

ASSESSING THE QUALITY OF SCHOOL LIFE:
SOME TECHNICAL CONSIDERATIONS

Sid Bourke & Jennifer Frampton
University of Newcastle

Abstract

The Quality of School Life of students in Years 7 to 12 at two secondary schools was investigated using a variety of response formats, including four, five and six categories, and reversed scales.

The factor structure, reliability coefficients and scale means were

compared for the different formats, and then examined by sex, year

level and ability of the students. Although there were some differences in factor structure and scale means, these were relatively

consistent across formats. Scale reliabilities, however, differed with

formats having an even number of categories exhibiting higher reliabilities than the 5-category format, including an "Undecided"

category. Other differences related to format are also described and discussed.

Paper presented at the AARE Annual Conference,
Deakin University, November 1992

Introduction

Items assessing attitudes, with responses coded in the Likert tradition (Likert, 1932), have been employed in research and

practical applications in the social sciences for many years. The items are normally summed subsequently to form scales. However, the most desirable format for categorising responses is still subject to dispute. In maximising reliability of the scale developed, whether to use a neutral category (Matell & Jacoby, 1972; Andrich, 1978 and Andrich & Masters, 1988) and the overall number of categories offered to respondents (Champney & Marshall, 1939; Komorita, 1963; Komorita & Graham, 1965; Peabody, 1962; Matell & Jacoby, 1972) are two of the most common concerns. The effect of format changes on the factor structure and stability of instruments has also been investigated (Schutz & Rucker, 1975; Martin, Fruchter & Mathis, 1974; Flamer, 1983). If there is any consensus on these concerns about format, it is that both instrument and respondent characteristics should be considered before a decision is made on the format to be used for items.

If usage by researchers were the only guide, the question of whether to offer a neutral category would seem to have been decided in favour of an "undecided" or "uncertain" category. The argument is that forcing a decision from respondents, who were genuinely undecided about an item, by not offering a neutral category would not improve scale reliability (Bendig, 1953; Finn, 1972; McKelvie, 1978). However, the opposing argument is that respondents who have a slight tendency to agree or disagree with an item might use a neutral category to avoid thinking about that item. Use of a neutral category becomes a response set for some, invalidating their results (Cronbach, 1946; Andrich, 1978; Andrich & Masters, 1988). The efficacy of offering a neutral category is linked with the number of categories offered. Matell & Jacoby (1972) found that an increase in the number of categories resulted in a decrease in the use of the neutral category. The neutral category had a 20 per cent usage for items with three or five categories, and a seven per cent usage for formats with more categories (up to a total of 18 categories).

Increasing the number of response categories offers the possibility of increasing the discrimination of the scale by providing a wider variation in descriptors. The age and sophistication of respondents are relevant here, with younger and less sophisticated respondents perhaps being comfortable with fewer categories, and hence answering such items more consistently. Masters (1974) reported conflicting results for reliability, dependent upon the degree to which opinion was divided on the topic under consideration. When opinion was widely divided the number of categories offered had little effect but, when there was less divergence among respondents, more categories were linked with higher reliability. The descriptors given to categories are also likely to be

important, quite independently from the number of categories offered. For example, offering a category "tend to agree" or "mostly agree" rather than "agree" may attract tentative respondents to use that category.

The order of categories offered for responses may also be important. An early study by Mathews (1929) indicated that some response categories were marked more often when they were placed in a different position among the responses. However, the study did not compare reliabilities of normal and reversed scales. The effect of order was investigated in the present study by reversing the four response categories in one format such that the "disagree" categories were placed on the left and the "agree" categories were placed on the right, that is the "disagree" categories came first.

The Study and Methodology

This study focuses on an instrument measuring secondary students' perceptions of the quality of their school life which has been used in studies of school organization (Ainley, Reed & Miller, 1986), staying on at high school (Ainley, Batten & Miller, 1984), intentions for further education (Bourke & Smith, 1989), and progress through secondary school (Ainley & Sheret, 1992). In this paper the effects on the factor structure and scale reliability of varying the number of response categories offered, of offering a neutral category, and of reversing the order of categories are reported for the "Quality of School Life" (QSL) questionnaire.

