DOCUMENT RESUME

ED 409 362 TM 026 895

AUTHOR	Barnette, J. Jackson
TITLE	Effects of Item and Response Set Reversals on Survey Statistics.
PUB DATE	Mar 97
NOTE	21p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Likert Scales; *Reliability; *Responses; *Statistical Analysis; *Surveys; Test Construction; *Test Items
IDENTIFIERS	*Acquiescence; Stem Analysis

ABSTRACT

The controversy regarding reverse or negatively-worded survey stems has been around for several decades. The practice has been used to guard against acquiescent or response set behaviors. A 20-item, 5-point Likert item survey was designed and the stems and response sets were varied in a 2 by 3 design. One independent variable was type of item stem: one level had all direct-worded stems and the other had, randomly determined, half direct and half reverse-worded stems. The other independent variable was response set type. One level had all response sets going "strongly disagree" (SD) to "strongly agree" (SA), one had all response sets going SA to SD, and the third had, randomly determined, half going SD to SA and half going SA to SD. The surveys were administered to 687 subjects. The form each subject received was determined randomly. Responses were scored so that all were in agreement with the direct or positive form of the item stem. Item means were lower for the all direct-worded surveys compared with the half direct, half reverse-worded stems. The survey with the all direct stems and half SD-SA, half SA-SD response sets had the highest item variable. However, the most important finding was that the survey with the lowest reliability was the one with half direct and half-reverse worded stems with half SD-SA and half SA-SD response sets, while the survey with the highest reliability was the survey with all direct-worded stems with half SD-SA and half SA-SD response sets. This would indicate that the use of a combination of all direct-worded stems and half of the response sets going in one direction and half in the other direction may be a better way of guarding against acquiescence and response set behaviors. (Contains eight tables and nine references.) (Author/SLD)

Effects of Item and Response Set Reversals on Survey Statistics

by

J. Jackson Barnette

P. O. Box 870231

University of Alabama

Tuscaloosa, AL 35487-0231

(205) 348 1184

(205) 348 0683 (FAX)

JBARNETT@UA1VM.UA.EDU

A paper presented at the 1997 Annual Meeting
of the
American Educational Research Association
Chicago, IL, March 1997

4/97

ABSTRACT

The controversy regarding using reverse or negatively-worded survey stems has been around for several decades. The practice has been used to guard against acquiescent or response set behaviors. A 20-item, five-point Likert item survey was designed and the stems and response sets were varied in a 2 by 3 design. One independent variable was type of items stem: one level had all direct worded stems and the other had, randomly determined, half direct and half reverse-worded stems. The other independent variable was response set type: one level had all response sets going SD to SA, one had all response sets going SA to SD, and the third had, randomly determined, half going SD to SA and half going SD to SA. The surveys were administered to 687 subjects. The form each subject received was determined randomly. Responses were scored so that all were in agreement with the direct or positive form of the item stem. Item means were lower for the all direct worded surveys compared with the half direct, half reverse-worded stems. The survey with the all direct stems and half SD-SA, half SA-SD response sets had the highest item variance. However, the most important finding was that the survey with the lowest reliability was the one with half direct and half-reverse worded stems with half SD-SA and half SA-SD response sets while the survey with the highest reliability was the survey with all direct-worded stems with half SD-SA and half SA-SD response sets. This would indicate that the use of a combination of all direct-worded stems and half of the response sets going in one direction and half going in the other direction may be a better way of guarding against acquiescence and response set behaviors.

Reverse or negatively-worded stems have be used extensively in educational surveys to guard against acquiescent behaviors or the tendency for respondents to generally agree with survey statements more than disagree. Also, such item stems are used to guard against subjects developing a response set where they pay less attention to the content of the item and provide a response that relates more to their general feelings about the subject than the specific content of the item. Reverse-worded items were used to attempt respondents to attend more to the survey items. Most of the research on this practice has pointed out problems with reliability, factor structures, and other statistics.

