The Measurement of Values in Surveys: A Comparison of Ratings and Rankings

Duane F. Alwin; Jon A. Krosnick

The Public Opinion Quarterly, Vol. 49, No. 4 (Winter, 1985), 535-552.

Stable URL:
http://links.jstor.org/sici?sici=0033-362X%28198524%2949%3A4%3C535%3ATMOVIS%3E2.0.CO%3B2-L

The Public Opinion Quarterly is currently published by Oxford University Press.

The Measurement of Values in Surveys: A Comparison of Ratings and Rankings

DUANE F. ALWIN AND JON A. KROSNICK

TO MANY social psychologists the concept of values is crucial to the understanding of human behavior. Values are generally defined as standards of desirability invoked in social interaction to evaluate the preferability of behavioral goals or modes of action (Williams, 1968). According to this perspective, values are assumed to be central to the cognitive organization of the individual and to serve as a basis for the formation of attitudes, beliefs, and opinions (see Rokeach, 1970).

Survey researchers have generally measured values using standardized techniques of ordering (or ranking) a set of competing alternatives provided by the investigator. Other methods have been used, but rankings have been the method of choice most frequently. However, rankings

Abstract Social values are most commonly measured using ranking techniques, but there is a scarcity of systematic comparisons between rankings and other approaches to measuring values in survey research. On the basis of data from the 1980 General Social Survey, this article evaluates the comparability of results obtained using rankings and ratings of valued qualities. The comparison focuses on (1) the ordering of aggregate value preferences and (2) the measurement of individual differences in latent value preferences. The two methods are judged to be similar with respect to ordering the aggregate preferences of the sample, but dissimilar with regard to the latent variable structure underlying the measures.

Duane F. Alwin is Research Scientist and Professor of Sociology, Survey Research Center, Institute for Social Research, The University of Michigan. Jon A. Krosnick is Assistant Professor of Psychology, The Ohio State University. This article is a revised version of a paper presented at the annual meetings of the American Association for Public Opinion Research, May 17–20, 1984. An earlier version of the paper was reprinted as Technical Report No. 40, General Social Survey, NORC, Chicago. The research was supported by a grant from the National Institute of Mental Health (MH-37289). An important debt is owed NORC and the staff of the General Social Survey for cooperation in conducting this research. The authors wish to thank Frank Andrews, David Jackson, Melvin Kohn, and Tom Smith for helpful comments on earlier drafts and Lynn Dielman and Marion Wirick for help in manuscript preparation. Address all correspondence to Duane F. Alwin, Survey Research Center, Institute for Social Research, P.O. Box 1248, Ann Arbor, MI 48106.

Public Opinion Quarterly Vol. 49:535–552 © by the American Association for Public Opinion Research
Published by Elsevier Science Publishing Co , Inc. 0033-362X/85/0049-535/$2.50

have a number of problems tied to their degree of difficulty, and some have suggested that other techniques, such as ratings, could serve as effective alternatives. In this article, we compare the rating and ranking approaches to measuring values using data on parental orientations toward children from a randomized split-ballot experiment carried out in the 1980 General Social Survey. Before reporting our findings, we place the work in context by reviewing the literature that has discussed the relative advantages and disadvantages of rating and ranking techniques for measuring values.

The Measurement of Values: Ratings vs. Rankings

Value researchers have consistently argued on theoretical grounds that ranking techniques provide the most appropriate conceptual mapping to conceptions of values. For example, Rokeach (1973:6) points out that values are often thought to be inherently comparative and competitive, and thus the "choice" nature of the ranking task fits nicely with this conceptualization. Also, Kohn (1977:19) observes that the ranking approach to measuring values is demanded by their very nature, in that "a central manifestation of value is to be found in *choice*." This point of view is validated to some extent by the sheer prevalence of the use of rankings to measure values (e.g., Allport et al., 1960; Kluckholn and Strodtbeck, 1961; Lenski, 1961; Bengston, 1975).

Despite these arguments regarding the conceptual mapping of ranking techniques to the concept of values, this approach has a number of significant practical drawbacks when used in survey research. First, rankings are often difficult and taxing for respondents, demanding considerable cognitive sophistication and concentration. This is particularly problematic when the list of concepts to be ranked is lengthy (Rokeach, 1973:28; Feather, 1973:228). Second, the use of ranking techniques is time-consuming and may therefore be more expensive to administer (Munson and McIntyre, 1979:49). Third, since rankings often require the use of visual aids, or "show-cards," it is difficult to gather such information using telephone methods of data collection (Groves and Kahn, 1979:122-33). And finally, the sum of the ranks for any individual respondent equals a constant, so there is a linear dependency among the set of ranked items (Clemans, 1966; Jackson and Alwin, 1980). Consequently, it may not always be possible to employ conventional statistical techniques in the analysis of the latent content of ranked preference data.

In contrast to rankings, rating scales are easy to present to respondents. Munson and McIntyre (1979:49) estimate that ranking tasks take three times longer than similar rating tasks and involve a considerable decrease in respondent burden. Also, ratings can readily be administered over the

telephone and do not involve the linear dependency problem inherent in rankings. So ratings have none of the major disadvantages of rankings and might therefore be a good substitute for survey measurements of values.