The 40-item QSL instrument, designed for secondary students, consists of two general scales (General Satisfaction and Negative Affect) and five specific scales (Relationships with Teachers, Opportunity, Achievement, Status, and Identity). Items have four response categories: "Definitely Agree, Agree, Disagree, Definitely Disagree", that is, there is no neutral category. Reliabilities of the scales have been found to be satisfactory, although not always as high as one would wish. For example, the study by Ainley et al (1986, p.143) found reliabilities ranging from 0.84 for the Opportunity scale to 0.76 for the Negative Affect scale.

In the methodological study reported here, a total of 1518 students in Years 7 to 12 at two high schools in the Hunter Region of NSW completed the QSL questionnaire. This represented about 85 per cent of the total numbers of students at the two schools. Although one school was much larger than the other, the schools were adjacent and shared students from the outer suburbs of Newcastle. In both cases the students had a range of socioeconomic backgrounds. There were approximately equal numbers of male and female students at

both schools.

The four formats of response categories for the QSL questionnaire used are shown in Table 1, together with the numbers of students who responded to each format at each school. In addition to these formats, it was possible to develop other formats artificially by collapsing the four categories to two, the six categories to four and to two categories, and the five categories to three, the latter retaining the neutral category. However, it must be recognised that the artificial collapsing of categories would be unlikely to produce the same results as using fewer categories on the questionnaire itself.

TABLE 1. NUMBER OF RESPONDENTS BY FORMAT BY SCHOOL 1

CATEGORY FORMAT 2 \ SCHOOL	Schl.A	Schl.B	Total
4-point (Strongly Agree to Strongly Disagree)	286	134	420
6-point (Very Strongly Agree to Very Strongly Disagree)	245	168	413
5-point (St.Agree to St.Disagree including Undecided)	241	161	402
Reverse 4-point (Strongly Disagree to Strongly Agree)	151	132	283
Total	923	595	1518

1. Adapted from Frampton (1992).
2. It will be noted that the category labels differed slightly from those used by Ainley et al (1986, p.161).

First the results of structural analyses of the scales, including both factor and reliability analyses were considered in developing the "best" independent measure for each scale, involving the reduction of the number of items in most scales.

Subsequently analyses were undertaken using all items in developing the scales for the calculation of reliability and the comparison of levels of each aspect of the quality of school life by sex, year level and ability.

Results of Structural Analyses

The stability of the factor structure of the QSL instrument over the different formats was first investigated as a measure of construct validity (Williams & Batten, 1981, pp.22-25). Principal components factor analysis (SPSS, 1988, pp.480-499) was used in an attempt to confirm the factor structure established by Ainley et al (1986, pp.139-140) following an oblimin rotation. The analyses determined which items met the criteria for all formats, which met criteria for most formats, and which generally failed to meet the criteria. As the tendency noted by Ainley and Bourke (1992) for items from the two general scales to load also on some of the specific scales was again observed, the procedure they recommended was followed. That is, items belonging to the general scales were analysed separately from those for the specific scales.

TABLE 2. INTENDED AND CROSS-LOADINGS FOR ITEMS EXHIBITING SOME INSTABILITY (An asterisk indicates the intended factor) 1

ITEM BY FORMAT	SCALE				
	Teachr	Opport.	Achieve.	Status	Identity
4-point					
Item 15				.42*	.44
Item 17				.52*	.54
5-point					
Item 7		.45	.40*		
Item 35			.59	.32*	
6-point					
Item 25	.48	.37*			
Item 26		.49	.40*		
Item 35			.47	.43*	
Reverse 4					
Item 15				.42*	.48
Item 17				.44*	.61

1. Adapted from Frampton (1992).

For the general scales (General Satisfaction and Negative Affect), all 10 items then loaded as expected. Of the 30 items belonging to the five specific scales, most items loaded as expected. That is, they had a loading of at least .30 (and normally much higher than this) on the intended factor and less than .30 (and normally less than .15) on any other factor. However, a total of six items cross-loaded at greater than .30 on a different factor in one of the formats in addition to their loading on the intended factor. The intended and aberrant loadings for these six items are shown in Table 2. As can be seen in the Table, although these items

did load on an inappropriate scale, they also loaded on the expected scale at an acceptable level. It was also clear that, of the five specific scales, the Status scale had the most items which cross-loaded onto other scales.

There were only small differences between formats when the eigenvalues and the total percentages of the variance explained by the factors were examined (see Table 3). The variance explained by the five factors was slightly higher for the 6-point format than for any of the other formats, while the 5-point format had marginally the lowest proportion of explained variance. The two 4-point formats explained almost identical proportions of the variance. For a detailed listing of factor loadings for each item under each format see Frampton (1992, pp.50-55).

