

The Dimensionality of Bipolar Scales in Self-Description

Alan J. Klockars
University of Washington

Daniel W. King and Lynda A. King
Central Michigan University

Sets of bipolar scales were constructed for self-description for 13 different traits. Within each trait four scales, which differed in the relationship between the trait dimension and the desirability of the endpoints, were developed. Two scales had either both desirable or both undesirable endpoints. The other two scales had one desirable and one undesirable endpoint but differed from one another in the direction of the trait dimension. Self-descriptions were obtained from 606 students on the 13 sets of scales and bipolar marker scales to measure the dimensions of evaluation, potency, activity, and familiarity. In addition, each student answered a true-false social desirability scale. The data were factor analyzed and rotated to simple structure. The factors closely reflected the trait dimensions. There were no factors that could be interpreted as either a social desirability or evaluation factor. The correlations of the bipolar scales that had differences in desirability of the endpoints averaged .12 with the social desirability scale and .13 with the evaluation marker scale. The correlations between the scales within each trait set reflected primarily the trait relationships but seemed to be moderated by the effects of evaluation or desirability. Scores were obtained on the sum of the four scales within each trait dimension. These scores were reasonably internally consistent and uncorrelated with social desirability. The potential for this method of personality assessment is discussed.

Bipolar rating scales anchored by semantic labels at the endpoints have been primarily associated with Osgood, Suci, and Tannenbaum's (1957) semantic differential technique to determine the underlying structure of connotative meaning. In their classic work, Osgood et al. (1957) demonstrated an underlying three-factor structure of generalized semantic space: evaluation, potency, and activity (EPA), with evaluation accounting for the greatest amount of variance. The present study explored the dimensionality of bipolar scales when used in a more restricted sense: in self-descriptive personality assessment. Factor analyses have been reported by Warr and Haycock (1970) and Gray (1973), which were interpreted as supporting an EPA structure in self-description with bipolar scales. Because of the scales used in the ratings and the specific context of the ratings, neither can be considered as specifically addressing the dimensionality of a cross-section of traits used solely for self-description.

In traditional personality assessment, Edwards and his associates (see Edwards, 1970) have argued for the confounding effect of a social desirability (SD) response tendency on responses to a variety of structured personality instruments. When personality inventories such as the Minnesota Multiphasic Personality Inventory (Hathaway & McKinley, 1951) are factor analyzed, the loadings of the scales on the first

factor tend to be proportional to the extent to which the scale keying is confounded with keying for SD responses. A scale constructed by Edwards (1957) to measure the rate of SD responding has typically loaded highly on the first factor of personality inventories.

The SD of a bipolar rating scale has been defined by Ford and Meisels (1969) as the difference in the desirabilities of the adjectives associated with the two endpoints. A rating scale marked at one end by a highly desirable adjective and at the other end by a highly undesirable adjective would be more susceptible to the influences of SD than a scale with less of a difference between the endpoints. This definition of a socially desirable bipolar rating scale is thus synonymous with Osgood et al.'s (1957) evaluative scale. Ford and Meisel (1969) found a correlation of .92 between the differences in the desirability of the endpoints of a scale and its loading on the evaluative factor of the semantic differential.

Based on the factor analysis of semantic differential data and the SD research with personality data, it might be expected that a large evaluation-SD factor would be the primary dimension in self-description with bipolar rating scales. However, in both of these research areas, the potential for trait factors has been minimized because of the nature of the stimuli used. The bipolar scales used in semantic differential studies were selected to cover a wide variety of attributes useful in describing a wide variety of concepts. There was no intent to build into the set of scales the potential for measuring personality traits. Under these conditions the ability to obtain specific trait factors was obviated and the underlying dimensions of evaluation, potency, and activity emerged. With SD research, the question has not been the obtaining of a trait factor, but being able to unambiguously interpret the factor as such. Given the same factor structure, Edwards and Heathers (1962) concluded that the first factor measured SD, while Block (1965) provided a trait interpretation and called the factor ego resiliency.