While there are ample examples of the use of reverse-worded item stems, no examples were found where response sets were reversed or where various combinations of reverse worded stems and reversed response sets were used. This research seeks to systematically examine effects of stem and item reversals on commonly used survey statistics of internal consistency reliability, survey means and survey variances.

Relevant Literature

The controversy associated with the use of direct and negatively-worded or reverse-worded survey stems has been around for the past several decades. Reverse-wording items has been used to guard against respondents providing acquiescent or response set related responses. Two general types of research has been conducted. One has looked at effects on typical survey statistics, primarily reliability and item response distributions and the other type has looked at factor structure differences.

Chamberlain and Cummings (1984) compared reliabilities for two forms of a course evaluation instrument. They found reliability has higher for the instrument when all positively-worded items were used. Benson (1987) used confirmatory factor analysis of three forms of the same questionnaire, one where all items were positively-worded, one where all were negatively-worded, and one where half were of each type to examine item bias. She found different response patterns for the three instruments which would lead to potential bias in score interpretation.

Barnette (1996) compared distributions of direct-worded and reverse-worded items on surveys completed by several hundred students and another one completed by several hundred teachers. He found that a substantial proportion of respondents in both cases provided significantly different distributions. On the student survey, which had 14 reverse-worded items out of 57, 31.3% of the respondents provided different distributions at p < .05, 17.7% had different distributions at p < .01, and 9.7% had different distributions at $p < .001$. There were lower proportions for the survey taken by the teachers. At the $p < .05$ level, 25.8% of the teachers had different distributions, 10.3% were different at $p < .01$, and 1.6% were different at $p < .001$.

Marsh (1986) examined the ability of elementary students to respond to items with positive and negative orientation. He found that preadolescent students had difficulty discriminating between the directionally oriented items and this ability was correlated with reading level; students with lower reading levels were less able to respond appropriately to negatively-worded item stems.

As pointed out by Benson and Hocevar (1985) and Wright and Masters (1982), the use of mixed items is based on the assumption that respondents will respond to both types as related to the same construct. Pilotte and Gable (1990) examined factor structures of three versions of the same computer anxiety scale: one with all direct-worded or positively-worded stems, one with all negatively-worded stems, and one with mixed stems. They found different factor structures when mixed item stems were used on a unidimensional scale. Others have found similar results. Knight, Chisholm, Marsh, and Godfrey (1988) found the positively-worded items and negatively worded items loaded on different factors, one for each type.

Methods

A 20-item survey designed by the author for assessing attitudes toward year-round schooling was used, modified with different item and response sets. The response set was a five-point scale of Strongly Disagree (SD), Disagree (D), Neutral (N), Agree (A), and Strongly Agree (SA). The original version of this survey had a Cronbach Alpha of .85. There were six versions of this survey as follows:

Form A: Original survey with no negatively worded items with response set of SD on left to SA on right

Form B: Original survey with no negatively worded items with response set of SA on left to SD on right

Form C: Original survey with no negatively worded items with (randomly determined) half SD-SA and half SA-SD

Form D: Half (randomly determined) direct-worded and half reverse-worded with response set of SD on left to SA on right

Form E: Half (randomly determined) direct-worded and half reverse-worded with response set of SA on left to SD on right

Form F: Half (randomly determined) direct-worded and half reverse-worded with response set with (randomly determined) half SD-SA and half SA-SD

This results in a two-by-three factorial design as follows:

Response→ ↓Stem	SD to SA	SA to SD	Half SD to SA Half SA to SD
All Direct-worded	Form A	Form B	Form C
Half Direct-Half Reverse-worded	Form D	Form E	Form F

The dependent variables are survey reliability, survey means and survey variances. Three types of reliability were computed: Cronbach's Alpha, split-half odd-even, and split-half first half-second half. The split-half correlations were compared to determine significant differences using Fisher z tests with transformed coefficients. Respondent survey means were compared using factorial analysis of variance. Respondent survey variances were compared using Bartlett homogeneity of variance tests. At least forty-five subjects per cell were sought to meet sample size requirements for the basic two-by-three design to detect an effect size of 0.75, with $\alpha = 0.05$, and a power of 0.90.

