compute this similarity coefficient one needs to possess information about or estimates of the reliabilities of the items that are analyzed. Such information is not required for the second type of similarity coefficient, the semi-parallel similarity coefficient. This coefficient, which was developed by Steenbergen (2000), assumes that all scale items are equally reliable.

Which similarity coefficient performs better in detecting unidimensionality? Hunter's similarity coefficient may require more or less arbitrary guesses of the item reliabilities. When these guesses are incorrect, this coefficient potentially will yield incorrect conclusions about unidimensionality. On the other hand, the semi-parallel similarity coefficient requires the assumption that items are equally reliable, which is not always plausible. When this assumption is wrong, this coefficient also may yield incorrect conclusions about unidimensionality. Which of these problems is worse? The following Monte Carlo simulation will shed light on this question.

DESIGN

The objective of the Monte Carlo simulations is to create measurement models that imply a particular correlational structure between items. I assume that researchers know this structure, except for the item reliabilities. That is, the true inter-item correlations are given in the correlation matrix, but the researcher does not know the diagonal elements of this matrix, which give the reliabilities of the items. These diagonal elements either can be ignored, as is done in the semi-parallel similarity coefficient, or they can be guessed, as is done in Hunter's similarity coefficient. Since the true correlational structure is known, the true similarity coefficients are also known. These can be compared with the semi-parallel similarity coefficients and with Hunter's similarity coefficients as estimated under different guesses of item reliabilities.

The simulation consists of computing Hunter's similarity coefficient for a large number of different guesses of the item reliabilities. For each item I conceptualize these guesses as random draws from a normal distribution with a given variance and a mean that is equal to the true reliability of the item plus bias. The bias indicates a systematic error that is repeated in all iterations of the simulation. The variance is an indication of uncertainty in the guesses of reliabilities: the greater the uncertainty is about an item's reliability, the larger the variance will be in the reliability guesses. The simulations consist of 10,000 draws from the normal distribution (draws are independent across items). Bias is systematically varied at levels of -.2, -.1, -.05, 0, .05, .1, and .2, while the variance takes on values of 0 (certainty), .00025, .000625, .0025, and .01.

The performance of Hunter's similarity coefficient and the semi-parallel similarity coefficient is likely to be affected by several attributes of the measurement models. In the simulations, I manipulate five different attributes of the measurement model. First, this model can contain items of a single construct or items of two different constructs. Second, when there are two constructs, these constructs can be correlated at three different levels—0, .25, and .75. Third, the number of items in the analysis varies; this number is fixed at 6 for models involving two constructs, but can take on values of 3 or 6 for models with a single construct. Fourth, the average reliability across the items may be low (.5) or high (.8). Finally, the dispersion of the reliabilities across items may be low or high.

The latter manipulation captures violations of the semi-parallelism assumption and is hence particularly relevant. Under semi-parallelism we should observe no dispersion in item reliabilities. In the simulation, this dispersion is set to a positive value. Specifically, item reliabilities are chosen either in a range between ±.05 from the mean reliability, representing a mild violation of semi-parallelism, or in a range between ±.20 from the mean reliability, representing a much more severe violation of semi-parallelism. One important question is how well the semi-parallel similarity coefficient holds up (relative to Hunter's similarity coefficient) when violations of the assumption of equal item reliabilities become more severe.

Combined with the selection of different values for the bias and variance of the distribution from which guesses about item reliabilities are generated, the various specifications of the measurement level produce 700 different simulations (as stated earlier, each simulation consists of 10,000 iterations).[1] These simulations provide the data on which I shall base comparisons of the performance of the semi-parallel similarity coefficient versus Hunter's similarity coefficient.

METHODS

To compare the performance of the two types of similarity coefficient I rely on estimates of the mean square error (MSE) for each item pair i, j, which are given by

$$MSE_{ij} = \frac{1}{K}\sum_{k=1}^{K}\left(\hat{S}_{ijk} - S_{ij}\right)^2 .$$

Here K denotes the number of iterations in the simulation (i.e., 10,000), S_{ij} denotes the true similarity between items i and j, and $\hat{S}_{ijk}$ denotes the estimated similarity between these items in iteration k of the simulation. The estimated similarity is either Hunter's similarity based on various guesses of the item reliabilities, or the semi-parallel similarity. In the case of semi-parallel similarity coefficients, the formula for the MSE may also be written as $\left(\hat{S}_{ij} - S_{ij}\right)^2$, since $\hat{S}_{ijk}$ is invariant across iterations (the same elements of the correlation matrix are ignored in each iteration).