In order to distinguish trait from evaluation-SD, there should be a number of items that measure the same trait where the trait keying is not completely confounded with the SD of the items. Adjective sets from Peabody (1967) provide ideal stimuli for constructing these bipolar scales. Each set of four adjectives represents a single descriptive trait dimension, with two of the words marking one end of the dimension and the other two words marking the opposite end of the dimension. Within the two words marking a common end of the dimension, one adjective is positively valued (high social desirability scale value or SDSV) and the other is negatively valued (low SDSV). For each set of four adjectives it is possible to construct four bipolar scales, all representing opposites in descriptive content but varying in their evaluative patterns. For example, one set of words representing a risk-taking dimension consists of the quadruplet CAUTIOUS-TIMID-BOLD-RASH. From these words it is possible to form the following four bipolar scales: (1) CAUTIOUS (high SDSV)—BOLD (high SDSV); (2) CAUTIOUS (high SDSV)—RASH (low SDSV); (3) TIMID (low SDSV)—BOLD (high SDSV); and (4) TIMID (low SDSV)—RASH (low SDSV). Scales 1 and 4 eliminate evaluation by having similar SDSVs for the two endpoints. Scales 2 and 3 both increase with respect to risk-taking, but Scale 2 goes from a positively valued to a negatively valued endpoint, while Scale 3 goes from a negatively valued to a positively valued endpoint.

If the SD or evaluation dimension is predominant in self-ratings, scales such as 2 and 3 above should load on a major factor with opposite signs to reflect the evaluative keying of the scales. This factor should have high loadings for the SD scale and for scales that normally mark the evaluation dimension. Scales such as 1 and 4 should load on separate smaller factors reflecting trait effects. On the other hand, if the trait factors emerge, all four scales should load on a common factor, with their loadings reflecting the trait dimension.

Method

Using 13 of Peabody's (1967) complete sets of four-adjective trait terms, a series of 7-point bipolar scales was constructed. Each set of four adjectives yielded four scales, two with endpoints equated on evaluation and two with endpoints of opposite evaluation. The four scales of each descriptive content area were divided into two subsets of two scales each, such that each adjective was used as an endpoint only once in each subset. The two subsets for each content area were randomly assigned to one of two groups. The total of 52 scales was thus organized into two groups of 26 scales each.

An additional 12 scales were selected to provide three marker scales each for Osgood et al.'s (1957) EPA structure and the familiarity factor found by Nunnally (1961). These scales were randomly distributed to one of the two scale groups such that all of the factors were represented in each group. The final two scale groups, containing 32 scales each, were separately placed in random orders and used as Parts 1 and 3 of the instrument.

Parts 1 and 3 instructions to respondents were similar to the standard semantic differential procedure, with emphasis on the requirement for rating the concept "myself." The instructions for Part 3 asked respondents not to refer to their Part 1 ratings or to try to recall their responses to the earlier (and similar) scales. Part 2 of the instrument contained 38 items from the Edwards (1957) SD scale¹ given under standard self-description instructions.

The booklets were administered in classroom groups to undergraduate psychology students at Central Michigan University, Mt. Pleasant, Michigan, and undergraduate business students at Northwood Institute, Midland, Michigan. A total of 606 respondents voluntarily completed the booklets, which required approximately 25 minutes. After the data were collected, classes were given a description of the experiment.

Responses to the SD items were coded and summed to obtain a single score for each respondent. Responses to the 64 bipolar scales and SD score were intercorrelated and submitted to principal components analysis with varimax rotation.

Results and Discussion

Table 1 presents the results of the rotated principal components analysis. Eighteen factors had eigenvalues greater than 1.0 and were retained in the rotated factor structure. As is evident from the pattern of loadings of the four scales from each common trait domain, the factors produced by this analysis are most easily interpreted as reflecting traits. In none of the factors are the loadings of scales contradictory to a trait interpretation in favor of an evaluative or SD interpretation. The last two columns of Table 1 present the correlations between the bipolar scales and the SD scale and between the bipolar scales and an evaluative scale produced by summing the three evaluative marker scales included in the instrument. When only the 26 bipolar scales that had the differences in the SDSVs of the endpoints were considered, the average correlation between ratings on the bipolar scales and the SD score was .12. Of the 26 correlations, 20 were in a direction consistent with an evaluative-SD interpretation, while 6 were not. The average correlation for bipolar scale ratings with the evaluative score was .13, with 22 in an evaluatively consistent direction and 4 not. The correlation between the SD score and the evaluative score was .30.

