Educational and Psychological Measurement
http://epm.sagepub.com

Effects of Response Order on Likert-Type Scales

Li-Jen Weng and Chung-Ping Cheng
Educational and Psychological Measurement 2000; 60; 908
DOI: 10.1177/00131640021970989

EFFECTS OF RESPONSE ORDER ON LIKERT-TYPE SCALES

LI-JEN WENG AND CHUNG-PING CHENG
National Taiwan University, Taipei, Taiwan

The study investigated whether the change of response order in a Likert-type scale altered participant responses and scale characteristics. Response order is the order in which options of a Likert-type scale are offered. The sample included 490 college students and 368 junior high school students. Scale means with different response orders were compared. Structural equation modeling was used to test the invariance of interitem correlations, covariances, and factor structure across scale formats and educational levels. The results indicated that response order had no substantial influence on participant responses and scale characteristics. Motivating participants and avoiding ambiguous items may minimize possible effects of scale format on participant responses and scale properties.

Likert-type scales have been very popular as a means of measuring human attitudes. Since Likert (1932) introduced the summative method to measure attitudes, this method has had an enduring impact on social science research (Likert, Roslow, & Murphy, 1934, 1993). During the past 60 years, the effects of scale format on participant responses on Likert-type scales, as well as reliability and validity of the scale scores, have been intensively researched. Among the factors studied, the influences of number of response categories and choice of verbal labels attached to the scales have received much attention (e.g., Bendig, 1954; Champney & Marshall, 1939; Chang, 1994; French-Lazovik & Gibson, 1984; Halpin, Halpin, & Arbet, 1994; Hancock & Klockars, 1991; Jenkins & Taber, 1977; Klockars & Hancock, 1993; Klockars & Yamagishi, 1988; Komorita & Graham, 1965; Matell & Jacoby, 1971; Newstead & Arnold, 1989; Spector, 1976; Wildt & Mazis, 1978; Wong, Tam, Fung, & Wan, 1993). The present study investigated the effect of

We thank all the students who participated in this study. Correspondence concerning this article should be addressed to Li-Jen Weng, Department of Psychology, National Taiwan University, Taipei 106, Taiwan, R.O.C.; e-mail: ljweng@ccms.ntu.edu.tw.

Educational and Psychological Measurement, Vol. 60 No. 6, December 2000 908-924

response order of the scale on participant responses and psychometric properties of the scale. This facet of Likert-type scale format has not been studied in such great detail.

Response order refers to the order in which response options of a Likert-type scale are presented. Response order may influence participant responses when participants lack motivation to attend to all the options given. Participants are expected to consider all the options provided and select the most appropriate one. A participant with limited motivation may instead choose the first option that appears acceptable to him or her without examining all the options. Previous research on response order effects has led to inconsistent conclusions (Belson, 1966; Chan, 1991; Johnson, 1981; Mathews, 1927, 1929). (Although Mathews conducted his studies prior to Likert's introduction to the summative method in 1932, the item format he used to measure student responses was in accordance with Likert-type scale.) Certain studies demonstrated that participant responses changed as options of the scales were altered, and others found participant responses robust to change of response order. If response order influences participant responses and the psychometric properties of the scale scores, conclusions from previous Likert-scale research might be called into question and future research involving Likert scales may need to be designed differently. On the other hand, if response order does not affect participant responses, researchers need not consider the order in which alternative options are presented. A comprehensive study is needed to clarify the effects of response order on the popular Likert-type scale.

The form of a Likert-type scale can be classified as positive (or traditional, ordinary) or negative (or reversed), according to the order in which alternative responses are presented (Belson, 1966; Chan, 1991). The positive form presents the positive or the favorable response labels (such as *like greatly*) first, whereas the negative form places the negative labels (such as *dislike greatly*) first. Earlier research investigated whether response order influenced participant choices (Belson, 1966; Chan, 1991; Johnson, 1981; Mathews, 1927, 1929), the interitem correlations, and factor structure of the scales (Chan, 1991), but the findings were inconclusive. Belson (1966), Chan (1991), and Mathews (1927, 1929) found that response order had a statistically significant and moderate effect on participant responses. Chan (1991) found that the interitem correlations obtained from positive and negative forms yielded different factor structures among high school students. Johnson, on the other hand, found very minor influence of response order among highly educated respondents. An examination of previous studies suggested that the discrepancy in conclusions might result from differences in research design.

