The Impact of Format-Specific Response Bias on Faces Scales

Samuel J. Yeager and John W. Bardo

This study examines the impact of subject bias in responding to faces formats, Likert formats, and other types of commonly used response formats. Students (*N* = 292) in a large urban university were instructed to respond randomly to several pages of contentless items. This procedure makes it possible to determine response formats bias without contamination from item content. Result indicate that faces formats are subject to considerably more positive bias than Likert and numerical formats and, consequently, that faces formats should not be used in empirical research.

A. INTRODUCTION

The present study describes the relative effects of two types of systematic response bias affecting faces and other types of measurement scales, a method for directly assessing their impact, and the results of applying the technique to a large sample of students.

One widely used measure of job satisfaction of the faces scale developed by Kunin (15) which consists of several drawings of human faces ranging in expression from a deep frown through neutral to a very broad smile or grin. Kunin argued that this measure has the advantage of requiring little verbal or reading ability. While original development was based on little evidence of reliability and validity, its use has subsequently included cross validation of several different kinds of other measures such as communications (19) and job satisfaction measures (20). A female faces scale (5) has also been developed. Others (4, 8, 12, 21) beside Kunin have used his measure to assess job satisfaction and satisfaction with past and future advancement opportunities. Faces scales have also been used to measure attitudes of school children toward classroom subjects (7).

One recent study has raised serious questions about the usefulness of faces scales. Ironson and Smith (11) investigated the impact of shifting the neutral point or anchor of verbal- and faces-formatted scales in either a positive or negative direction. They found that faces scales are affected more than verbal scales by these changes in contexts and that faces scales do not retain their meaning as well as verbal scales when shifts of this kind are made. In addition, they report that faces scales as they presently exist may not be useful in other societies, such as Hong Kong and the Philippines, where Anglo-American or western faces all appear alike. Since faces-formatted scales continue to be used in recent analyses of organizational behavior, further validation and assessment of the procedure's adequacy are needed. Here the impact of certain types of systematic response bias on faces formats will be examined and compared with other response formats.

Measurement error is a major problem in social psychological research. Its occurrence simply means that an individual's responses to questionnaire items do not accurately reflect his or her actual feelings, perceptions, or "true scores" (18). This results in less reliable and less useful research.

Several types of systematic response bias exist-see (1) and (18) for a thorough review of these measurement problems-the most significant of which are leniency error-the tendency of a respondent to rate an item too high or too low-and central tendency error-the hesitancy of respondents to give extreme scores. Both errors result from the respondents' orientation or psychological reaction to the form of the test or response format. These form-related errors do not involve respondents' orientation toward item subject matter. Thus, leniency error results from the respondent not wanting to be a "nay sayer," and central tendency error is a consequence of the respondent not wanting to be perceived as extreme (6).

In addition, these different types of format-related errors are expected to have different effects. Central tendency and leniency errors should affect the mean and variance of any distribution of scores in which they occur. Leniency effort should cause the distribution of scores to be skewed low. Central tendency error should reduce the expected variance of the distribution. For example, any random variable with a range of 1 to 5 has an expected mean (μ,) of 3 and an expected variance of (σ^2) of 2. Where the errors of leniency and central tendency affect its distribution, $\mu_, \neq 3$ and $\sigma^2 \neq 2$.

Reduction of variance and shifting of the expected mean are particularly problematic in empirical research. Summated rating scales are most often employed as variables in multivariate research designs based on measurement statistics and their derivatives. Reduction in variance caused by a restricted range decreases the efficiency of measurements which are already limited by the predetermined range of the response format (13, 16). This may have disastrously misleading effects on correlation and regression analyses (14) and may confound significance of difference tests, such as *t*, *F*, and Cooper's (3) Exact Probability Test, because underlying assumptions concerning both expected means and variances are violated.