The findings are somewhat difficult to reconcile with the amassed literature that shows the importance of SD in self-description and the importance of the evaluative dimension within the semantic differential. There was no large first factor with high loadings for either the SD scale or the evaluative scales. Rather, the first factor of the unrotated factor structure had an eigenvalue of 6.95, accounting for only 10.7% of the variance. The SD scale had loadings on a num-

¹Item 38 was omitted because of its offensive character.

Table 1
Factor Loadings and Correlations with SD and
Evaluation for Bipolar Rating Scales

Set Bipolar Scale	Factor (Loading)	r SD	r EVAL
1 Cautious-Bold	1 (-33); 2 (45)	20	11
Rash-Timid	1 (56)	-19	-04
Cautious-Rash	2 (49)	09	-01
Bold-Timid	1 (55)	-27	-14
2 Idealistic-Pragmatic	13 (81)	06	-01
Opportunistic-Unrealistic	12 (59); 15 (-32)	-14	-23
Idealistic-Opportunistic	13 (73)	06	00
Pragmatic-Unrealistic	12 (70)	-21	-11
3 Self-controlled-Uninhibited	2 (70); 10 (-60)	-03	-07
Impulsive-Inhibited	2 (-48)	-06	-02
Self-controlled-Impulsive	2 (57); 9 (30)	-14	-14
Uninhibited-Inhibited	2 (-49)	-26	-09
4 Skeptical-Trusting	3 (-58)	15	22
Gullible-Distrustful	3 (80)	07	-01
Gullible-Skeptical	3 (74)	13	09
Distrustful-Trusting	3 (-58)	17	36
5 Thrifty-Generous	5 (80)	-01	15
Extravagant-Stingy	5 (-80)	-04	-02
Extravagant-Thrifty	5 (-79)	-03	01
Stingy-Generous	5 (49)	08	21
6 Selective-Tolerant	4 (69)	04	-14
Undiscriminating-Choosy	4 (-79)	-08	05
Selective-Undiscriminating	4 (72)	01	08
Tolerant-Choosy	4 (-65)	-07	-05
7 Lenient-Firm	11 (66)	08	-02
Severe-Lax	11 (-69)	-04	04
Lenient-Severe	11 (75)	-05	-09
Firm-Lax	9 (35); 11 (-39)	-15	-08
8 Individualistic-Cooperative	7 (67)	-01	02
Conforming-Uncooperative	7 (-62)	01	-05
Individualistic-Conforming	7 (73)	-19	-14
Cooperative-Uncooperative	3 (30); 7 (-44); 16 (44)	-09	-22
9 Relaxed-Alert	6 (-42); 15 (50)	-06	-04
Tense-Lethargic	6 (78)	27	17
Tense-Relaxed	6 (77)	43	27
Lethargic-Alert	15 (75)	24	30

(continued on next page)

Table 1 - Continued
 Factor Loadings and Correlations with SD and
 Evaluation for Bipolar Rating Scales

Set Bipolar Scale	Factor (Loading)	r SD	r EVAL
10 Modest-Confident	8 (-57)	22	23
Conceited-Self-disparaging	8 (81)	-19	-03
Modest-Conceited	8 (-83)	02	02
Confident-Self-disparaging	8 (31); 15 (-35)	-37	-30
11 Frank-Discrete		-12	-08
Secretive-Indiscrete	10 (71)	10	07
Secretive-Frank	10 (83)	16	11
Indiscrete-Discrete	14 (53)	03	14
12 Steady-Flexible	9 (60)	-05	20
Vacillating-Inflexible	9 (-48)	-05	-20
Steady-Vacillating	9 (56)	-11	-03
Flexible-Inflexible	9 (-35); 11 (30)	-04	-26
13 Open-minded-Committed	16 (73)	-09	-21
Fanatical-Noncommittal	18 (67)	01	08
Fanatical-Open-minded	16 (-70)	04	36
Noncommittal-Committed	18 (-77)	-01	05
Evaluation			
Good-Bad	17 (-68)	-25	
Unfair-Fair	17 (63)	16	
Worthless-Valuable	12 (-31); 17 (51)	26	
Potency			
Hard-Soft	1 (61)	-07	
Rugged-Delicate	1 (75)	-14	
Weak-Strong	1 (-48); 15 (37); 17 (34)	32	
Activity			
Active-Passive	1 (36); 15 (-47)	-19	
Slow-Fast	15 (61)	25	
Calm-Excitable	2 (33); 6 (-38); 11 (31)	-16	
Familiarity			
Predictable-Unpredictable	7 (-41); 17 (-33)	-00	
Simple-Complex	14 (63)	12	
Unusual-Usual	7 (49); 14 (-31)	02	
Social Desirability Scale	6 (51); 16 (48); 17 (30)		

Note Factor number reported for loading .30 or greater, decimals omitted.