The designs of previous research on response order differed in two aspects: first, the treatments of response order as a between-subjects variable or a within-subjects variable, and second, the education level of the sample used. *Response order* was treated as a between-subjects variable in Belson (1966), Johnson (1981), and Mathews's (1929) primary school sample. Different groups of participants responded to positive and negative forms. On the other hand, Chan (1991) and Mathews (1929) took response order as a within-subjects variable in their junior high school and college samples. The same participants responded to both positive and negative forms. When a control group that responds to the same forms repeatedly is absent in the design, changes of participant responses observed in repeated-measures design might result from factors other than response order. Hence, response order was treated as both a between-subjects variable and a within-subjects variable in this study, and control groups responding to the same forms across time were also included as the basis for comparison.

The *educational levels* of the participants in past research ranged from primary school to college and beyond. Mathews (1927, 1929) used students at primary school, junior high school, and college. About 90% of Belson's (1966) participants had a high school education or less. Chan (1991) collected data from high school students. Johnson's (1981) participants were mainly male elites who were readers of *Horizons USA* and had occupations as educators, government officials, mass communicators, defense leaders, civic leaders, labor leaders, businessmen, professionals, artists, writers, and students. Although amount of education might account for the presence or absence of response-order effects (e.g., Krosnick & Alwin, 1987; McClendon, 1986), findings from past research identified no consistent influence of education. Johnson's samples of male elites showed no clear evidence for response-order effects, but Chan's (1991) sample of high school students demonstrated response-order effects. All of Mathews's (1927, 1929) participants—including primary school students, junior high school students, and college students—showed response-order effects. Some earlier research (e.g., Belson, 1966; Krosnick & Alwin, 1987) classified level of education into two categories: high school or less and college and beyond. If we categorized participant education in previous studies accordingly, participants of a high school education or less tended to exhibit response-order effects, whereas participants of a college education or beyond seemed robust to such effects, except for the college sample in Mathews's (1929) study.

A systematic study of the influence of education level on response order effect is warranted. Therefore, the present study includes participants at two educational levels, college and junior high school. If level of education is a plausible explanation for the presence of response-order effects, such effects are expected to appear in the junior high school sample but not in the college sample.

But why would level of education explain the presence or absence of response-order effects? Let us consider the theory proposed by Krosnick and

Alwin (1987). Krosnick and Alwin's research on response-order effects with ranking data suggested that people tended to choose a satisfactory or an acceptable answer instead of an optimal one so as to minimize the psychological costs required to respond. When encountering an acceptable option, participants were likely to select that option without examining through all possible options. The phenomenon was more likely to occur when examining all the options required much more cognitive demands on the participants than simply checking the first acceptable option without examining the rest. The response-order effects were anticipated to appear among participants with less cognitive sophistication because choosing the optimal option required more psychological costs for them relative to capacities than did seeking a satisfactory answer. According to their findings, participants with a high school education or less and more limited vocabularies were more likely to be influenced by response order, which further supported their hypothesis.

If cognitive sophistication is relevant to response-order effects on ranking data, it is likely to affect response-order effects with ratings on Likert-type scales as well. According to Krosnick and Alwin (1987, 1988), the amount of formal education is an important indicator of the degree of cognitive sophistication. Hence, the junior high school sample in our study was expected to show a stronger response-order effect than the college sample.

The purpose of the present research was to understand whether response order affected participant responses, and how education level mediated the presence of response-order effects. It was hypothesized that the response-order effects would appear in the junior high school sample but not in the college sample. Earlier research examined whether response order affected response means (Belson, 1966; Chan, 1991; Johnson, 1981; Mathews, 1929), interitem correlations, and factor structure among items (Chan, 1991). These three aspects of participant responses and scale properties were examined in the present study. In short, this study investigated systematically the possible effects of response order and amount of education on participants' responses on Likert-type scales, including response means, interitem correlations and covariances, and factor structure of the items.

Method

Participants

The entire sample consisted of 858 participants with complete data. The college sample consisted of 490 students in a university in Taipei, including 220 males and 270 females. The junior high school sample included 173 boys and 195 girls, a total of 368 students from a junior high school in Taitung, a city located on the east coast of Taiwan.

Instrumentation

Five items from the Personal Distress Scale (Chan, 1986; Chan, 1991), a subscale of the Interpersonal Reactivity Index, were used for the replication of previous findings. Chan (1986) translated the items into Mandarin. The present study used the Mandarin version of the Personal Distress Scale. One example of the items is "When I see someone who badly needs help in an emergency, I go to pieces." Two forms of this scale were constructed. The positive form of the scale presented *describes me very well* (4), the most positive alternative, at left followed by *describes me quite well* (3), *describes me well* (2), *describes me slightly well* (1), and *does not describe me well* (0). The negative form placed the response alternatives in an opposite order beginning with *does not describe me well* (0) at the left. Chan (1986) conducted exploratory factor analyses on the five items together with items from other subscales, and the results indicated that all five items saturated primarily the Personal Distress factor, but the fourth and the fifth items had small structure coefficients on the Fantasy factor as well. The same results were found in both the college sample and the high school sample.