When the scale associated with each of the specific factors was examined, the Teacher scale was always the first factor, Opportunity and Identity were either the second or third factor in each case, Achievement was consistently the fourth factor, and Status was invariably the fifth factor. Status was also the weakest factor as indicated by the number of items which also loaded on other factors (see also Table 2). Despite the minor differences in item loadings between factors reported above, there was a high level of stability between formats. Finally five items were removed because of cross-loadings within some of the formats investigated: three items were removed from the Status scale, and one item from each of the Achievement and Opportunity scales. For details of actual items removed see Frampton (1992, p.43).

TABLE 3. COMPARISON OF EIGENVALUES AND VARIANCE EXPLAINED FOR THE DIFFERENT FORMATS OF THE SPECIFIC SCALES

FACTOR	EIGENVALUES (EV) X FORMAT			
	4-point	5-point	6-point	Reverse 4-pt
1	7.34	7.44	8.24	7.95
2	2.51	2.60	2.51	2.42
3	1.87	1.62	1.69	1.64
4	1.67	1.34	1.98	1.36
5	1.07	1.14	1.15	1.05
Sum of EV	14.46	14.16	15.74	14.42
% Variance	57.8	56.5	59.9	57.6

Reliability of the QSL scales was of primary importance in determining their usefulness in research. Reliabilities for the four formats for each of the seven scales, of which three scales had been reduced, are shown in Table 4. The 4-point format had the highest reliability for three of the scales, the 6-point format had the highest for three scales, the Reverse 4-point for 2 scales, and the 5-point format had the equal highest reliability for one scale. Because of the wider variation of reliability of the 4-point format for the different scales, the 6-point format had the highest mean reliability, both 4-point formats had the next highest and the 5-point format had the lowest mean reliability. However the differences, although reasonably consistent, were not large. The reliabilities were, in the main, very similar to those recorded for the Victorian study.

Any collapsing of the response categories from 6-point and 4-point to 2-point, and from 5-point to 3-point consistently reduced scale reliabilities, the 2-point format having the lowest reliabilities overall.

In summary, reliabilities for scales with an even number of categories were consistently higher than for the 5-point scale. On average, scales with a higher number of categories were more reliable, although differences were small. Finally when the ordering of categories was considered, the scales where the disagreement categories preceded the agreement categories (the reverse of the usual order) had slightly higher reliability.

TABLE 4. RELIABILITIES OF REDUCED QSL SCALES BY FORMAT

SCALE	FORMAT				No. of Items
	4-pt	5-pt	6-pt	Rev4-pt	
General Scales					
Gen.Satisfaction	0.819	0.834	0.835	0.821	5
Negative Affect	0.772	0.736	0.807	0.812	5
Specific Scales					
Teacher	0.860	0.824	0.845	0.828	6
Opportunity	0.844	0.821	0.842	0.874	5 *
Achievement	0.822	0.774	0.813	0.803	5 *
Status	0.686	0.713	0.713	0.687	3 *
Identity	0.779	0.806	0.832	0.821	6

Mean Reliability 0.797 0.787 0.812 0.807

* These scales had been reduced from six items.

Reliabilities With All Items Included

In order to make comparisons with the characteristics of the QSL scales found by Ainley et al (1986) for a large sample of 8464 secondary students in Victoria, the reliabilities for the scales containing all items were calculated and are reported in Table 5. The reliabilities were generally consistent with those reported in the study by Ainley et al (1986), although the Achievement scale was consistently higher in the present study, and the Status scale was higher for all formats except the 5-point.

TABLE 5. RELIABILITIES OF FULL SCALES BY FORMAT

SCALE 1	FORMAT				AINLEY
	4-pt	5-pt	6-pt	Rev4-pt	(4-pt)
General Scales					
Gen.Satisfaction	0.817	0.819	0.838	0.821	0.83
Negative Affect	0.772	0.705	0.809	0.812	0.76
Specific Scales					
Teacher	0.859	0.803	0.846	0.828	0.83
Opportunity	0.840	0.817	0.833	0.863	0.84
Achievement	0.838	0.803	0.830	0.826	0.77
Status	0.794	0.776	0.810	0.821	0.77
Identity	0.775	0.796	0.831	0.821	0.78
Mean Reliability	0.814	0.788	0.828	0.827	0.80

1. These were extracted from Ainley et al (1986, p.143).

As might be expected because of the three additional items included, reliabilities for the Status scale were considerably higher than for the reduced Status scale. Differences for other scales tended to be minor. Overall, the 6-point and the reverse 4-point formats had the highest mean reliabilities (approximately 0.83), with the 4-point format a little less (about 0.81), and the 5-point format had the lowest mean

reliability (0.79).

