

THE MEASUREMENT OF VALUES: WHEN THE LATENT BECOMES MANIFEST

Allen R. Wilcox, University of Nevada, Reno
Department of Political Science, Reno, Nevada 89557

Key Words: Measurement, Values

Values, however conceptualized and measured, are clearly important in the study of public opinion. The research presented here is intended to move the measurement and thus the usefulness of values as explanatory variables one or two steps forward. There are two primary ways by which values have been measured. One is to elicit responses to a set of value terms in the tradition of Milton Rokeach (1973). The second is to elicit responses to a set of statements, usually presented in a "agree-disagree" format, that are collectively thought to tap a specific value. The latter scale approach, through the logic of multiple indicators, is thought to provide more reliable measurement. Unfortunately, it also takes up much more questionnaire space, thus limiting the number of values that can be included in a survey. This is an example of a more general problem of under sampling of theoretically relevant variables. The representative sampling of individuals (or other units of analysis) typically receives considerable attention in large-scale social research. The sampling of variables receives less systematic attention, the result being that many statistical analyses are probably misspecified because crucial variables are not represented in the survey by appropriate questions. The task, then, is to develop procedures that will allow the use of single item measures (or perhaps shorter scales). In the present context, if this can be successfully accomplished, more value terms can be included in a survey, increasing theoretical adequacy. The result should be surveys that can do greater justice to the rich value context in which opinions are embedded.

COMPARATIVE ADVANTAGES

The inspiration for this paper came primarily from an article by Stanley Feldman on "Structure and Consistency in Public Opinion: the Role of Core Beliefs and Values" (1988). In that piece, concerned with the validation of value scales developed for an NES Pilot Study, Feldman observed that multi-item value scales might be more useful than the familiar single item value labels used by Milton Rokeach in the *Nature of Human Values* and many other publications. Although he did not make this specific argument, greater utility might come from the enhanced reliability obtained when multiple indicators of a

concept (in this case, a value) are combined into a scale, thus mitigating the random error contaminating each individual indicator. With this perspective there is no argument.

However, the use of multi-item scales does have the disadvantage of using up precious survey space. The inclusion of single item indicators allows the inclusion of a wider range of concepts and enhances the possibility of more complex and better specified analyses. In addition, single item indicators may have greater face validity, particularly if they are comprised of a value label as in the Rokeach tradition. Other possible advantages of single items include the use of archival data in which single-item measures are present, higher response rates, and easier communication to a non-technical audience (Johnson and Rejda, 1996). The question thus becomes what criteria might be used to judge when a single item is an adequate substitute for a scale? This paper suggests what such criteria might be and provides a set of preliminary examples of their application.

METHODOLOGY

The analysis utilizes data produced by a self-administered questionnaire collected from a convenience sample of students taking social science courses at a western university. The questionnaire included several multi-item scales intended to measure values in the standard agree-disagree format. It also included a set of Rokeach value items presented, not as a ranking exercise, but as eleven point rating scales ranging from a negative -5 through 0 to a positive +5. Several other sets of items used in validation will be described below.

The criteria used to judge the potential adequacy of a single-item substitute for a multi-item scale are as follows:

Face Validity. The single item value label must have a plausible, and preferably compelling semantic equivalence to the content of the multi-item scale.

Convergent Validity. The single item must satisfy the following statistical criteria:

Factor Analysis. The item must have approximately as high (and preferably higher) a loading on a single factor as any of the items included in the competing multi-item scale.

Reliability Analysis. The item must not decrease the alpha coefficient of the scale when added to the scale items.

Correlation Analysis. The item must be correlated with other items taken from the instrument at a level approximately as high as any scale item and high enough compared to the scale score so as not to significantly impinge on substantive conclusions. Item sets used for this purpose in this study were:

1. 23 policy items using the NES 7-point scale with verbal summaries at each end.
2. 48 value items using an 11-point negative to positive scale (The large number of response categories was used to compensate for the tendency of respondents to congregate at the positive end of the scales).
3. 26 group thermometer items.
4. 20 institutional confidence items using a 5-point scale anchored by "hardly any" and "a great deal".

Discriminant Validity. The item must satisfy the above criteria better than any proposed rivals. This criterion, for reasons of space, will be explored systematically in another article.

RESULTS

The application of these criteria will be illustrated with four scales which, it is suggested, may represent the measurement of the values of equality, racial equality, free enterprise, and either national security or patriotism.

1. EQUALITY: The Best Case

Single Item Wording (Factor loadings in parentheses):

Equality (brotherhood, equal opportunity for all) (.735)

Scale Item Wordings:

If people were treated more equally in this country, we would have many fewer problems. (.693)

We should give up on the goal of equality, since people are so different to begin with. (-.685)

Our society should do whatever is necessary to make sure that everyone has an equal opportunity to succeed. (.650)

One of the big problems in this country is that we don't give everyone an equal chance. (.626)

Cronbach's Alpha:

Without single item: .759 With single item added: .771

Correlations: Correlations of single item with other items is comparable to (at least as high on average) all scale items.