Design and Procedures

All the participants were administered the scale twice, each 1 week apart to avoid possible changes in responses due to experience or maturation. The positive form and the negative forms were randomly distributed to participants in the first administration. In the second administration, some participants received whichever form they had not previously taken, whereas others received the same form as taken before. Response order, therefore, could be treated as both a between-subjects variable and a within-subjects variable. When data from the first administration were analyzed, response order was treated as a between-subjects variable. When data from two administrations were compared, response order was treated as a within-subjects variable. According to the education levels of the participants and the forms received in the two administrations, the whole sample were classified into eight groups, ranging from university sample with negative forms on both administrations (UNN) to junior high school sample with positive forms on both administrations (JPP).

Analyses

Scale means. ANOVA and dependent t tests were used to compare means of the scale. When response order was treated as a between-subjects variable, a form by education two-way ANOVA was performed against data collected at the first administration of the scale. With response order as a within-subjects variable, dependent *t* tests were employed to compare differences in scale means across time. Dependent *t* tests were also conducted on participants responding to the same form in two administrations to examine the stability of scores across time.

Interitem correlations and covariances. Structural equation modeling was used to test the equality of interitem correlations and covariances across forms. When response order was treated as a between-subjects variable, multisample structural equation modeling was performed on data collected at the first administration. The analyses included four samples with the combination of two levels of education and two forms of response order. When response order was treated as a within-subjects variable, structural equation modeling was conducted to test the stability of correlations and covariances across forms. The stability of correlations and covariances with the same form across time was also tested for the sake of completeness. All the analyses were carried out by EQS (Bentler & Wu, 1995). LISREL8 (Jöreskog & Sörbom, 1996) was also used to help calculate some of the fit indexes for model evaluation. Weng and Cheng (1997) showed that relative fit indexes obtained from least squares (LS), generalized least squares (GLS), and maximum likelihood (ML) estimation methods differ due to the differences in parameter estimates of the null model. Weng and Cheng (1996) and Fan, Thompson, and Wang (1999) showed that the values of fit indexes of a model depended on the estimation method used. Hu and Bentler (1999) indicated that ML-based fit indexes outperformed those obtained from GLS and asymptotically distribution-free estimator (ADF) and should be preferred indicators for evaluating model fit. According to these findings, the maximum likelihood estimation method in EQS was employed throughout the study.

The chi-squared statistics (Chi) and various fit indexes were used to evaluate the fit of proposed models to the data. A cutoff of .95 for the nonnormed fit index (NNFI), the comparative fit index (CFI), the incremental fit index (IFI), and the relative noncentrality index (RNI); a cutoff of .06 for root mean squared error of approximation (RMSEA); and a cutoff of .09 (or .10) for the standardized root mean squared residual (SRMR), as suggested by Hu and Bentler (1999), were adopted for model evaluation. Other frequently used fit indexes such as goodness-of-fit index (GFI), the ratio of chi-squares to degrees of freedom (Chi/*df*), and normed fit index (NFI) were also presented for model assessment. Because the sample size of each group in this study was fewer than 250, the Satorra-Bentler scaling corrected (SCALED) test statistic was also used when applicable, as Hu and Bentler suggested.

Factor structure. Structural equation modeling was again used to test the equality of factor structure with positive and negative forms. To explore the factor structure of the Personal Distress Scale in junior high school and college samples, exploratory factor analysis on the five items was conducted prior to the application of confirmatory factor analysis. Two methods of analysis were used: the maximum likelihood estimation method with oblimin rotation and the iterative principal factor method with promax rotation. The analysis was performed on data collected from each administration of the scale in the eight groups, respectively. A total of 32 (8 groups by 2 administrations by 2 methods) analyses were completed. An examination of the results suggested that a two-factor model outperformed the one-factor model in explaining the interitem correlations. The first three items were measures of the first factor and the remaining two items were measures of the other factor.

These results were different from Chan's (1986). Although Chan found that the five items saturated one factor when items from other subscales were included in the analyses, the last two items were found to be correlated with another factor too. This result implied that the construct underlying the first three items might not be identical to the construct underlying the last two items. The present analyses with only the five items clearly revealed this underlying factor structure. Therefore, the two-factor structure model, instead of the one-factor model, was employed to test the invariance of factor structure against response order. Response order was again treated as both a between-subjects variable and a within-subjects variable the same way as in test of invariance of interitem correlations and covariances across forms. The maximum likelihood estimation method in EQS (Bentler & Wu, 1995) and LISREL8 (Jöreskog & Sörbom, 1996) was used.