B. METHOD

In order to analyze the impact of formate-related errors on faces scales, a test free of subject matter was devised consisting of a series, of response formats for faces and other formats frequently used in empirical research. Use of contentless items is not a new technique. It has frequently been used in the past to estimate the possible effects of response bias on various forms of fixed response formats (2, 10, 17, 22). These earlier studies paid no attention to possible effects of variations in response style across formats. Here, this issue is more closely examined.

The test used here contains the following response formats: schematic drawings of human faces (five faces to a line) ranging from broad smiles to deep frowns drawn so that one set had long hair (female faces), one had short hair (male faces), and a third had no hair (neuter faces); five-position Likert-formats in which responses ranged from "strongly agree" to "strongly disagree"; and five-point numerical formats, all starting with the number 1.

Each format was presented in two clusters of 10 lines. Half of the formats were reversed in each case. In total, the "questionnaire" consisted of 11 pages of clusters of response formats. Pages within questionnaires were randomized so that respondents would not view the formats in the same order.

Instructions accompanying the packet of response formats were designed to maximize the likelihood of generating approximations of random variables. After the *E*s introduced themselves, they instructed respondents to complete the forms so that their responses would show no pattern or would be random. The following instructions were read aloud:

> Many things in this world occur in such a way that they do not have a particular pattern or relationship to one another. For instance, scuff marks on the floor or wall do not appear in any particular pattern. Where things do not have a identifiable pattern, we say that they are random. We would like you to mark the pages in front of you such that your marks make no pattern; that is, you should respond randomly. Please make only one mark per line; work quickly and do not change your marks.

Since this was the initial test of a technique and because a large population was available from which to draw respondents, a sample of undergraduate social science students (*N* = 292) was administered the questionnaire. Ages of respondents ranged from 17 to 57 with a mean of 28.7. About 45 percent of the respondents were male, and 79 percent were right-handed. All questionnaires were administered in class groups and there were very few problems with the forms or instructions. In all cases completion took the groups less than 10 minutes. We assume that completing the questionnaire this rapidly increased the randomness of responses.

Responses were assigned integer values, the highest value being assigned to broadly smiling faces, to "SA" (from a format using the abbreviation "SA" for strongly agree and "SD" for strongly disagree), and to "strongly agree" (from a format with the words spelled out). Numeric values were arbitrarily assigned with 1 as the low score and 5 as the high score. Numerical scores were then summed for each set of items and the percent response for each category was calculated as were the summated score means (*M*) and variances (s^2). Obtained means (*M*) were then compared with expected means of the random variables (μ,) using significance of difference texts for independent sample means (*Zm*). The variance ratios (s^2 to σ^2) were also calculated. These ratios were tested for significance using one-tailed *Z* scores (Zs^2) transformation of chi square tests (9) since central tendency and agreement errors are both expected to reduce the observed variance (s^2).

C. RESULTS

Initial examination of the data revealed a strong bias among respondents to mark the first, left-most response on the first line of each cluster of items. Hence, first lines were eliminated from these analyses and treated separately. Data for the remaining nine items in each cluster are presented in Table I.

The faces formats were the most problematic formats of all. The differences between the observed and expected means on these items were greater than for any other type of response format. In all three sets (male, female, and neuter faces) there were significant preferences for the smiling faces, which accounted for about half of all responses. Hair length or gender apparently did not affect the results; all the means (*M*) obtained for faces formats were significantly different from the expected mean (μ). These results indicate the occurrence of leniency errors in faces scales. In contrast, no significant leniency error was found for standard five-point Likert and numeric formats. All obtained means (*M*) of these formats varied only slightly from the expected mean (μ).