Table 2
Intercorrelations Among Scales in Each Set

Set	Correlated Pairs			
	++with--	++with+-	--with+-	+--with+-
1	.39	.40	.33	.29
2	.00	.51	.57	.11
3	.19	.44	.20	.11
4	.36	.18	.30	.25
5	.52	.39	.47	.15
6	.40	.20	.29	.27
7	.49	.27	.67	.18
8	.34	.58	.43	.18
9	.24	.54	.59	-.14
10	.34	.45	.34	.16
11	.27	.37	.62	.05
12	.25	.45	.39	.05
13	.10	.48	.47	.00
		.37	.42	
		.46	.40	
		.30	.40	
		.26	.50	
		.26	-.05	
		.48	.61	
		.41	.24	
		.47	.47	
		.17	.22	
		.20	.26	
		.32	.34	
		.46	.16	
		.21	.30	

ber of factors instead of a major loading on one factor.

Table 2 presents the intercorrelation among the four scales within each descriptive content area. The correlations in column 1 are between the scale with both endpoints positive and the scale with both endpoints negative. Column 2 presents the correlations of the two scales having one positive and one negative endpoint with the scale having both endpoints positive. Column 3 presents the correlations between the two scales having a positive and a negative endpoint and the scale having both endpoints negative. The

correlations in column 4 are between the pair of scales each with a positive and a negative endpoint.

The correlations in column 1 represent solely trait relationships. Both scales are independent of SD, since they have equally desirable endpoints. In addition, each of the four adjective endpoints occurs once and only once within the two scales; that is, there is no repetition of endpoints between scales. The average of the correlations in column 1 is .30.

The correlations in columns 2 and 3 are influenced both by the trait and by a common end-

point within the pair of correlated scales (e.g., CAUTIOUS-BOLD and CAUTIOUS-RASH). They should not be influenced by SD since one of the correlated scales is neutral on SD. Although scales having a common endpoint were physically separated by the SD scale on the instrument, the possibility that ratings were somehow anchored by having the same endpoint presented a second time cannot be dismissed. The average correlations for columns 2 and 3 are .37 and .38, respectively.

The correlations in column 4 place evaluation-SD and the descriptive dimension in direct opposition. With one exception, the correlations are positive, suggesting that the descriptive dimension is stronger than the evaluative dimension. However, the average of the correlations is .13, indicating a significant reduction in the expected correlation due to the descriptive content alone. Although the descriptive dimensions may be more important than evaluation, it is not possible to completely discount the influence of evaluation on ratings.

The question remains as to why the usual evaluation factor associated with semantic differential ratings and SD factor associated with personality measurement did not appear in the present study. A rationale for this outcome may be derived from the arguments of Peabody (1970) and Felipe (1970) with regard to trait inferences. Both argued that evaluation and description are involved when individuals use bipolar scales. In general, their proposition is that when the rating or inference task involves relating some concept or characteristic to a dimension that is descriptively pertinent (or highly content-associated), the rater attends to the descriptive aspects of the scale. Given the task of relating a concept or characteristic to a scale that is descriptively inappropriate, the rater will seek to achieve evaluative consistency in the absence of descriptive consistency.

It may be surmised that given the task of rating "myself" on a series of scales that were specifically person-related or personality descriptive, the raters behaved according to the Peabody-Felipe proposition. Raters attended to

the descriptive aspects of the scales causing the usual evaluation-SD factor to fail to emerge in the factor structure.