2. FREE ENTERPRISE: A Close Case

Single Item Wording (Factor loadings in parentheses):

Free Enterprise (absence of government control) (.670)

Scale Item Wordings:

The less government gets involved with business and the economy, the better off this country will be. (.800)

There are many goods and services that would never be available to ordinary people without governmental intervention. (-.564)

There should be no interference with business and trade. (.736)

Putting governmental regulations on business does not endanger personal freedom. (-.645)

Government intervention leads to too much red tape and too many problems. (.776)

Cronbach's Alpha:

Without single item: .830 With single item added: .753

Correlations: Correlations of single item with other items is comparable to (at least as high on average) all scales.

3. RACIAL EQUALITY: A Bad Case

Single Item Wording (Factor loadings in parentheses):

Racial Equality (equal treatment and opportunity for the races) (-.501)

Scale Item Wordings:

Irish, Italians, Jewish and many other minorities overcame prejudice and worked their way up. Blacks should do the same without any special favors. (.775)

Over the past few years blacks have gotten less than they deserve. (-.809)

It's really a matter of some people not trying hard enough; if blacks would only try harder they could be just as well off as whites. (.866)

Generations of slavery and discrimination have created conditions that make it difficult for blacks to work their way out of the lower class. (-.738)

Equal opportunity for blacks and whites is very important but it's not really the government's job to guarantee it. (.702)

Cronbach's Alpha:

Without single item: .880 With single item added: .841

Correlations: Correlations of single item with other items is inconsistent with scale item correlations (sometimes higher, sometimes lower) but in ways that make sense if one assumes the single item and the scale items are measuring different concepts.

4. NATIONAL SECURITY and PATRIOTISM: Alternative Cases

Single Item Wording (Factor loadings in parentheses):

National Security (protection from attack) (.424)

Patriotism (devotion to country) (.612)

Scale Item Wordings:

The U.S. should maintain its position as the world's most powerful nation even if it means going to the brink of war. (.587 - Nat) (.656 - Pat)

Any time a country goes communist, it should be considered a threat to the vital interests and security of the United States. (.815 - Nat) (.776 - Pat)

The United States should do everything it can to prevent the spread of communism to any other part of the world. (.881 - Nat) (.835 - Pat)

Cronbach's Alpha:

Without single item: .796 (Nat) .792 (Pat)
With single item added: .698 (Nat) .724 (Pat)

Correlations: Correlations of both single items with other items is similar but Patriotism correlations are fairly consistently higher than National Security and more similar to scale item correlations

SUMMARY

The four examples provide a range of possible outcomes when trying to match a concept label (with a short parenthetical definition) to a set of Likert scale statements. In the "Equality" case, all criteria were well met and the single item can probably be substituted for the four Likert items without serious loss of conceptual cogency or statistical power. (Note: The original Likert scale was intended to measure "Egalitarianism" and it is not clear whether that label would perform as well.) In the "Free Enterprise" case, the criteria are less well met with the decline in Cronbach's alpha when adding the single item to the scale particularly disturbing. The poorly performing "Racial Equality" case is one in which the Likert items were originally intended to measure "Effort" and

where another label such as "Racial Individualism" might fare better. In the "National Security" vs. "Patriotism" case, the latter concept is a better match to the content of the Likert items, but it still falls short of the performance of "Equality". If the NES items explicitly designed to measure patriotism had been included on this survey, the surmise is that an equally successful case would have resulted. New research under way will test the viability of single item substitutions with a much larger set of values and other concepts and will utilize multi-trait multi-methodology validation procedures.

CONCLUSION

The research reported here establishes at least some of the conditions under which a single item may be usefully substituted for a multi-item scale (These criteria may, of course, evolve and others may be added). In the past, names or labels applied to sets of items have been theoretically derived or, in the case of much exploratory factor analysis, have been based upon perceived commonalities among the items. Considerable uncertainty often remains as to whether the items are truly measuring what is implied by the label attached to the scale. That uncertainty can be reduced a great deal by applying the above validity criteria. When those criteria are satisfied, a latent variable can become manifest (or, at least, very close to it). A single item that names a factor can replace a set of items thought to represent that factor. More concepts can be included in surveys, thus enriching analysis. On the other hand, a researcher who chooses to stay with a multi-item scale because of its greater reliability can also have more confidence in its validity. Both approaches benefit.

The above logic holds, of course, only for concepts that have names that are common currency in the subject sample and population. Most importantly, if the concept's label is understandable in the general population (equality, say, as opposed to Machiavellianism), relatively small convenience samples can be used to establish validity instead of large probability samples. This, in turn, means that more researchers can contribute to the measurement of values (and other concepts) in a decentralized and collaborative manner. All that remains is the small matter of establishing a mechanism for sharing the results of such research. Finally, it should be noted that the above procedures are not static. The equivalence of single items and multi-item scales should be tested on a periodic basis for changes in the semantic connotations of either or both.

References:

Feldman, Stanley (1988) , "Structure and Consistency in Public Opinion: the Role of Core Beliefs and Values," *American Journal of Political Science*, 32, 416-440.

Johnson, David R., and Karen Rejda (1996), "Single-Item Measures: When are they valid?", Poster Presentation at the American Association of Public Opinion Research Annual Meeting, Salt Lake City, Utah.

Rokeach, Milton (1973), *The Nature of Human Values*, New York: Free Press.