Collection and Scoring of Data

The six different instruments were randomly mixed in sets of ten of each type. They were then administered to classes of undergraduate students, graduate students and inservice teachers in five different locations. No names or any other identifiers were used. All instruments were computer-scored using a program written by the author, and responses were reverse-scored as needed to have the lowest response (one) being indicative of not agreeing with the positive or direct state of the item content and the highest response (five) being indicative of agreeing with the positive or direct state of the item content. The results presented below are on the reverse-scored, as needed, responses.

Results

Table 1 presents the means and standard deviations for the six instrument configurations. The treatment means were all very close to the midpoint of the scale which was 3.0. The lowest mean was 2.957 for the all direct stems-half SD-SA and half SD-SA response sets treatment and the highest was 3.114 for the half direct and have reversed stems-all SD-SA response sets. Mean differences were tested with two-way ANOVA testing for interaction and main effects. Results for the one-way ANOVA are presented in Table 2. The interaction effect was not significant, $F(2, 681) = 1.05$, $p > .05$, nor was the main effect relating to response set order, $F(2, 681) = 0.88$, $p > .05$. However, there was a significant difference in the means on the all direct-half direct and half reversed stem variable, $F(1, 681) = 5.12$, $p < .05$. The mean for the half direct and have reversed stems was 3.107 while the mean for the all direct stems was 3.027.

Table 3 presents the item variances for the six instrument configurations. Variances were tested for the interaction cells and for main effects using a series of Bartlett tests. As reported in Table 4, there was a significant difference for the six interaction cells, $\chi^2(5, N= 687) = 19.08$, $p < .05$. Follow-up was conducted using three Bartlett tests comparing the all direct with the half direct-half reversed stems within each response set variable. As indicated in Table 5, only one contrast was significant and that was within the half SD-SA and half SA-SD response sets, $\chi^2(1, N= 229) = 13.313$, $p < .05$. The variance for the all direct item stems configuration was higher (0.3186) than for the half direct-half reversed item stems configuration (0.1600)

Reliability coefficients were computed for all instrument configurations and are reported in Table 6. Three coefficients were computed: Cronbach's Alpha internal consistency and two split-half coefficients, one correlating total scores on the odd and even numbered items and the other correlating total scores of the first ten items and the second ten items. The split-half coefficients incorporated a Spearman-Brown correction to double the length to twenty items, the number used for the Cronbach Alpha coefficient.

The range of Alpha coefficients was from .6525 in the half direct, half reversed stem condition with half SD-SA, half SA-SD response set condition to .8569 in the all direct stem condition with half SD-SA, half SA-SD response set condition. In every case the Cronbach Alpha was higher for the all direct stem condition than the half direct, half reversed stem condition. All of the Alpha's were above .80 for the all direct stem conditions and lower than .73 for the half direct-half reverse-worded stems. The range of odd-even split-half coefficients was from .7666 to .8906 and the range for first half-second-half split-half coefficients was .6742 to .8310. Of these reliability coefficients, the odd-even split-half had less variability among the six treatment conditions.

Since the split-half coefficients are based on Pearson's product-moment correlation, comparisons are made using z tests of Fisher-transformed coefficients. Table 7 presents results of the z tests for the odd-even correlations. For the total all direct stems compared with the total half direct-half reverse worded stems, there was a significant difference ($z = 2.890$, $p < .05$) with the all direct having a higher odd-even reliability (.8766) than the half direct-half reversed (.8142). There was a significant difference between these two levels of item stem type for the all SD-SA response set ($z = 2.124$, $p < .05$), where the odd-even reliability was higher for the all direct stems (.8605) compared with the half direct-half reversed (.7666). In addition, there was a significant difference between these two levels of item stem type for the half SD-SA and half SA-SD response sets ($z = 2.144$, $p < .05$) where the odd-even reliability was higher for the all direct stems (.8906) compared with the half direct-half reversed (.8135).