However, ratings have two potential drawbacks which should be evaluated before they are employed. First, since ratings require less effort, the quality of data may be reduced relative to rankings. And as Feather (1973:229) points out, making the task easier may also reduce respondents' willingness to make more precise distinctions about the relative importance of valued qualities. Thus, although easier to administer and perform, ratings may compromise the level of precision of the data.

A second potential drawback of ratings is their susceptibility to problems of response style or response sets (Berg, 1966; Block, 1965; Phillips, 1973). When the rated qualities are all considered good or socially desirable, ratings tend to fall within a rather restricted range of the available scale points (Feather, 1973). The particular center or anchor-point of an individual's ratings may be due to extremity response style (Hamilton, 1968), individual interpretations of the meaning of judgment categories (Cronback, 1946, 1950; Messick, 1968), or group response sets (Cunningham et al., 1977). Variation across persons in such response tendencies may lead to correlated response patterns, or what Costner (1969) referred to as differential bias, producing spuriously positive correlations among ratings due to the common method of measurement (see also Alwin, 1974; Andrews, 1984). These positive correlations may complicate the analysis of the latent structure of value preferences.

EMPIRICAL COMPARISONS OF RATINGS AND RANKINGS

The major research comparing ratings and rankings as measures of values has focused on Rokeach's rankings of "instrumental" (modes of behavior) and "terminal" (end-states of existence) values using his *Value Survey* (Rokeach, 1967) and rating techniques adapted from it. This research has shown the following: (1) the aggregate or average preference orders measured by ratings and rankings have generally been found to be quite similar (Feather, 1973, 1975; Moore, 1975); (2) individual-level preference orders tend to be much less similar across ratings and rankings (Moore, 1975; Rankin and Grube, 1980), primarily because in using ratings respondents can score valued qualities equally; (3) over-time relationships among identical measures are slightly higher for rankings than for ratings, although the differences tend to be small (Munson and McIntyre, 1979; Rankin and Grube, 1980; Reynolds and Jolly, 1980); and (4) the predictive validity of ratings is somewhat higher than that of rankings (Rankin and Grube, 1980).

This research suggests that rating and ranking techniques may be inter-

changeable for the purpose of measuring aggregate preference orderings. However, there are other purposes for which researchers may wish to assess the value-preferences of respondents in surveys, and the validity of various measurement approaches depends on the extent to which these objectives are accomplished.

Another important purpose for measuring values is to study their latent content as reflected in their covariance or correlational structure (e.g. Kohn, 1969; Rokeach, 1973). Past research has failed to focus explicitly on differences in the covariance structures of measures produced by ratings and rankings. Although some of the research cited above deals superficially with this issue, there is a virtual absence of research that has compared the covariance properties of ratings and rankings and the consequences of these differences for analysis of the comparative measurement differences between the two techniques. This is especially remarkable since the major anlaytic technique used to decompose measures of value preferences and to ascertain their underlying latent dimensionality, namely factor analysis, depends intimately on the covariance structure of the measures of preferences.[1]

In the following analysis of value measures, we compare the results obtained with a rating technique to those obtained using a more conventional ranking procedure. We first examine measures of the relative importance of value-preferences in the aggregate in order to verify in our data the finding in the literature that ratings and rankings provide similar pictures of relative preference orders. Second, we compare the intercorrelations among rated items with those among rankings. These analyses illustrate some major differences in the covariance properties of measures obtained via the two procedures. Third, we examine results obtained from the factor analysis of the two types of measures, and we examine these differences in light of the differences in their covariance properites. And fourth, we examine the relationships of the latent factors identified through factor analyses with criterion variables thought to be related to these values.

Data

The 1980 General Social Survey (see NORC, 1982) conducted an experimental comparison of three forms of a measurement technique

[1] One study referred to by Munson and McIntyre (1979) factor-analyzed rating and ranking measures of Rokeach's terminal and instrumental values. Three times as many factors were apparently required to account for the correlations among ratings compared to the number required for correlations among rankings. This suggests that the latent structure of the two types of measures are quite different. One problem with generalizing much from

prominent in the literature on the measurement of parental valuation of child qualities (see Kohn, 1969). Two forms were quasireplications of the ranking method originally used by Kohn, and a third form used a rating scale format.

The question used in prior GSS surveys and for one ranking form (Form X) of the current experiment is as follows:

a. The qualities listed on this card may all be important, but which *three* would you say are the *most desirable* for a *child* to have?
b. Which *one* of *these three* is the *most* desirable of all?
c. All of the qualities listed on this card may be desirable, but could you tell me which *three* you consider *least important*?
d. And which *one* of these three is *least important* of all?
   (1) that he has good manners
   (2) that he tries hard to succeed
   (3) that he is honest
   (4) that he is neat and clean
   (5) that he has good sense and sound judgment
   (6) that he has self-control
   (7) that he acts like a boy (she acts like a girl)
   (8) that he gets along well with other children
   (9) that he obeys his parents well
   (10) that he is responsible
   (11) that he is considerate of others
   (12) that he is interested in how and why things happen
   (13) that he is a good student.