Results

Scale Means

The means and standard deviations of the scale scores obtained from each administration of the scale were summarized in Table 1. Scores from the second administration of the scale were lower than those from the first administration except junior high school negative-negative (JNN) and junior high school positive-negative (JPN) groups. The correlations of scores from two administrations of the scale ranged from .70 to .77 except in the junior high school negative-positive (JNP) group. Four independent *t* tests were conducted on data collected at Time 1 to test the hypothesis that, within each educational level, participants taking the same form in the first administration but different forms at the second administration responded to the scale similarly. The results supported the hypothesis.

Response order as a between-subjects variable. Table 2 presents the results from the form by education two-way ANOVA, treating response order

Table 1
Mean Scale Scores, Paired t *Test, and Correlation Across Administration of Scale for Each Group*

			M		*SD*			
Education	Fm1-Fm2	*n*	Adm 1	Adm 2	Adm 1	Adm 2	*t*	*r*
University								
	Neg-Neg	93	8.581	7.237	3.579	3.481	4.76*	.703
	Neg-Pos	184	8.092	7.364	3.724	3.567	3.52*	.705
	Pos-Neg	104	8.308	7.673	3.912	4.307	2.23	.754
	Pos-Pos	109	7.835	6.661	3.755	3.963	4.33*	.732
Junior high school								
	Neg-Neg	78	8.256	8.692	3.849	4.418	–1.20	.707
	Neg-Pos	105	7.848	6.952	3.647	3.696	2.76*	.592
	Pos-Neg	83	7.301	7.313	3.571	3.910	–0.04	.734
	Pos-Pos	102	7.480	6.588	3.332	3.649	3.72*	.762

Note. Fm1-Fm2 = forms received at the first and the second administration of the scale; Adm1 = data collected at the first administration of the scale; Adm2 = data collected at the second administration of the scale; Neg = negative form; Pos = positive form.
*$p < .01$.

as a between-subjects variable. The analysis was conducted on data collected at the first administration of the scale. The results indicated that neither the interaction effect nor the main effects of order and education was significant. Because a two-factor model represented the scale items better than the one-factor model, additional ANOVAs were performed on the sum of the first three items (Factor I) and sum of the last two items (Factor II) separately. Results of the two ANOVAs indicated that only education had a main effect on mean scores of Factor I ($p < .01$), but its associated correlation ratio was small ($\eta^2 = .012$). The overall results suggested that when different participants responded to positive and negative forms of the scale, response order did not influence participant responses.

Response order as a within-subjects variable. Results of the dependent *t* tests as shown in Table 1 indicated that five out of the eight groups showed a statistically significant difference ($p < .01$) in scores obtained from two administrations of the scale. Three out of the five differences were found in groups responding to the same form in two administrations. Further analyses showed that inequality of scores mainly came from differences in Factor I. Scores on Factor II showed no statistically significant differences across two administrations in all eight groups. With the mixed results obtained, it seems difficult to conclude that response order had any systematic effects on mean scores. The properties of items may play a role in mediating the presence or absence of response-order effects.

Table 2
Analysis of Variance for Total Scale and Factor Means Collected at First Administration

	Total		Factor I		Factor II	
Source	*F*	η^2	*F*	η^2	*F*	η^2
Form	2.547	.003	1.514	.002	2.800	.003
Education	3.127	.004	10.684*	.012	0.869	.001
Form × Education	0.718	.001	2.677	.003	0.302	.000

Note. All *F*s were tested with 1 and 854 degrees of freedom.
*$p < .01$.

Interitem Correlations

Response order as a between-subjects variable. Model 1 (M1) in Table 3 presents the results of testing the equality of correlation matrix among items across four groups (two education levels combined with two forms) collected at Time 1. Although the chi-square value was statistically significant due to large sample size, fit indexes suggested an acceptable fit of the model to the data. Response order did not affect interitem correlations when it was treated as a between-subjects variable.

Response order as a within-subjects variable. Table 4 presents the results of testing the model of equal correlation matrix across two administrations of the scale. The results indicated that interitem correlations remained the same across two administrations of the scales, regardless of the form used in seven out of the eight groups analyzed. The JPN group showed only a marginal fit of the model to the data. Response order taken as a within-subjects variable did not result in substantial changes in correlations among items.