Table I
Comparison of Faces Formats to Other Five-Point Response Formats

	Response								
Format	1	2	3	4	5	M- μ	Z_m	s^2/σ^2	Z_s^2
Faces									
NH	11.9	15.9	23.4	25.5	23.4	.329	2.59*	.86	-1.69*
SH	11.8	16.8	23.4	25.8	22.4	.294	2.23*	.86	-1.69*
LH	10.8	15.5	23.4	25.6	24.6	.337	3.01*	.84	-1.93
Likert	16.9	20.3	24.25	22.2	16.3	.179	0.05	.86	-1.63
Numeric	17.5	20.5	24.2	21.9	15.8	-.031	0.27	.87	-1.55

Note: Higher codes designate more positive or "agreeable" responses. A "5" is a broadly smiling face. In numeric formats, higher numbers were arbitrarily considered positive. NH = no hair or neuter; SH = short hair or male; LH = long hair or female. M-μ. = difference between observed and expected means; *Zm* = difference tests for independent sample means· s^2/σ^2 = ratio between observed variance and expected variance; Zs^2 = *Z* score transformation of chi square tests.
*p < .05.

Faces scales also were more likely to produce central tendency errors than were Likert and numerical formats. On each type of format there were substantial and consistent central tendency errors that resulted in average variance ratios (s^2/σ^2) between .84 and .89. All Zs^2s (described above) for the faces scales were significant. One of the four other formats (numeric 5 to 1) was significant; Likert five-point formats were only marginally insignificant (*p* = . 0516).

Respondents' answers on the first item in each set are reproduced in Table 2. These data indicate a strong bias or tendency toward marking the first occurring response, regardless of format, though it is stronger for faces scales, regardless of gender, than for any of the Likert or numerical formats. This tendency persists regardless of where a set of items is located on a questionnaire page; however, it is greater when "strongly agree" or a smiling face appears in the left-most position. Similarly, higher numbers beginning with "SA" received higher proportions of left-most marks than did the reversed items.

Table 2
First Item Responses by Format (in %)

	Order of response				
Format	1st	2nd	3rd	4th	5th
Faces[a]					
Neuter	50.9	18.9	14.7	8.3	7.2
Male	43.4	21.9	18.9	7.5	7.9
Female	49.3	21.9	13.9	6.3	8.7
Five-point					
Likert: SA to SD	41.7	17.7	19.8	10.8	10.1
Likert: SD to SA	33.3	26.0	18.1	13.2	9.4
Numeric: 1 to 5	39.2	16.7	13.2	14.6	16.0
Numeric: 5 to 1	42.1	19.7	16.9	6.6	14.8

Note: SA = strongly agree; SD = strongly disagree.
[a] Smiling to frowning.

D. DISCUSSION

These results demonstrate, first, that faces scales contained greater leniency errors than any of the other response formats tested. Their strong positive bias means that, when employed with measures of job satisfaction and employee perceptions of other aspects of organizational behavior, they are less valid and less reliable than the more common Likert and numerical formats.

Second, respondents showed a moderate tendency to over-select the center-most responses. This resulted in a reduction in variance from what would be expected under normal assumptions of randomness. Faces formats suffered more from this central tendency error than did Likert and numerical formats. The faces scales in Ironson and Smith's data (11) apparently suffer from this same problem.

Third, it is also significant that empirical measurements may be somewhat biased by the tendency of respondents to select first-appearing responses on any page or at the beginning of any set of items. We call this new type of systematic response bias "first-choice error." This tendency is much stronger for faces than for Likert and numerical formats. This result suggests that researchers might be well advised to employ initial filler items as controls and to use Likert or numerical rather than faces format scales. Use of adjusted response anchors (11) is another way this problem might be corrected but this possibility needs to be more fully investigated.

Fourth, the presence of leniency, central tendency, and first choice errors may confound statistical analyses by artificially restricting the range and amount of variance in the data. This may limit the usefulness of the higher order statistical analyses. This problem is greater for measures using faces formats than for other kinds of measurement schemes.