On the second ranking form (Form Y), "a child" was substituted for "he" in the usual NORC format in order to remove any gender connotation in the list of qualities. Schaeffer (1982) has recently demonstrated very little difference in the mean rankings using these two forms, and our results were the same for both, so we combine them for presentation below.[2]

The rating form (Form Z) of the values measure asked the following question:

Please look at the qualities listed on this card. All of the qualities may be desirable for a child to have, but could you tell me whether the quality is extremely important, very important, fairly important, not too important, or not important at all?

The list of qualities presented was the same as those presented with the Form X ranking version, that is, phrased with the pronoun "he."

The population sampled in the 1980 NORC survey was the total

---

this type of comparison is that the ordinary factor analysis of rankings is inappropriate without modifying the common factor model to accommodate the ipsative properties of the data (See Jackson and Alwin, 1980).

[2] We also confirmed this finding using more appropriate Hotelling T-tests (see Anderson, 1958) on the vectors of means between forms and found no significant difference.

noninstitutionalized English-speaking population of the continental United States, 18 years of age or older. The 1980 sample was produced by full-probability cluster sampling methods.[3] One-third of the sample each received (on a random basis) one of the following three forms:

1) Form X—The standard reduced-ranking form using the "he" pronoun.
2) Form Y—The standard reduced-ranking form using "a child."
3) Form Z—The use of five-point rating scales for each quality separately.

For the analyses reported below, the 13 valued qualities measured using Form X and Y were scored as follows:

5—The trait or quality most valued of all
4—One of the three most valued qualities, but not the most valued
3—Neither one of the three most nor one of the three least valued qualities
2—One of the three least valued, but not the least valued quality
1—The quality least valued of all

In Form Z the response categories are scored as follows:

5—Extremely important
4—Very important
3—Fairly important
2—Not too important
1—Not at all important

These coding schemes are relatively arbitrary and may not necessarily represent interval-level metrics. However, we are convinced by O'Brien's (1979) analysis that the assumption of interval-level scoring is robust with respect to monotonic transformations of the scale units which preserve the ordinal character of the data.

Results

THE IMPORTANCE OF CHILD QUALITIES

Consistent with earlier research (e.g., Feather, 1973), we found that ratings and rankings produce very similar results when considered in

[3] The sampling details are given in NORC (1982:207–212). Earlier analyses of these measures in the pre-1980 GSS data restricted the analysis to parents (Kohn, 1976; Alwin and Jackson, 1982b). Consistent with this previous research we present results for the subsample of parents only, defined as respondents who reported ever having had children, regardless of whether they were currently living with them. In addition, we exclude cases for which there is not complete data on the 13 parental value items. In the ranking forms this necessitated deleting 13 percent of the cases, and in the rating form 3 percent of the cases were deleted.

terms of their measurement of the relative importance of the child qualities. Table 1 presents the mean ranks and ratings for each quality studied and the percentage of respondents choosing a particular quality as "most important" (in Forms X and Y) and "extremely important" (in Form Z). Using either technique, the quality estimated to be the most valued in the population is "honesty" and the quality estimated to be the least valued is "acting like a child should." The rank-order of the remaining qualities in terms of their overall relative importance is very similar. The Spearman rank-order correlation between the mean rankings and mean ratings in Table 1 is .966.

Table 1. Descriptive Statistics for Ratings and Rankings of Child Qualities by Experimental Condition: NORC General Social Survey, 1980

	Form			
	Ranking (n=655)		*Rating (n=318)*	
Quality	*Mean Reduced-Rank Score*	*% Ranking Quality Among 3 Most Important*	*Mean Rating*	*% Rating Extremely Important*
Manners	2.05	24.5	3.13	29.6
Tries hard	1.97	17.2	3.05	29.2
Honest	3.12	66.1	3.64	66.0
Neat & clean	1.48	6.4	2.93	24.8
Good sense	2.46	41.1	3.31	40.3
Self-control	2.01	13.7	3.22	35.5
Role	.84	3.4	2.71	24.8
Gets along	1.96	13.6	2.96	24.2
Obeys	2.38	31.6	3.30	44.7
Responsible	2.35	34.2	3.29	38.1
Considerate	2.21	25.3	3.29	36.8
Interested	1.64	16.9	3.00	27.0
Studious	1.65	5.8	2.88	23.6

The ranking technique produced sharper distinctions among the various qualities. Table 1 shows that ratings tend to be somewhat skewed, with the bulk of respondents responding positively, as suggested in the existing literature. For most rated qualities, at least one-quarter of the respondents selected the "extremely important" category, and in several cases, the figure is upwards of 40 percent. These results stand in clear contrast to the ranking of child qualities, wherein the procedure requires that some qualities be ranked below others. Here the percentage of respondents choosing qualities among the "three most important" is quite variable over items, suggesting the greater differentiation of preferences by the ranking approach.