Interitem Covariances

Response order as a between-subjects variable. Model 2 (M2) in Table 3 presents the results of testing the equality of covariances among items, when response order was taken as a between-subject variable. The chi-square statistic and various fit indexes indicated that the model of identical covariance matrix across four samples was rejected by the data. Two additional analyses were conducted. The model of common covariance matrix for two university samples and two junior high school samples (M3) was supported. Another model in which groups receiving the same form at Time 1 had a common covariance matrix (M4) was rejected. The results suggested that the difference in covariance matrix among groups was due to the difference in

Table 3

*Test Statistics and Fit Indexes for Test of Four-Sample Models on Data Collected at First Administration of Scale (*N *= 858)*

Model	*df*	Chi	*p*1	Chi/*df*	GFI	SRMR	RMSEA	*p*2	NFI	NNFI	CFI	IFI	RNI
M1. Invariant correlation matrix across groups	30	60.65	.001	2.022	.963	.143	.035	.980	.953	.967	.975	.983	.975
M2. Invariant covariance matrix across groups	45	148.53	.000	3.301	.944	.113	.052	.352	.885	.926	.917	.913	.917
M3. Common covariance matrix with same education level	30	23.70	.785	.790	.986	.061	.000	1.000	.982	1.007	1.000	1.013	1.005
M4. Common covariance matrix with same form	30	139.62	.000	4.654	.944	.109	.065	.010	.892	.883	.912	.920	.912
M5. Invariant factor loadings across groups	25	34.64	.095	1.386	.967	.071	.021	.999	.973	.988	.992	1.004	.992
M6. Invariant factor loadings and covariance across groups	34	41.77	.169	1.229	.963	.105	.016	1.000	.968	.993	.994	.999	.994
M7. Invariant factor loadings, covariance, and common error variance with same education level	44	51.80	.196	1.177	.957	.111	.014	1.000	.960	.994	.994	.991	.994
M8. Invariant factor loadings and covariance, and common error variance with same form	44	160.58	.000	3.650	.923	.119	.056	.147	.875	.915	.907	.903	.907
M9. Invariant factor loadings, covariance, and error variances	49	163.03	.000	3.327	.928	.116	.052	.327	.873	.925	.909	.901	.909

Note. Chi = chi-square test statistic; *p*1 = *p* value associated with chi-square statistic; Chi/*df* = ratio of chi-square statistic to degrees of freedom; GFI = goodness-of-fit index; SRMR = standardized root mean squared residual; RMSEA = root mean squared error of approximation; *p*2 = *p* value for test of RMSEA < .05; NFI = normed fit index; NNFI = nonnormed fit index; CFI = comparative fit index; IFI = incremental fit index; RNI = relative noncentrality index; M = Model.

Table 4
Test Statistics and Fit Indexes for Test of Invariant Correlation Matrix Across Administration for Each Group

Education	Fm1-Fm2	Chi	$p1$	Chi/df	Chi2	$p2$	GFI	SRMR	RMSEA	$p3$	NFI	NNFI	CFI	IFI	RNI
University															
	Neg-Neg	19.22	.038	1.922	15.86	.104	.961	.056	.100	.108	.960	.905	.979	1.059	.979
	Neg-Pos	9.55	.481	0.955	8.50	.580	.990	.043	.000	.782	.991	1.002	1.000	1.034	1.000
	Pos-Neg	20.08	.029	2.008	19.89	.030	.963	.067	.099	.097	.971	.931	.985	1.038	.985
	Pos-Pos	17.80	.059	1.780	14.89	.136	.969	.066	.085	.169	.972	.941	.987	1.046	.987
Junior high school															
	Neg-Neg	19.00	.040	1.900	17.45	.065	.955	.097	.108	.099	.945	.865	.970	1.087	.970
	Neg-Pos	7.87	.642	0.787	6.12	.805	.985	.045	.000	.801	.979	1.030	1.000	1.115	1.007
	Pos-Neg	32.03	.000	3.203	24.85	.006	.932	.082	.163	.003	.926	.744	.943	1.033	.943
	Pos-Pos	16.91	.076	1.691	13.44	.200	.968	.052	.083	.195	.954	.904	.979	1.087	.979

Note. *df* for all is 10. Fm1-Fm2 = forms received at the first and the second administration of the scale; Chi = chi-square test statistic; $p1 = p$ value associated with chi-square statistic; Chi/*df* = ratio of chi-square statistic to degrees of freedom; Chi2 = Satorra-Bentler scaled test statistic; $p2 = p$ value associated with Satorra-Bentler scaled test statistic; GFI = goodness-of-fit index; SRMR = standardized root mean squared residual; RMSEA = root mean squared error of approximation; $p3 = p$ value for test of RMSEA < .05; NFI = normed fit index; NNFI = nonnormed fit index; CFI = comparative fit index; IFI = incremental fit index; RNI = relative noncentrality index; Neg = negative form; Pos = positive form.

educational levels rather than form of the scale were used. Response order did not affect interitem covariances.