The full scales were used in further analyses which were undertaken of the mean levels of responses for the total group and then by sex, year level and ability. The full scales were used in preference to the reduced scales for two reasons. First, the improved reliability coefficients when all items were included meant that error would be lessened when relationships between the scales and other variables were analysed. Secondly, the five items were removed initially because, although they loaded on the factor intended, they also cross-loaded on another factor, thus reducing the independence of the factors involved. If the deleted items had not loaded on the intended factor, perhaps the reduced scales would have been preferred.

Comparison of Scale Means

It would be of some concern to researchers and others using the QSL instrument if the number of categories offered grossly affected any interpretation of level of responses given. Of course the numerical descriptors given to responses would be higher for formats with more categories, but it is the interpretation of the descriptors which is important. This could be checked by comparing the means calculated for the different formats in terms of the category headings used ("Very strongly agree" to "Very strongly disagree"). In order to make more direct comparisons of levels of response across formats, it was considered desirable to re-scale some of the numerical equivalents allocated to response categories. Given that the original questionnaire had four categories, this was the standard used. Consequently the 5-point format items were scaled down such that category 5 (Strongly agree) became 4, category 3 (Undecided) became 2.5, and other categories were scaled accordingly. For the 6-point format items, category 6 (Very strongly agree) became 4 and other categories were scaled proportionally.

The adjusted mean scores for each scale under each format are shown in Table 6. On the whole, differences were small between formats, although when tested by an analysis of variance the difference reached statistical significance at the 0.05 level for three scales. For the Negative Affect scale, the mean for the 6-point format was significantly higher than the other formats; the mean for the Status scale for the 6-point format was significantly higher than for the two 4-point formats; and the mean for the Identity scale for the 5-point format was significantly higher than the mean for the reverse 4-point format.

TABLE 6. COMPARISON OF SCALE MEANS BY FORMAT, ALL SCALES
COMBINED AND WITH THE VICTORIAN STUDY

SCALE MEANS	FORMAT				AINLEY	
	4-pt	5-pt	6-pt	Rev4-pt	All	(4-pt)
1						
General Scales						
Gen.Satisfaction	2.53	2.58	2.64	2.55	2.58	2.74
Negative Affect *	1.93	1.94	2.07	1.96	1.96	1.94
Specific Scales						
Teacher	2.78	2.83	2.79	2.71	2.78	2.93
Opportunity	3.13	3.18	3.13	3.01	3.15	3.12
Achievement	2.95	2.98	2.89	2.92	2.92	3.08
Status *	2.58	2.51	2.60	2.54	2.51	2.54
Identity *	3.04	3.09	3.03	2.97	3.04	3.13
Mean (Excluding Negative Affect)	2.82	2.86	2.85	2.77	2.83	2.92

1. Adapted from Ainley et al (1986, p.143).

* Significant difference between formats.

When the present results were compared with those for the earlier Victorian sample, the Victorian responses in general had the higher means. This was true for four scales: the General Satisfaction, Teacher, Achievement and Identity scales. The Negative Affect, Opportunity and Status scales had about the same means. Before becoming concerned about what were minor differences in mean scores on the QSL scales between this study and the study by Ainley et al (1986), it should be remembered that the data were gathered in two different locations and about seven years apart. There was also a difference in the response category headings used. The Ainley study used "Definitely agree" and "Mostly agree", while the present study used "Strongly agree" and "Agree" for the 4-point scale format. This change, made in this study to better accommodate naming the response categories for the 6-point format, may well have slightly altered the level of agreement of respondents.

Scale Mean Differences by Sex. When scale means were compared for male and female students, there was generally little

difference across formats (see Frampton, 1992, p.82). In only two cases was there a significant difference between means: on the Negative Affect scale responses by females were more favourable for the 5-point format, and on the Status scale the mean score for females was higher for the 6-point format. In addition, female students generally gave more favourable, although not significantly different responses for the General Satisfaction, Teacher and Identity scales.