CORRELATIONS AMONG VALUE ITEMS

Rankings have built-in negative correlations among them since the rankings for any given respondent sum to a constant (see Cattell, 1944; Horst, 1965; Clemans, 1966; Jackson and Alwin, 1980; Alwin and Jackson, 1982a; 1982b). And as discussed above, ratings of value preferences may have built-in positive correlations among them due to response biases (e.g., Bentler, 1969). This difference between the covariance

Table 2. Bivariate Correlations Among Child Qualities for Ratings and Rankings: NORC General Social Survey, 1980

Quality	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
1 Manners	1.0	.385	.204	.517	.183	.320	.400	.444	.444	.304	.259	.249	.389
2 Tries hard	- 067	1.0	.352	.455	.336	.311	.427	.351	.336	.307	.176	.378	.492
3 Honest	-.030	-.166	1.0	.166	.388	321	.182	.146	279	.359	.288	.268	.261
4. Neat & Clean	-.092	- 090	.005	1.0	.306	.378	.539	.525	.412	.286	.320	.314	.535
5 Good sense	- 186	-.025	-.204	-.152	1.0	.484	276	.329	.210	446	.380	.442	.334
6. Self-control	-.106	-.141	003	-.145	-.043	1.0	.355	.326	.406	.418	331	.387	.446
7 Role	-.088	-.020	-.053	-.105	-.128	-.046	1.0	.457	.486	.198	.135	.339	.577
8 Gets along	- 020	-.084	-.093	-.139	- 101	-.044	- 100	1.0	.377	348	.372	.368	502
9 Obeys	043	-.195	-.109	.089	-.278	- 054	-.006	-.057	1.0	.310	290	.248	.485
10. Responsible	-.236	- 072	-.049	-.203	.065	-.014	-.087	-.066	-.259	1.0	514	.430	409
11. Considerate	-.162	-.131	-.063	-.137	-.010	-.061	-.151	.010	-.052	.002	1.0	.380	.359
12 Interested	-.185	002	-.193	-.190	.025	-.083	-.229	-.025	-.186	.032	-.005	1.0	.409
13. Studious	-.073	-.043	-.093	-.084	-.080	-.053	-.143	-.028	.051	-.060	- 073	-.147	1.0

NOTE: Form X and Y N=655 (below diagonal).
Form Z N=318 (above diagonal).

properties of ratings and rankings is illustrated in Table 2. The correlations among the (nonipsative) rating scales are all positive and show considerable intercorrelation in all cases, whereas the intercorrelations among ranked items tend toward zero, and in most instances are negative. As indicated in the foregoing, these covariance properties are directly linked to fundamental differences between the measuring procedures themselves. Consequently, it is difficult to compare the two techniques without taking into account these inherent sources of differences in their covariance structures. Therefore, the factor analysis of ratings must allow for positive correlations due to correlated response errors, and the factor analysis of rankings must take into account the negative correlations of ranked qualities.

ANALYSIS OF LATENT STRUCTURE

The Factor Analysis of Rankings. To analyze the factor structure of the ranking data, we applied the Jackson-Alwin ipsative common factor model (see Jackson and Alwin, 1980; Alwin and Jackson, 1982a; 1982b), which imposes a set of constraints to produce the negative correlations

among disturbances which are inherent in ipsative data.[4] For purposes of comparison, we also evaluated a "nonipsative" factor model, which assumes the disturbances in the factor model are uncorrelated. For the purpose of estimating these parameters, we have arbitrarily constrained the variance of the latent factor to unity. Parameter estimates were obtained using maximum-likelihood confirmatory factor analysis (Jöreskog and Sörbom, 1981), and the results are presented in Table 3.

Table 3. Parameter Estimates of a Single-Factor Model for Rankings of Child Qualities Using Ipsative and Nonipsative Factor Models: NORC General Social Survey, 1980

Parameter	*Nonipsative Model*	*Ipsative Model*
Factor loadings		
Manners	-.300*	-.281*
Tries hard	.091*	.086
Honest	-.110*	-.095
Neat & clean	-.304*	-.280*
Good sense	.360*	.339*
Self-control	.033	.033
Role	-.103*	-.092
Gets along	.014	.013
Obeys	-.391*	-.385*
Responsible	.312*	.307*
Considerate	.097*	.097*
Interested	.326*	.309*
Studious	-.025	-.051
Factor variance	1.000[a]	1.000[a]
Disturbance variances		
Manners	.595	.609
Tries hard	.636	.647
Honest	.758	.753
Neat & clean	.598	.620
Good sense	.712	.726
Self-control	.418	.420
Role	.861	.860
Gets along	.401	.400
Obeys	.581	.580
Responsible	.493	.493
Considerate	.467	.466
Interested	.855	.858
Studious	.491	.484
Goodness of Fit		
Ipsative model:	L^2 = 100.750, df=53, L^2/df= 1.90	
Nonipsative model:	L^2 = 1290.931, df=54, L^2/df=23.91	

[a] Fixed parameter.
* Factor loading is at least twice its standard error.

[4] The rationale for this model and the details involved in its estimation may be found in Alwin and Jackson (1982a).