Response order as a within-subjects variable. Table 5 presents the results of testing the model of equal covariance matrix across two administrations of the scale. The chi-square statistics and various fit indexes suggested that the model that interitem covariances remained the same across two administrations was acceptable in most groups. The JPN group had the poorest fit. The UNN group showed a marginal fit of the model. Response order did not demonstrate systematic influences on covariances among items.

Factor Structure

Response order as a between-subjects variable. Multisample confirmatory factor analysis was used to test the invariance of factor structure when response order was treated as a between-subjects variable. The bottom five rows of Table 3 present the fit of various models (M5 to M9) to the data from the first administration. The results indicate that a model of invariant factor pattern coefficients and interfactor covariance across four groups and invariant error variances within each education level (M7) best fits the data. The results explain why the covariance matrices from the university sample and the junior high school sample differ (M3). The difference results from unequal error variances. Neither response order nor level of education had any effects on major parameters of the factor model, including factor pattern coefficients and interfactor covariances.

Response order as a within-subjects variable. Table 6 presents the fit of the model of invariant factor pattern coefficients and interfactor covariance across two administrations of the scale. The results suggest that the model shows an adequate fit to the data for all the samples. Response order and level of education do not show any substantial effects on the factor structure of the scale.

Discussion

The present study systematically investigated the effect of response order on participant responses on a 5-point Likert-type scale. Previous research on the effects of response order on Likert-type scales employed different designs, and the conclusions were inconsistent. In the present study, response order was treated as both a between-subjects variable and a within-subjects variable. It was shown that response order did not affect scale means, interitem correlations, interitem covariances, factor pattern coefficients, and interfactor covariance of the scale, when taken as a between-subjects vari-

Table 5
Test Statistics and Fit Indexes for Test of Invariant Covariance Matrix Across Administration for Each Group

Education	Fm1-Fm2	Chi	$p1$	Chi/df	Chi2	$p2$	GFI	SRMR	RMSEA	$p3$	NFI	NNFI	CFI	IFI	RNI
University															
	Neg-Neg	35.30	.002	2.353	30.90	.009	.934	.042	.120	.016	.927	.861	.954	.957	.954
	Neg-Pos	35.79	.002	2.386	34.02	.003	.964	.039	.087	.048	.967	.939	.980	.980	.980
	Pos-Neg	24.79	.053	1.653	27.01	.029	.956	.062	.080	.180	.965	.955	.985	.986	.985
	Pos-Pos	23.24	.079	1.549	21.38	.125	.961	.055	.071	.245	.963	.958	.986	.987	.986
Junior high school															
	Neg-Neg	24.72	.054	1.648	25.89	.039	.944	.110	.092	.143	.928	.903	.968	.971	.968
	Neg-Pos	16.45	.353	1.097	14.22	.508	.971	.053	.030	.603	.955	.987	.996	.996	.996
	Pos-Neg	40.32	.000	2.688	35.57	.002	.920	.096	.140	.004	.907	.804	.935	.939	.935
	Pos-Pos	20.92	.139	1.395	18.11	.257	.962	.061	.062	.337	.943	.945	.982	.983	.982

Note. df for all is 15. Fm1-Fm2 = forms received at the first and the second administration of the scale; Chi = chi-square test statistic; $p1 = p$ value associated with chi-square statistic; Chi/df = ratio of chi-square statistic to degrees of freedom; Chi2 = Satorra-Bentler scaled test statistic; $p2 = p$ value associated with Satorra-Bentler scaled test statistic; GFI = goodness-of-fit index; SRMR = standardized root mean squared residual; RMSEA = root mean squared error of approximation; $p3 = p$ value for test of RMSEA < .05; NFI = normed fit index; NNFI = nonnormed fit index; CFI = comparative fit index; IFI = incremental fit index; RNI = relative noncentrality index; Neg = negative form; Pos = positive form.