Scale Mean Differences by Year Level. There were several cases where significant differences were found between formats for students at different levels of the secondary school: junior (Years 7 and 8), mid level (Years 9 and 10), and senior (Years 11 and 12). At the junior level, the highest mean scores were found for the 5-point format, and these were significant for the Teacher, Opportunity, Achievement and Identity scales. There was only one significant difference at the mid-secondary level: the 6-point format had the highest means for the Status scale. Again there was only one significant difference at the senior level: the 6-point format had the highest means for the General Satisfaction scale (Frampton, 1992, p.84). However, students at the junior level generally gave more favourable responses than the other students for the General Satisfaction, Negative Affect, Teacher, Opportunity and Achievement scales.

Scale Mean Differences by Ability. Although no intentional measure of student ability was used, in both schools the mathematics classes for Years 8 to 10 were graded. This information was used to develop three groups: higher, medium and lower ability students. There were no significant differences in the mean level of responses for any of these groups by format (Frampton, 1992, p.86). However, the lower ability group generally gave less-favourable responses than the other two groups.

Summary and Concluding Comments

On the basis of the present small study and the literature available, it would seem that two intersecting trends are apparent. First, scales with more categories have higher reliability than scales with fewer categories (Green & Rao, 1970; Finn, 1972; Lissitz & Green, 1975; and Wylie, 1976), particularly for heterogeneous items (Komorita & Graham, 1965). Secondly, scales with an even number of categories are more reliable than scales with an odd number of categories (Andrich, 1978; Andrich & Masters, 1988).

There is also the suggestion that reverse scales may be slightly more reliable. Whether a reverse scale, simply

because it is different, causes a respondent to pause and think more carefully about their answer, is a possibility that requires further investigation.

When scale means are compared across formats, those with more categories tend to be higher, as found by Komorita & Graham (1965) when there were few items in a scale, although differences were small in the present study. The reverse format produced the lowest mean scores, suggesting that there may have been a tendency for respondents to tick more frequently the first boxes they came to (that is, those on the left). Most of the differences were not statistically significant.

For the development of attitude questionnaires in general, the most vexed question remains whether to use a neutral or undecided category for each item. For this particular questionnaire relating to the quality of school life of secondary students, the evidence suggests that an "undecided" category should be omitted if the reliability of the scales measured is to be maximised. However, there are always a few respondents who are concerned about the absence of an "undecided" category in the questionnaire. This reaction could be simply because the respondents are accustomed to being offered such a category, or because they are disinclined to think sufficiently to decide about the issue, or because they are genuinely unable to decide. However, the need, expressed by a few, for a neutral category should be addressed, albeit in the knowledge that use of an undecided category is to be avoided, if possible.

Of course there is always the option of placing in the instructions the statement that respondent should: "Try to make up your mind, but, if you really cannot decide, omit the item". If made explicit, this is the simplest and perhaps the best course of action. Experience suggests that this type of instruction results in respondents answering almost all the items yet, in permitting the possibility of genuine uncertainty, it satisfies the concerned person. Other possibilities include the provision of a neutral category, but not one centrally located in the response format. It is possible to isolate the neutral response from the other categories by placing it after a gap between it and the "normal" agree - disagree categories. Again such a procedure will usually reduce use of the undecided category. Another option is to use combinations of categories such as "agree" and "mostly agree" rather than "definitely agree" and "agree".

Offering the "mostly agree" and "mostly disagree" categories may reduce concern at being required to make a decision.

It has been assumed that the major requirements of an attitude questionnaire is for the scales developed to assess different aspects of the attitude to be based on a sound theory, to measure what they purport to measure (to have content validity), to be relatively independent of other scales (construct validity), and to measure consistently (to be reliable). This paper has focussed on construct validity of the questionnaire (through factor analysis) and reliability of the scales. It has not addressed the theoretical bases of the instrument (see Williams & Batten, 1981), or content validity.

Investigation of content validity had previously involved not only inspection of the items but qualitative methods such as interviews with students who had completed the questionnaire to check their understandings (see Williams & Batten, 1981, pp.28-29; Ainley & Bourke, 1992, p.110).