The relative sizes of the factor loadings shown in Table 3 are consistent with the notion of a single underlying self-direction/conformity factor in these data, and there is strong similarity between the two factor solutions in this regard. These numbers resemble those obtained both by Kohn (1976) and by Alwin and Jackson (1982b). Negative loadings are associated with conformity items, such as obedience, manners, and cleanliness. Positive loadings are associated with self-direction items, such as good sense and sound judgment, responsibility, and curiosity.[5]

We also examined the residual covariances among the variables under the two models. In virtually every case, the residual covariances are negative in the nonipsative model, suggesting the need for negatively correlated disturbances. Because the ipsative model incorporates such correlations, no such pattern is obvious from the inspection of the residual covariances of that model. Also, because the ipsative common factor model does not ignore the "built-in" negative correlations among disturbances, the fit of this model to the data is literally more than 10 times better than the fit of the nonipsative model (see Table 3). The fit of the ipsative factor model is acceptable (L^2/df=1.90), whereas the fit of the nonipsative model is clearly unacceptable (L^2/df=23.91).[6]

The Factor Analysis of Ratings. We analyzed the factor structure of the rating data using two types of common factor models: (1) a model based upon preliminary exploratory factor analyses specifying two latent factors, and (2) a model that incorporates these two factors and a general method factor, uncorrelated with the other two, upon which all items had equal loadings.[7] This third factor was included to incorporate explicitly the spurious positive correlations which might result from differential response bias (Costner, 1969; Bentler, 1969; Alwin, 1974; Andrews, 1984).

Table 4 displays the factor pattern coefficients in standard form for the

[5] Because of the properties of the common factor model for rankings, the factor pattern coefficients must sum to zero. Therefore, the coefficients in Table 3 *cannot* be interpreted in absolute terms; they simply reflect the relative ordering of items in relation to one another (see Jackson and Alwin, (1980:222-23).

[6] We use the notation L^2 to refer to the sample estimate of the population likelihood-ratio χ^2 value for the model. We should point out that the use of L^2 as an estimate of χ^2 depends upon its maximum likelihood properties. In order for L^2 to estimate χ^2, the variables must have multivariate normal distributions. This assumption is likely to be violated in the case of rankings (but not necessarily for ratings) since rankings are often relatively skewed. We acknowledge this distributional requirement and therefore place less weight on the statistical tests in evaluating the fit of the models to the data. The Bentler-Bonett normed fit index for this model using the null model suggested by Jackson and Alwin (1980:235) is .66. By this criterion this may not appear to be a very good fit to the data, but given the difficulty of modeling ipsative data (see Alwin and Jackson, 1982b:211-12), it is not clear what an acceptable alternative would be.

[7] The results obtained by relaxing this constraint are not very different from those reported here.

two-factor model, along with the estimated factor intercorrelations and disturbance variances. As in the case of rankings, the rating data tend to cluster in a manner consistent with Kohn's conceptual framework. That is, the conformity-related characteristics of children tend to cluster together—obedience, good manners, neatness and cleanliness—and self-direction qualities also tend to cluster—curiosity, consideration, self-control, responsibility, and good judgment. However, instead of the clusters being opposed at polar ends of the same continuum as in the case of rankings, here the clusters are positively correlated (r=.65), which is

Table 4. Parameter Estimates for Two- and Three-Factor Models for Ratings of Child Qualities: NORC General Social Survey, 1980

	Two-Factor Model		*Three-Factor Model*		
Parameter	*I*	*II*	*I*	*II*	*III*
Factor Loadings					
Manners	.411*	.000[a]	.340*	.000[a]	.587*
Tries hard	.402*	.095	.149	-.197*	.587*
Honest	-.009	.266	-.363*	-.187	.587*
Neat & clean	.608*	-.038	.513*	.078	.587*
Good sense	.000[a]	.423*	.000[a]	.153*	.587*
Self-control	.175*	.321	.071	.079	.587*
Role	.945*	-.238*	.490*	-.182	.587*
Gets along	.425*	.090	.463*	.247*	.587*
Obeys	.431*	.035	.280*	.000[a]	.587*
Responsible	-.046	.503*	-.067	.328*	.587*
Considerate	-.030	.411*	.000[a]	.511*	.587*
Interested	.125	.396*	.064	.160*	.587*
Studious	.544*	.107	.377*	.053	.587*
Factor covariances	1.000[a]		1.000[a]		
	.648*	1.000[a]	-.223	1.000[a]	
			.000[a]	.000[a]	1.000[a]
Disturbance variances					
Manners		.632			.633
Tries hard		.632			.566
Honest		.748			.499
Neat & clean		.453			.433
Good sense		.570			.612
Self-control		.596			.614
Role		.414			.393
Gets along		.566			.495
Obeys		.612			.616
Responsible		.456			.482
Considerate		.592			.442
Interested		.619			.638
Studious		.419			.445
Goodness of fit					
Two-factor model: L^2=142.962, df=53, L^2/df=2.70					
Three-factor model: L^2=120.750, df=54, L^2/df=2.24					

[a] Fixed parameter.
* Factor loading is at least twice its standard error.

clearly inconsistent with the structure of ranking data. The fit of this model is only marginally acceptable; the ratio of the L^2 to degrees of freedom is 2.70.