Table 6
Test Statistics and Fit Indexes for Test of Invariant Factor Loadings and Factor Covariance Across Administration for Each Group

Education	Fm1-Fm2	Chi	*p*1	Chi/*df*	Chi2	*p*2	GFI	SRMR	RMSEA	*p*3	NFI	NNFI	CFI	IFI	RNI
University															
	Neg-Neg	32.70	.336	1.090	29.09	.513	.936	.041	.031	.649	.932	.991	.994	.994	.994
	Neg-Pos	34.75	.252	1.158	33.92	.284	.964	.028	.029	.794	.968	.993	.995	.995	.995
	Pos-Neg	50.54	.011	1.685	53.10	.006	.909	.085	.082	.097	.928	.953	.969	.969	.969
	Pos-Pos	35.26	.233	1.175	29.25	.504	.941	.065	.040	.597	.944	.987	.991	.991	.991
Junior high school															
	Neg-Neg	44.41	.044	1.480	40.37	.098	.903	.123	.079	.168	.871	.928	.952	.954	.952
	Neg-Pos	37.30	.169	1.243	30.28	.451	.940	.070	.048	.488	.898	.966	.977	.978	.977
	Pos-Neg	47.47	.022	1.582	39.26	.120	.904	.085	.084	.114	.890	.932	.955	.957	.955
	Pos-Pos	46.10	.030	1.537	35.11	.239	.921	.071	.073	.179	.875	.925	.950	.952	.950

Note. *df* for all is 30. Fm1-Fm2 = forms received at the first and the second administration of the scale; Chi = chi-square test statistic; *p*1 = *p* value associated with chi-square statistic; Chi/*df* = ratio of chi-square statistic to degrees of freedom; Chi2 = Satorra-Bentler scaled test statistic; *p*2 = *p* value associated with Satorra-Bentler scaled test statistic; GFI = goodness-of-fit index; SRMR = standardized root mean squared residual; RMSEA = root mean squared error of approximation; *p*3 = *p* value for test of RMSEA < .05; NFI = normed fit index; NNFI = nonnormed fit index; CFI = comparative fit index; IFI = incremental fit index; RNI = relative noncentrality index; Neg = negative form; Pos = positive form.

able. When response order was treated as a within-subjects variable, positive and negative forms resulted in different scale means. The difference was mainly due to a shift of scores on Factor I, and the mean of Factor II was unaffected by response order.

Response order again did not show any substantial influence on interitem correlations and covariances, factor pattern coefficients, and interfactor covariance, when taken as a within-subjects variable. In essence, the present study suggests that response order is not a critical factor in affecting participant responses on Likert-type scales and factor structure of the scale. The hypothesis that the junior high school sample was more likely to exhibit response-order effects is not supported by our study.

Why is response-order effect absent in our junior high school sample? To answer this question, we probably need to consider a more basic question: How and when do response-order effects occur? Mathews (1929) indicates that reading habits can be the reason for response-order effects. Johnson (1981) suggests that response-order effects may occur with ambiguous questions or unstructured situations. Literature on response-order effects with Likert-type scales offers only limited discussion on this topic. However, appropriate identification of the circumstances under which response-order effects may occur is of a great value to researchers. Researchers with such knowledge can avoid conditions that lead to unstable participant responses.

Krosnick and Alwin's (1987) research on response-order effects with ranking data provides a helpful line of thought addressing this issue. Krosnick and Alwin (1987) indicate that participant motivation is an important factor for the presence or absence of response-order effects. They also indicate that participants with less formal education are more likely to be affected by change of response order of the scale. In sum, earlier research suggests that the characteristics of *both* the participants and the items are relevant to the presence or absence of response-order effects.

Characteristics of participants include motivation, reading habits, and education level, and characteristics of items include clarity of items and degree of specificity of situations. An examination of the items used in this study suggests that the items of Factor II are less ambiguous and participants are more likely to respond to them consistently regardless of the scale format. The quality of items is a possible explanation for our finding that when response order is treated as a within-subject variable, Factor II is less affected by change of format than Factor I. On the other hand, motivation of the participants may explain the absence of response-order effects in our junior high school sample. The junior high school students in our study came from a city located on the east coast of Taiwan. Unlike students in Taipei, they seldom participated in any research and showed great interest in our study. They were highly motivated and they responded to the questionnaires with full attention throughout the study. Their motivation made the effects of response order less likely to occur.

Our results of no obvious response-order effects seem to suggest that participant motivation and clarity of items are perhaps more critical than education level for the presence or absence of response-order effects. If the participants are not motivated to respond to the scales with attention, the results may be unreliable. Good item-writing skill to avoid ambiguous items is also important in preventing response-order effects. If researchers can do a good job in motivating the participants and in writing clear and unambiguous questions, any change in response order should not be a crucial factor in affecting participants' responses on Likert-type scales and properties of the scales. Low motivation and ambiguous items tend to lead to unstable results and should be avoided.