An implied issue in these findings concerns the usefulness of bipolar rating scales as a method of personality assessment. As with most types of stimuli, the individual bipolar scale item would not be expected to have sufficient reliability to be useful as a general measure of personality. Instead, the values of responses to several bipolar scales would be summed. For the stimuli provided by Peabody (1967), four scales can be summed for each trait area. Using the present data, coefficient alpha estimates of internal consistency for each set were calculated. These values are present in Table 3. In general, the values are reasonably high, considering that only four items contribute to the summative score. The lower values are for those sets that did not load on a common factor in the factor analysis of the bipolar scales and in general seem most poorly matched on the content dimension. It appears that for well-defined content dimensions, it is possible to obtain internally consistent measures of personality with a small number of scales.

The sums of the four bipolar scale ratings should be independent of SD, since the individual bipolar scales were relatively uncorrelated with the SD scale, and the sum counterbalances what minor effects might be attributed to desirability. The correlations of the sums with the SD scale and the sum of the three evaluative scales are presented in the last two columns of Table 3.

The data on the sums of the bipolar scales as measures of personality indicate considerable potential. Using the results of taxonomies of trait descriptive terms such as provided by Goldberg (1977) and Wiggins (1979), adjectives could be selected to construct bipolar scales that reflect a common trait and retained the configuration of endpoints necessary to control the influence of social desirability. Based on the Spearman-Brown formula, the median internal consistency of the sums of the scales reported in Table 3 could be raised to approximately .8 by constructing 8 scales for each trait. With highly

Table 3
Coefficient Alpha, Correlation with SD and
Correlation with Evaluation for Sums of Four Scales

Set	Alpha	r SD	r EVAL
1	.73	.26	.10
2	.51	.17	.12
3	.59	-.05	.05
4	.75	.04	.17
5	.75	.05	.13
6	.74	.07	.05
7	.73	-.08	.01
8	.66	-.06	.03
9	.48	.25	.11
10	.69	.28	.21
11	.61	-.13	-.04
12	.51	-.04	.22
13	.47	-.06	-.23

internally consistent scales free from the influence of social desirability, the concerns of retest reliability and construct validity could be addressed. Only when all these steps have been completed will the potential of bipolar scales as a method of personality assessment be known.

References

- Block, J. *The challenge of response sets*. New York: Appleton-Century-Crofts, 1965.
- Edwards, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.
- Edwards, A. L. *The measurement of personality traits by scales and inventories*. New York: Holt, Rinehart, & Winston, 1970.
- Edwards, A. L., & Heathers, L. B. The first factor of the MMPI: Social desirability or ego strength? *Journal of Consulting Psychology*, 1962, 26, 99-100.
- Felipe, A. Evaluative versus descriptive consistency in trait inferences. *Journal of Personality and Social Psychology*, 1970, 16, 627-638.
- Ford, L. H., & Meisels, M. Social desirability and the semantic differential. In J. G. Snider & C. E. Osgood (Eds.), *Semantic differential technique: A sourcebook*. Chicago: Aldine, 1969.
- Goldberg, L. R. *Language and personality: Developing a taxonomy of trait-descriptive terms*. Invited address to the division of Evaluation and Measurement at the 86th annual convention of the American Psychological Association, San Francisco, August 1977.
- Gray, J. E. Dimensions of personality and meaning in self-ratings of personality. *British Journal of Social and Clinical Psychology*, 1973, 12, 319-322.
- Hathaway, S. R., & McKinley, J. C. *Minnesota Multi-phasic Personality Inventory* (rev. ed.). New York: Psychological Corporation, 1951.
- Nunnally, J. C. *Popular conceptions of mental health: Their development and change*. New York: Holt, Rinehart, & Winston, 1961.
- Osgood, C. E., Suci, G., & Tannenbaum, P. *The measurement of meaning*. Urbana: University of Illinois Press, 1957.
- Peabody, D. Trait inferences: Evaluative and descriptive aspects. *Journal of Personality and Social Psychology Monograph*, 1967, 7 (Whole No. 644).
- Peabody, D. Evaluative and descriptive aspects in personality perception: A reappraisal. *Journal of Personality and Social Psychology*, 1970, 16, 639-646.

- Warr, P., & Haycock, V. Scales for a British personality differential. *British Journal of Social and Clinical Psychology*, 1970, 9, 328-337.
- Wiggins, J. S. A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, 1979, 37, 395-412.

Author's Address

Send requests for reprints or further information to Alan J. Klockars, 312 Miller Hall, DQ-12, University of Washington, Seattle, WA 98195.