Some of the lack of fit evident here may be due to this model's failure to explicitly take into account the effects of correlated response bias; we therefore estimated the parameters of a three-factor confirmatory factor model designed to account for these effects. The results for this three-factor model (shown in Table 4) and the improved goodness-of-fit measures demonstrate the importance of positing such a general method factor for rating data. The loadings on the method factor are statistically significant, and the improvement in fit of this model as compared to the two-factor model is significant ($\Delta L^2=22.21$, df=1).[8] Moreover, the correlation between the latent self-direction and conformity factors is now negative ($r=-.22$), rather than highly positive, although this correlation is not statistically significant. Thus, by including this method factor we improve the fit to the data and reduce the correlation between the two substantive latent factors to what we would expect on the basis of the ranking results. So by modeling the methodological properites of value ratings, we have found evidence of a latent structure in these data similar though not identical to that of the ranking data.

CRITERION-VALIDITY OF RATINGS AND RANKINGS

Even though the results of the factor analysis of ratings and rankings are different, we have identified concepts of self-direction and conformity as "latent" dimensions in both types of data. It is therefore worthwhile to examine the correlations between the latent dimensions and theoretically relevant predictor variables. For this purpose, we correlated the parental values factors with a set of predictors used in previous studies, namely, parental occupational prestige, education, and income. These socioeconomic indicators were chosen because their relationship to parental values has been the focus of much previous research (e.g., Lynd and Lynd, 1929; Duvall, 1946; Miller and Swanson, 1958; Lenski, 1961; Kohn, 1969, 1976, 1977, 1981; Alwin and Jackson, 1982b; Alwin, 1984).

The measurement of these variables in the 1980 NORC survey is as follows:

1. *Respondent's occupational prestige*—measured in the metric of Hodge-Siegel-Rossi scores (Siegel, 1971) assigned to 1970 U.S. Census detailed occupation codes.

[8] This is not a hierarchical test, since the models are not nested, but the improved fit is evident from the reduction in the ratio of L^2 to degrees of freedom. The Bentler-Bonett normed fit index for these models, using a null model which posits only random variance in the measures, are .91 and .93 for the two- and three-factor models, respectively.

2. *Respondent's education*—measured as the number of years of formal schooling completed.
3. *Family income*—measured as the total family income, from all sources, before taxes in the year preceding the survey. Our analysis assigns the midpoints of 12 income categories using $100 units.

Past research indicates that the contrast between self-direction and conformity, as measured by the latent factors identified above for the ranking data, correlates positively with these socioeconomic variables. On this basis we would expect the self-direction dimension of the ratings to correlate positively and the conformity dimension to correlate negatively with these variables. We estimated the relationships of these predictor variables with the latent self-direction/conformity factors underlying the rating and ranking data using linear structural equation (LISREL) models (Jöreskog and Sörbom, 1981).

Table 5. Zero-order Correlations of Latent Factors with Criterion Variables: NORC General Social Survey, 1980

Model	*Predictor*		
	Education	*Occupational Prestige*	*Income*
1. Rankings—single-factor nonipsative model			
I. Self-direction/conformity	.591*	.422*	.344*
2. Rankings—single-factor ipsative model			
I. Self-direction/conformity	.624*	.448*	.367*
3. Ratings—two-factor model			
I. Self-direction	.053	-.045	-.058
II. Conformity	-.355*	-.321*	-.319*
4. Ratings—three-factor model			
I. Self-direction	.036	.210*	-.062
II. Conformity	-.532*	-.267*	-.401*
III. Differential bias	-.060	-.181*	-.120

* Coefficient is greater than twice its standard error.

As anticipated, the self-direction/conformity factor underlying the rankings shows strong correlation with these indicators of socioeconomic position. Table 5 presents the zero-order correlations of the latent factors underlying our models for rankings and ratings with the set of theoretically relevant parental characteristics. The two factor models we have contrasted for ranking data, the ipsative common factor model and the nonipsative model, are quite similar in the magnitudes of correlations with these socioeconomic factors. The relationships are slightly higher for the ipsative factor model than for the nonipsative model, but the differences here are small. This is an important result because there

has been no research to date that has reported an empirical comparison of these two models.

In the rating data (see the results for the three-factor model) the "conformity" latent variable correlates negatively with the measures of socioeconomic position, as anticipated. The "self-direction" latent variable assessed by the ratings correlates positively with occupational prestige, but, contrary to our expectations, it is essentially uncorrelated with education and income. This is a striking difference between the results obtained using ratings and rankings. The comparable relationships for the two-factor version of the model are generally weaker, which again demonstrates the value of positing a general method factor for the rating data.

Summarizing these results, the latent factor underlying the ranking data is correlated with theoretically relevant predictor variables in a manner consistent with previous research. While the results for the rating data are not in dramatic contrast to these findings, generally weaker relationships are involved, as evidenced both by the weaker correlation between the latent self-direction and conformity dimensions and the correlations with the predictor variables. Moreover, the major difference between the results for ratings and rankings is the lack of correlation of the rating self-direction factor with the socioeconomic indicators. Thus, one would reach different substantive conclusions using ratings compared to the use of rankings.