References

Belson, W. A. (1966). The effects of reversing the presentation order of verbal rating scale. *Journal of Advertising Research*, *6*, 30-37.

Bendig, A. W. (1954). Reliability and the number of rating scale. *Journal of Applied Psychology*, *23*, 323-331.

Bentler, P. M., & Wu, E.J.C. (1995). *EQS for windows users' guide*. Encino, CA: Multivariate Software, Inc.

Champney, H., & Marshall, H. (1939). Optimal refinement of the rating scale. *Journal of Applied Psychology*, *23*, 323-331.

Chan, C. (1986). *Grade, sex role, relationship orientation, and empathy*. Unpublished master's thesis, National Chengchi University, Taipei, Taiwan.

Chan, J. C. (1991). Response-order effects in Likert-type scales. *Educational and Psychological Measurement*, *51*, 531-540.

Chang, L. (1994). A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement*, *18*, 205-215.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, *6*, 56-83.

French-Lazovik, G., & Gibson, C. L. (1984). Effects of verbally labeled anchor points on the distributional parameters of rating measures. *Applied Psychological Measurement*, *8*, 49-57.

Halpin, G., Halpin, G., & Arbet, S. (1994). Effects of number and type of response choices on internal consistency reliability. *Perceptual and Motor Skills, 79*, 928-930.

Hancock, G. R., & Klockars, A. J. (1991). The effect of scale manipulations on validity: Targeting frequency rating scales for anticipated performance levels. *Applied Ergonomics*, *22*, 147-154.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Jenkins, G. D., Jr., & Taber, T. D. (1977). A Monte Carlo study of factors affecting three indices of composite scale reliability. *Journal of Applied Psychology*, *62*, 392-398.

Johnson, J. D. (1981). Effects of the order of presentation of evaluative dimensions for bipolar scales in four societies. *Journal of Social Psychology*, *113*, 21-27.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL8: User's reference guide*. Chicago: Scientific Software International, Inc.

Klockars, A. J., & Hancock, G. R. (1993). Manipulations of evaluative rating scale to increase validity. *Psychological Report*, *73*, 1059-1066.

Klockars, A. J., & Yamagishi, M. (1988). The influence of labels and positions in rating scales. *Journal of Educational Measurement*, *25*, 85-96.

Komorita, S. S., & Graham, W. K. (1965). Number of scale points and the reliability of scales. *Educational and Psychological Measurement*, *15*, 987-995.

Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, *51*, 201-219.

Krosnick, J. A., & Alwin, D. F. (1988). A test of the form-resistant correlation hypothesis: Ratings, rankings, and the measurement of values. *Public Opinion Quarterly*, *52*, 526-538.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22,* 5-55.

Likert, R., Roslow, S., & Murphy, G. (1934). A simplified and reliable method of scoring the Thurstone attitude scales. *Journal of Social Psychology*, *5*, 228-238.

Likert, R., Roslow, S., & Murphy, G. (1993). A simplified and reliable method of scoring the Thurstone attitude scales. *Personnel Psychology*, *46*, 689-690.

Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert-scale items? Study I: Reliability and validity. *Educational and Psychological Measurement*, *31*, 657-674.

Mathews, C. O. (1927). The effect of position of children's answers to questions in two-response types of tests. *Journal of Educational Psychology*, *18*, 445-457.

Mathews, C. O. (1929). The effect of the printed response words on an interest questionnaire. *Journal of Educational Psychology*, *30*, 128-134.

McClendon, M. J. (1986). Response-order effects for dichotomous questions. *Social Science Quarterly*, *67*, 205-211.

Newstead, S. E., & Arnold, J. (1989). The effect of response format on ratings of teaching. *Educational and Psychological Measurement*, *49*, 33-43.

Spector, P. E. (1976). Choosing response categories for summated rating scales. *Journal of Applied Psychology*, *61*, 374-375.

Weng, L.-J., & Cheng, C.-P. (1996). On incremental fit indexes and estimation methods in structural equation modeling. *Survey Research*, *2*, 89-109.

Weng, L.-J., & Cheng, C.-P. (1997). Why might relative fit indices differ between estimators? *Structural Equation Modeling*, *4*, 121-128.

Wildt, A. R., & Mazis, M. B. (1978). Determinants of scale response: Label versus position. *Journal of Marketing Research*, *15*, 261-267.

Wong, C.-S., Tam, K.-C., Fung, M.-Y., & Wan, K. (1993). Differences between odd and even number of response scale: Some empirical evidence. *Chinese Journal of Psychology*, *35*, 75-86.