The estimated MSEs thus defined pertain to the similarity of a pair of items. To obtain the estimated MSE across all items in the analysis, the estimated MSEs of non-redundant item pairs can be averaged (ignoring the diagonal elements of the similarity matrix). The resulting statistic is the estimated average mean square error or AMSE:

$$AMSE = .5P\left(P-1\right)\sum_{i}\sum_{j<i} MSE_{ij} ,$$

where P is the total number of items in the analysis.

[1] Given the design parameters, there are 280 simulations/cases for the single-factor measurement model (items × mean reliability × reliability dispersion × bias in prior reliability estimates × variance in prior reliability estimates = 2 × 2 × 2 × 7 × 5 = 280). There are 420 simulations/cases for the two-factor measurement model (factor correlations × mean reliability × reliability dispersion × bias in prior reliability estimates × variance in reliability estimates = 3 × 2 × 2 × 7 × 5 = 420).

The performance of the semi-parallel similarity coefficient relative to Hunter's similarity coefficient can be assessed by considering the log-ratio of the AMSEs of these coefficients, which I shall refer to as RAMSE:

$$RAMSE = \log\left(\frac{AMSE_{Semi-Parallel}}{AMSE_{Hunter}}\right).$$[2]

When the semi-parallel similarity coefficient outperforms Hunter's coefficient, this means that the $AMSE$ of the parallel coefficient is smaller than that of Hunter's coefficient, so that $RAMSE < 0$. On the other hand, if Hunter's similarity coefficient performs better than the semi-parallel coefficient then $RAMSE > 0$. Finally, when both coefficients perform equally well, then $RAMSE = 0$.

RESULTS

In order to claim that semi-parallel similarity coefficients are an attractive alternative to Hunter's similarity coefficient we should find that $RAMSE < 0$ most of the time. The bottom row of Table A.1 suggests that this is indeed the case. For the 700 simulations the semi-parallel similarity coefficient outperformed Hunter's similarity coefficient almost 75% of the time. Moreover, the mean of the AMSE for semi-parallel similarities was on the whole only .000059, which compares favorably to the overall mean AMSE for Hunter's similarity coefficient, which is .000479. This is also evident from the mean value of the RAMSE across different simulations, which was -1.1424.

Hunter's similarity coefficient outperformed the semi-parallel similarity coefficient only under very limited circumstances. Specifically, as Figure A.2 shows, Hunter's similarity coefficient had superior performance when there was no bias in the specification of item reliabilities, and then only when there was little or no uncertainty (variance less than .0025).[3] It

[2] This statistic can only be calculated when the AMSE for Hunter's similarity coefficient is not 0. In some instances, this condition was not met because Hunter's similarity coefficient perfectly recovered the true similarities. In these instances, the AMSE for Hunter's similarity coefficient is set to an arbitrary low value in order to compute RAMSE.

[3] An ANOVA shows that bias has a significant main effect on the AMSE (F[6, 665] = 26.07, $p<.01$) and that variance has a significant main effect on the AMSE (F[4, 665] = 5.44, $p<.01$). There is no evidence

is unlikely that these conditions are common; after all, researchers generally approach the specification of the item reliabilities with a great deal of uncertainty, and may well be too optimistic (or pessimistic) about the reliability of their items.

Table A.1 Performance of Similarity Coefficients in Monte Carlo Simulations[a, b, c]

Reliabilities	Mean of AMSE Semi-Parallel	Mean of AMSE Hunter	Mean of RAMSE	Range of RAMSE	% Simulations RAMSE<0
Low Dispersion	.000033	.000450	-2.1591	-5.00 – 3.15	87.4 (N = 350)
High Dispersion	.000086	.000509	-0.1256	-1.87 – 3.44	62.3 (N = 350)
Combined	.000059	.000479	-1.1424	-5.00 – 3.44	74.9 (N = 700)

Notes: [a] Based on 700 simulations of 10,000 iterations each; [b] AMSE is the estimated mean square error across the iterations averaged over non-redundant item pairs (in this table the mean AMSE across 700 simulations is given); [c] RAMSE is the log of the ratio of the AMSE for the semi-parallel similarity and Hunter's similarity.