Summary and Conclusions

Although ranking methods tend to be preferred for measuring social values, the empirical evidence available from past research suggests that rating techniques may be used just as effectively. In our analyses of parental values for children, we found that ratings and rankings produced similar results in terms of ordering the relative importance of value choices in the aggregate but are dissimilar with regard to latent structure. The main reason for the observed differences seems to be the different constraints introduced by the two techniques in the measurement process. The ranking technique forces value-choices to be generally negatively correlated, whereas the rating approach encourages positive correlations. As a consequence, the latent variable structures for the two sets of measures are quite different.

However, even after we took the method-induced positive and negative correlations between the items into account in our analyses, the latent structures of the rating and ranking items were different. A single bipolar factor was found to represent the latent content underlying the ranking measures, whereas two separate and distinct substantive factors

were found underlying the rating data. Although these two factors corresponded to the two concepts contrasted in the bipolar factor for rankings, the correlation between the two rating factors was not as strongly negative as the ranking analysis suggests it would be. The two types of measures produce different correlations between values and theoretically relevant predictor variables. All of these differences mitigate against the conclusion that ratings and rankings are interchangeable in the measurement of values.

Value researchers have consistently argued that the ranking technique is uniquely suited to the measurement of values, owing to the inherent comparative nature of values. From this point of view the present findings may be interpreted as support for this premise. However, these results are also consistent with a somewhat different interpretation; that is, rankings may impose a somewhat artificial contrast on the data and, as a result, measure both the latent dimension of contrast as well as the ability to see logical contrasts in the list of ranked qualities. It seems worthwhile to ask, in part because of our results using ratings, whether the ranking approach may in fact create artificial contrasts among the latent content of the measures.

Recall that the correlation between self-direction and conformity in the ranking data is in theory -1.0, since the two concepts define a bipolar factor. The latent variable underlying the set of rankings of child qualities has therefore been thought of as a continuum that distinguishes between the two extremes on a standard of desirability, i.e., the contrast between self-direction and conformity (Kohn. 1969). If one assumes that self-direction and conformity are contrasting values, as would be expected on the basis of the ranking data, the results for the rating data—that self-direction and conformity are only slightly related in a negative way—may seem incongruous. And if the values in question are in fact contrasting in this sense, one would expect the self-direction cluster of the ratings to correlate more strongly than it does with the socioeconomic predictor variables. Perhaps instead, self-direction and conformity are in actuality not contrasting values, in the sense that there is not a strong negative correlation between them. Is it not possible that the ranking technique essentially forces a contrast between them by asking respondents to make choices that they may not otherwise make? Since the conformity factor underlying the rating measures correlates nearly as strongly with the socioeconomic predictors as the contrast between self-direction and conformity in the ranking data, the correlation between the latent factor underlying the rankings and the socioeconomic predictors may be solely due to an association between the latter variables and values for conformity.

Although we raise these two differing interpretations of the differences

between the rating and ranking results, it is not possible on the basis of our present analysis to choose between them. To conclude that ratings and rankings measure somewhat different things when viewed in terms of the latent content involved does not necessarily lead to the additional conclusion that one is the more valid approach to measuring values. Until further research evidence is developed on these issues, particularly with respect to the extent to which rankings impose contrasts which do not occur using other methods, such a conclusion would be premature.

It may be difficult to generalize from these findings to the measurement of other kinds of value phenomena. One's choice of measurement approach should depend upon theoretical considerations, as much as on knowledge of the properties of various measurement techniques. The problems we have addressed in this paper are closely tied to a well-defined set of theoretical issues, and our analysis informs the choice of measurement strategy in this context. To the degree that other theoretical problems are similar to those studied here, there may be some correspondence between the measurement issues involved. More generally, in the absence of a priori theoretical knowledge about the content one wishes to measure, generalizations about the relative advantages of various measurement strategies are difficult. We look forward to further research that will bring more theory-based evidence to bear on the relative advantages and disadvantages of these and other approaches to the measurement of values.

References

Allport, G.W., P.E. Vernon, and G. Lindzey
1960 A study of Values. Boston: Houghton Mifflin.

Alwin, D.F.
1974 "Approaches to the interpretation of relationships in the multitrait-multimethod matrix." In H.L. Costner (ed.), Sociological Methodology 1973-74. San Francisco: Jossey-Bass.
1984 "Trends in parental socialization values: Detroit, 1958 to 1983." American Journal of Sociology 90:359-82.

Alwin, D.F., and D.J. Jackson
1979 "Measurement models for response errors in surveys: issues and applications." Pp. 68-119 in K.F. Schuessler (ed.), Sociological Methodology 1980. San Francisco: Jossey-Bass.
1982a "Adult values for children: an application of factor analysis to ranked preference data." Pp. 311-29 in R.M. Hauser, D. Mechanic, A.O. Haller, and T.S. Hauser (eds.), Social Structure and Behavior: Essays in Honor of William Hamilton Sewell. New York: Academic Press.
1982b "The statistical analysis of Kohn's measures of parental values." Pp. 197-223 in K. G. Jöreskog and H. Wold (eds.), Systems Under Indirect Observation: Causality, Structure, and Prediction. Amsterdam: North-Holland.

Anderson, T.W.
1958 An Introduction to Multivariate Statistical Analysis. New York: Wiley.