Table 8 presents results of the z tests for the first half-second half correlations. For the total all direct stems compared with the total half direct-half reverse worded stems, there was a significant difference ($z = 2.092$, $p < .05$) with the all direct having a reliability (.7898) than the half direct-half reversed (.7214). The only significant difference between these two levels of item stem type within the response set type was for the half SD-SA and half SA-SD response sets ($z = 2.781$, $p < .05$) where the reliability was higher for the all direct stems (.8310) compared with the half direct-half reversed (.6742).

Significance tests comparing Cronbach Alpha coefficients will be conducted after programs have been developed by the author on procedures presented by Feldt, Woodruff, and Salih (1987).

Conclusions

There was evidence that commonly used survey statistics were affected by the various treatment conditions. There were differences in condition means and condition variances. These would affect score interpretability. There were also differences in reliability coefficients. Cronbach Alpha values varied considerably, as did half-half split-half coefficients. There was less variability in the odd-even split-half coefficients. This would likely be a function of difficulty in dealing with the item stem or response set reversals being distributed about evenly for the odd and even-numbered items.

The most important finding in this research relates to the use of mixed response sets rather than mixed item stems. The summary statistics were actually higher for the condition where all direct-worded item stems were used in combination with half of the response sets going from SD to SA and the other half going from SA to SD. This condition had the highest level of reliability and also higher item variance. This would seem to indicate that this condition was reliable and provided for higher discrimination of responses on the scale. Thus, reversing response sets seems to be a much better alternative than reversing item stems to reduce acquiescence or response set behaviors.

References

Barnette, J. J. (1996). Responses that may indicate nonattending behaviors in three self-administered educational attitude surveys. *Research in the Schools, 3*(2).

Benson, J. (1987). Detecting item bias in affective scales. *Educational and Psychological Measurement, 47*(1), 55-67.

Benson, J. & Hocevar, D. (1985). The impact of item phrasing on the validity of attitude scales for elementary school children. *Journal of Educational Measurement, 22*(3), 231-240.

Chamberlain, V. M. & Cummings, M. N. (1984). Development of an instructor/course evaluation instrument. *College Student Journal, 18*(3), 246-250.

Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical inference for coefficient Alpha. *Applied Psychological Measurement, 11*(7), 93-103.

Knight, R. G., Chisholm, B. J., Marsh, N. V., & Godfrey, H. P. (1988). Some normative, reliability, and factor analytic data for the revised UCLA Loneliness Scale. *Journal of Clinical Psychology, 44*(2), 203-206.

Marsh, H. W. (1986). Negative item bias in rating scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology, 22*(1), 37-49.

Pillote, W. J. & Gable, R. K. (1990). The impact of positive and negative item stems on the validity of a computer anxiety scale. *Educational and Psychological Measurement, 50*(3), 603-610.

Wright, B. & Masters, M. (1982). *Rating scale analysis*. Chicago: Mesa Press.

Table 1

Item Means and Standard Deviations by Instrument Configuration

		Left to Right Response Sets (SD - SA)	Right to Left Response Sets (SA - SD)	Half Each Direction Half SD - SA Half SA - SD	Total
All Direct Items	n	116	114	109	339
	M	3.0478	3.0728	2.9569	3.0270
	SD	0.4709	0.4961	0.5645	0.5118
Direct and Reversed Items	n	115	113	120	348
	M	3.1139	3.1027	3.1054	3.1073
	SD	0.4147	0.4325	0.4000	0.4145
Total	n	231	227	229	687
	M	3.0807	3.0877	3.0347	3.0677
	SD	0.4441	0.4647	0.4899	0.4668