Given the vastly increased use of attitude questionnaires in educational research, there is a need to replicate this type of study on several dimensions. Apart from the effects of reversing the order of categories referred to above, changes in the strengths of categories could be investigated further. For example, in a 4-point scale with changes in categories from "strongly agree" to "agree" and from "agree" to "mostly agree", it would be of interest to know the extent to which such a change affected the mean level of agreement for items. Finally with respect to the quality of school life questionnaire specifically, the format recommendations made here relate to secondary students in Years 7 to 12. Whether declining the use of a neutral category and having either four or six response categories would hold for primary students in Years 5 and 6 should be investigated. With respect to the latter, it might be anticipated that the younger primary students would be more comfortable with four rather than six categories per item.

It should be remembered that this study investigated the effects of changing the format of a multi-structural attitude questionnaire containing many heterogeneous items. The suggestion is that a format with either four or six categories provided the highest reliability. Other attitude measures consist of more homogeneous items. It would be necessary to work with a different questionnaire to determine the format effects on an instrument consisting solely of homogeneous items loading onto a single scale. As suggested by Komorita and Graham (1965), it may be that results of different item formats would be quite different for such questionnaires.

References

- Ainley, J., Batten, R. & Miller, H. (1984). Patterns of Retention in Australian Government Schools. Hawthorn, Victoria: ACER.
- Ainley, J. & Bourke, S. (1992). Student views of primary schooling. *Research Papers in Education*, 7(2), 107-128.
- Ainley, J., Reed, R. & Miller, H. (1986). School Organization and the Quality of Schooling: A Study of Victorian Government Secondary Schools. Hawthorn, Victoria: ACER.
- Ainley, J. & Sheret, M. (1992). Progress Through High School. Hawthorn, Victoria: ACER.
- Andrich, D. (1978). A Binomial Latent Trait Model for the Study of Likert-Style Attitude Questionnaires. *British Journal of Mathematical and Statistical Psychology*, 31(1), 84-98.
- Andrich, D. & Masters, G.N. (1988). Rating Scale Analysis. In J.P. Keeves (ed), *Educational Research, Methodology and Measurement: An International Handbook*. Oxford: Pergamon Press.
- Bourke, S. & Smith, M. (1989). Quality of School Life and Intentions for Further Education: The case of a rural high school. Paper presented at the Annual Conference of the AARE, Adelaide.
- Champney, H. & Marshall, H. (1939). Optimal Refinement of the Rating Scale. *Journal of Applied Psychology*, 23, 323-331.
- Cronbach, L.J. (1946). Response Sets and Test Validity. *Educational and Psychological Measurement*, 6, 475-494.
- Finn, R.H. (1972). Effects of Some Variations in Rating Scale Characteristics on the Means and Reliabilities of Ratings. *Journal of Educational and Psychological Measurement*, 32, 255-265.
- Frampton, J. (1992). The effect of changes in the response format of an attitude questionnaire: A methodological study. Unpublished dissertation, University of Newcastle.
- Green, P.E. & Rao, V.R. (1970). Rating Scales and Information Recovery - How Many Scales and Response Categories to Use? *Journal of Marketing*, 34, 33-39.

Komorita, S.S. (1963). Attitude, Content, Intensity and the Neutral Point on a Likert Scale. *Journal of Social Psychology*, 61, 327-334.

Komorita, S.S. & Graham, W.K. (1965). Number of Scale Points and the Reliability of Scales. *Journal of Educational and Psychological Measurement*, 25, 987-995.

Likert, R. (1932). A Technique for the Measurement of Attitude. *Archives of Psychology*, No.140.

Lissitz, R.W. & Green, S.B. (1975). Effect of the Number of Scale Points on Reliability: A Monte Carlo Approach. *Journal of Applied Psychology*, 60(1), 10-13.

Masters, J.R. (1974). The Relationship Between the Number of Response Categories and Reliability of Likert-type Questionnaires. *Journal of Educational Measurement*, 11(1), 49-53.

Matell, M.S. & Jacoby, J. (1971). Is There an Optimal Number of Alternatives for Likert Scale Items? Study 1: Reliability and Validity. *Journal of Educational and Psychological Measurement*, 31, 657-674.

Peabody, D. (1962). Two Components in Bi-polar Scales: Direction and Extremeness. *Psychological Review*, 69(2), 65-73.

SPSS Inc. (1988). *SPSSX User's Guide (3rd Edition)*. Chicago: SPSS Inc.

Wylie, P.B. (1976). Effects of Coarse Grouping and Skewed Marginal Distributions on the Pearson Product-Moment Correlation Coefficient. *Educational and Psychological Measurement*, 36, 1-7.