Andrews, F.M.
1984 "Construct validity and error components of survey measures: a structural modeling approach." Public Opinion Quarterly 48:409-42.
Bengtson, V.L.
1975 "Generation and family effects in value socialization." American Sociological Review 40:358-71.
Bentler, P.M.
1969 "Semantic space is (approximately) bipolar." Journal of Psychology 71:33-40
Berg, I.A.
1966 Response Set in Personality Assessment. Chicago: Aldine.
Block, J.
1965 The Challenge of Response Sets. New York: Appleton-Century-Crofts.
Cattell, R.B.
1944 "Psychological measurement: ipsative, normative and interactive." Psychological Review 51:292-303.
Clemans, W.V.
1966 "An analytical and empirical examination of some properties of ipsative measures." Psychometric Monographs 14.
Costner, H.L.
1969 "Theory, deduction, and rules of correspondence." American Journal of Sociology 75:245-63.
Cronbach, L.J.
1946 "Response sets and test validity." Educational and Psychological Measurement 6:475-94.
1950 "Further evidence on response sets." Educational and Psychological Measurement 10:3-31.
Cunningham, W. H., I. C. M. Cunningham, and R. T. Green,
1977 "The ipsative process to reduce response set bias." Public Opinion Quarterly 41:379-84.
Duvall, E. M.
1946 "Conceptions of parenthood." American Journal of Sociology 52:193-203.
Feather, N.T.
1973 "The measurement of values: effects of different assessment procedures." Australian Journal of Psychology 25:221-31.
1975 Values in Education and Society. New York: The Free Press.
Groves, R.M., and R. L. Kahn
1979 Surveys By Telephone: A National Comparison with Personal Interviews. New York: Academic Press.
Hamilton, D.L.
1968 "Personality attributes associated with extreme response style." Psychological Bulletin 69:192-203.
Horst, P.
1965 Factor Analysis of Data Matrices. New York: Holt, Rinehart and Winston.
Jackson, D.J., and D.F. Alwin
1980 "The factor analysis of ipsative measures." *Sociological Methods and Reseach* 9:218-38.
Jöreskog, K.G., and D. Sörbom
1981 LISREL—Analysis of Linear Structural Relationships by the Method of Maximum Likelihood. Version V. Chicago: National Educational Resources, Inc.
Kluckholn, F.R., and F.L. Strodtbeck
1961 Variations in Value Orientations. Evanston IL: Row, Peterson and Co.
Kohn, M. L.
1969 Class and Conformity: A Study in Values. Homewood, IL: Dorsey.
1976 "Social class and parental values: another confirmation of the relationship." American Sociological Review 41:538-45.
1977 "Reassessment 1977." In Kohn, Class and Conformity: A Study in Values. 2nd ed. Chicago: University of Chicago Press.

1981 "Personality, occupation, and social stratification: a frame of reference." In D. J. Treiman and R. V. Robinson (eds.), Research in Social Stratification and Mobility, Vol 1. Greenwich, CT: JAI Press.

Lenski, G.
1961 The Religious Factor. Garden City, NY: Doubleday.

Lynd, R., and H. Lynd
1929 Middletown: A Study in Contemporary American Culture. New York: Harcourt-Brace.

Messick, S.
1968 "Response sets." In D.L. Sills (ed.), International Encyclopedia of the Social Sciences, Vol 13. New York: Macmillan.

Miller, D.R., and G.E. Swanson
1958 The Changing American Parent. New York: John Wiley.

Moore, M.
1975 "Rating versus ranking in the Rokeach Value Survey: an Israeli comparison." European Journal of Social Psychology 5:405-08.

Munson, J.M., and S. H. McIntyre
1979 "Developing practical procedures for the measurement of personal values in cross-cultural marketing." Journal of Marketing Research 16:48-52.

National Opinion Research Center
1982 General Social Surveys, 1972-82: Cumulative Codebook. Chicago: National Opinion Research Center, University of Chicago.

O'Brien, R.
1979 "The use of Pearson's correlation with ordinal data." American Sociological Review 44:851-57.

Phillips, D.L.
1973 Abandoning Method. San Francisco: Jossey-Bass.

Rankin, W.L., and J.W. Grube
1980 "A comparison of ranking and rating procedures for value system measurement." European Journal of Social Psychology 10:233-46.

Reynolds, T.J., and J.P. Jolly
1980 "Measuring personal values: an evaluation of alternative methods." Journal of Marketing Research 17:531-36.

Rokeach, M.
1967 Value Survey. Sunyvale, CA: Halgren Tests, 873 Persimmon.
1970 Beliefs, Attitudes and Values. San Francisco: Jossey-Bass.
1973 The Nature of Human Values. New York: Free Press.

Schaeffer, N.C.
1982 "A general social survey experiment in generic words." Public Opinion Quarterly 46:572-81.

Siegel, P.M.
1971 "Prestige in the American occupational structure." Ph.D. dissertation, University of Chicago. Proceedings of the Association for Consumer Research.

Williams, R.M.
1968 "Values." In D.L. Sills (ed.), International Encyclopedia of the Social Sciences. New York: Macmillan.