Table 2

Analysis of Variance for Item Means by Instrument Configuration

Source	*df*	*MS*	*F*	*p*
Stem reversal	1	1.1083	5.12	0.0239
Response reversal	2	0.1894	0.88	0.4170
Interaction	2	0.2273	1.05	0.3503
Error	681	0.2163		
Total	686			

Table 3

Item Variances by Instrument Configuration

		Left to Right Response Sets (SD - SA)	Right to Left Response Sets (SA - SD)	Half Each Direction Half SD - SA Half SA - SD	Total
All Direct Items	n	116	114	109	339
	Var	0.2217	0.2461	0.3186	0.2620
Direct and Reversed Items	n	115	113	120	348
	Var	0.1720	0.1871	0.1600	0.1718
Total	n	231	227	229	687
	Var	0.1972	0.2160	0.2400	0.2179

Table 4

Bartlett's Test for Comparing Item Variances by Instrument Configuration

Comparison	df	χ^2	p
Stem reversal	1	15.16	.0001
Response reversal	2	2.22	.3301
Interaction	5	19.08	.0019

Table 5

Comparison of Item Reversal Configuration within Response Reversal Condition using Bartlett's Test

Comparison	df	χ^2	p
Direct vs. Half Direct-Half Reversed within all SD-SA	1	1.844	.1745
Direct vs. Half Direct-Half Reversed within all SA-SD	1	2.108	.1466
Direct vs. Half Direct-Half Reversed within Half SD-SA and Half SA-SD	1	13.313	.0003

Table 6

Reliability Coefficients by Instrument Configuration, all Adjusted to 20 Item Length

	Reliability Type	Left to Right Response Sets (SD - SA)	Right to Left Response Sets (SA - SD)	Half Each Direction Half SD - SA Half SA - SD	Total
All Direct Items	n	116	114	109	339
	Alpha	.8126	.8216	.8569	.8335
	Odd-even	.8605	.8716	.8906	.8766
	Half-half	.7704	.7587	.8310	.7898
Direct and Reversed Items	n	115	113	120	348
	Alpha	.7179	.7261	.6525	.6989
	Odd-even	.7666	.8624	.8135	.8142
	Half-Half	.7303	.7588	.6742	.7214
Total	n	231	227	229	687
	Alpha	.7686	.7772	.7837	.7772
	Odd-even	.8161	.8648	.8610	.8499
	Half-Half	.7480	.7589	.7766	.7616

Table 7

Comparison of Odd-Even Reliability Coefficients using Fisher z Tests

Response Set	All Direct	Half Direct Half Reversed	z	p
SD - SA	.8605	.7666	2.124	.0347
SA - SD	.8716	.8624	0.275	.7832
Half SD-SA Half SA-SD	.8906	.8135	2.144	.0331
Total	.8766	.8142	2.890	.0040

Table 8

Comparison of Half-Half Reliability Coefficients using Fisher z Tests

Response Set	All Direct	Half Direct Half Reversed	z	p
SD - SA	.7704	.7303	0.690	.4912
SA - SD	.7587	.7588	-0.002	.9986
Half SD-SA Half SA-SD	.8310	.6742	2.781	.0059
Total	.7898	.7214	2.092	.0368

AERA CONFERENCE 1997

TM026895

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Effects of item and response set reversals on survey statistics	
Author(s): J. Jackson Barnette	
Corporate Source:	Publication Date: March 1997

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following two options and sign at the bottom of the page.

☑ Check here

For Level 1 Release:
Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical) *and* paper copy.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 1

The sample sticker shown below will be affixed to all Level 2 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 2

☐ Check here

For Level 2 Release:
Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical), but *not* in paper copy.

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Sign here→ please

Signature: J. Jackson Barnett	Printed Name/Position/Title: J. Jackson Barnette, Professor	
Organization/Address: University of Alabama, P. O. Box 870231, Tuscaloosa, AL 35487-0231	Telephone: (205) 348 1184	FAX: (205) 348 0683
	E-Mail Address: JBARNETT@UA1VM.UA.EDU	Date: April 16, 1997

(over)