Performance of the semi-parallel similarity coefficient is affected by the severity of the violation of the assumption that items are equally reliable. As Table A.1 shows, when the reliabilities have limited dispersion, the relative performance of the semi-parallel similarity coefficient, as compared to Hunter's similarity coefficient, is the best. When greater discrepancies in item reliabilities exist, Hunter's similarity coefficient outperforms the semi-parallel similarity coefficient more frequently, as one should expect. These differences are statistically significant in an ANOVA ($F[1, 699] = 296.19$, $p<.01$). It is important to note, however, that even in this unfavorable case the semi-parallel similarity coefficient outperforms Hunter's similarity coefficient in a majority of the simulations.[4]

DISCUSSION

The point of these analyses is not to show that Hunter's similarity coefficient is bad. One should keep in mind that the values for the AMSEs reported in Table A.1 are extremely small,

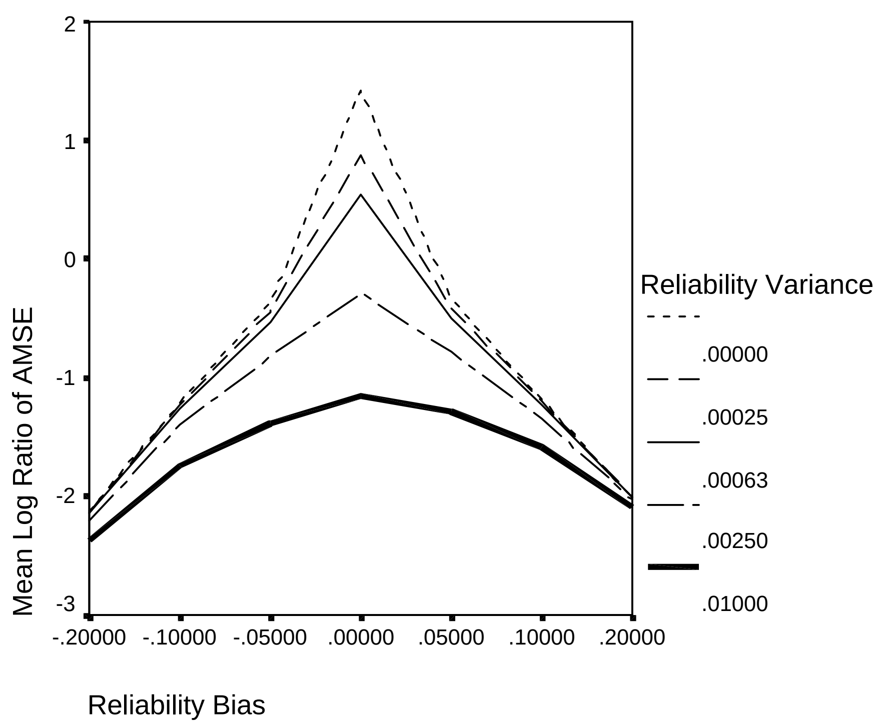
for an interaction effect between bias and variance in the guesses of item reliabilities ($F[24,665] = .80$, *ns*).

[4] Other aspects of the design also influence the relative performance of the parallel similarity coefficient. Details about these results are available upon request.

suggesting that poorly specified item reliabilities have little impact on Hunter's similarity coefficient—a result that Hunter (1973) also demonstrated. Rather, the point of these analyses is that one can get away not specifying item reliabilities at all. Under most circumstances, researchers are better off ignoring the item reliabilities by making a parallel test assumption then by trying to guess the item reliabilities. Even if the assumption that items are equally reliable is flawed, as it often will be, the statistical properties of the semi-parallel similarity coefficients are typically better then the properties of Hunter's similarity coefficients. If there is little payoff for specifying item reliabilities, except when one happens to guess them correctly, it may be good advice to ignore reliability information altogether.

Figure A.2:

The Impact of Bias and Variance in Reliability Estimates on the Performance of Similarity Coefficients



REFERENCES

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.

Steenbergen, Marco R. 2000. “Item Similarity in Scale Analysis.” *Political Analysis* 8: 261-283.