Fifth, what should be done about these measurement errors is a separate, more difficult question. Minimally, data should be examined to determine if first-choice error, leniency, and central tendency errors are present or not. Examination of frequency distributions is a simple procedure and should always be done for every variable regardless of format. Ideally, adjustments to the data should be made to correct this effect. How this can best be done is unknown and needs to be investigated. At this point, however, researchers might reduce these difficulties by using Likert and numeric instead of faces formats in their questionnaires, and take them into account in analyzing and presenting their results.

Finally, since faces scales are significantly more prone to suffer from leniency, central tendency, and first-choice errors and since the statistical impact of these errors is greatest on faces measures, we conclude that faces formats are less appropriate for empirical research than are standard Likert and numeric formats and that their use should be avoided.

REFERENCES

1. BARDO, J. W., YEAGER, S. J., & KLINGSPORN, M. J. Preliminary assessment of format-specific central tendency and leniency error in summated rating scales. *Percep. & Mot. Skills*, 1982, 54, 227-234.
2. BASS, B. M. Development and evaluation of a scale for measuring social acquiescence. *J. Abn. & Soc. Psychol.*, 1956, 51, 296-299.
3. COOPER, M. An exact probability test for use with the Likert-type scales. *Educ. & Psychol. Meas.*, 1976, 36, 647-655.
4. DUNHAM, R. B., & HERMAN, J. B. Development of a female faces scale for measuring job satisfaction. *J. Appl. Psychol.*, 1975, 60, 629-631.
5. DUNHAM, R. B., SMITH, F. J., & BLACKBURN, R. S. Validation of the index of organizational reactions with the JDI, the M.S.Q., and the faces scales. *Acad. Manag. J.*, 1977, 20, 420-432.
6. GUILFORD, J. P. Psychometric Methods. New York: McGraw-Hill, 1954.
7. HALADYNA, T., & THOMAS, G. The attitudes of elementary school children toward school and subject matters. *J. Exper. Educ.*, 1979, 48, 18-23.
8. HAMMER, T. H., & VARDI, Y. Locus of control and career self-management among non-supervisory employees in industrial settings. *J. Vocat. Behav.*, 1981, 18, 13-29.
9. HAYS, W. L. Statistics. New York: Holt, Rinehart, & Winston, 1981.
10. HUSEK, T. R. Acquiescence as a response set and as a personality characteristic. *Educ. & Psychol. Meas.*, 1961, 21, 295-307.
11. IRONSON, G. H., & SMITH, P. C. Anchors away-the stability of meaning of anchors when their location is changed. *Person. Psychol.*, 1981, 34, 249-262.
12. KAFRY, D., & PINES, A. The experience of tedium in life and work. *Hum. Relat.*, 1980, 33, 477-503.
13. KIM, J. O. Multivariate analysis of ordinal variables. *Amer. J. Soc.*, 1975, 81, 261-298.
14. KIRK, R. Introductory Statistics. Monterey, Calif.: Brooks/Cole, 1978.
15. KUNIN, T. The construction of a new type of attitude measure. *Person. Psychol.*, 1955, 8, 65-78.
16. LABOVITZ, S. The assignment of numbers to rank order categories. *Amer. Sociolog. Rev.*, 1970, 35, 514-515.
17. LIKERT, R. A technique for the measurement of attitudes. *Arch. Psychol.*, 1932, 140.
18. NUNALLY, J. Psychometric Theory. New York: McGraw-Hill, 1967.
19. ROBERTS, K. H., & O'REILLY, C. Measuring organizational communication. *J. Appl. Psychol.*, 1974, 59, 321-326.
20. SMITH, P. C., KENDALL, C. L., & HULIN, M. L. The Measurement of Satisfaction in Work and Retirement. Chicago: Rand McNally, 1969.
21. STOCKFORD, L., & BISSELL, H. W. Factors involved in establishing a merit rating scale. *Personnel*, 1949, 26, 94-118.
22. VAN HEERDEN, J., & HOOGSTRATEN, J. Response tendency in a questionnaire without questions. *Appl. Psychol. Meas.*, 1979, 3, 117-121.