RELIABILITY AND VALIDITY OF Q-METHOD RESULTS: SOME EMPIRICAL EVIDENCE

John R. Fairweather
University of Missouri-Columbia

It is commonly assumed, and for good reasons, that Q method is reliable. Reliability deals with the problem of whether a measuring instrument consistently measures whatever it is measuring. For Q method this most commonly relates to the issue of whether a given set of Q-sort statements produces results that are reliable over a period of time, either for the same sample of people or for another sample from the population. Proponents of Q methodology readily accept that Q method is reliable because of their understanding of the processes involved in the generation and interpretation of factors. Opponents, however, believe that there are confounding processes which render the factors, and their interpretations, idiosyncratic and unreliable (Cronbach & Gleser, 1954).

This issue of reliability is important in its own right because consistency is a desirable feature of any measure. Knowledge of the reliability of Q method is also important to obtain in order to provide a basis for validity. These issues cannot be resolved arbitrarily, nor in an a priori manner, regardless of the apparent demonstrability of reliability and validity. It is highly desirable that empirical evidence

Operant Subjectivity, 1981(Oct), 5(1), 2-16.

be brought to bear on these issues, although no empirical test can make a conclusive judgment. In the long term it is necessary to accumulate evidence, preferably from a variety of sources (Gronlund, 1971).

Little evidence on these issues is available. However, some studies are based on the assumption that Q method is reliable. For example, Brenner (1972) studied attitudes toward the Vietnam War over a one-year period. In 1967 a Q study yielded four types, described as a single Hawk position and three Dove positions. A year later a similar study yielded three types: The same Hawk and Dove positions were found although these were more sharply characterized, reflecting a polarization of public opinion; one of the Dove positions was not found in the second study. It was assumed that any changes in the results could be completely explained by changes in public opinion and that there were no sources of error derived from lack of reliability.

A study directly focusing on test-retest reliability considered the replicability of individual Q sorts (Kahle & Lee, 1974). Subjects repeated Q sorts at one and two week intervals. It was found that subjects sorted statements the same way, with correlation coefficients between respective Q sorts higher than 0.95. This similarity between Q sorts was taken to indicate that type descriptions developed from the data would be the same, and that Q method was reliable. In addition, Kerlinger (1973: 596n) mentions one person who resorted a 90-item Q sample after eleven months with 0.81 reliability. It is not yet known whether people with unchanged attitudes will produce the same Q sorts after a long time interval when they have had more time to forget their original performances. Neff and Cohen (1967) do not pursue this aspect of reliability in their treatment of item selection by the researcher.

The research reported here focuses on the reliability of Q method over a one- to two-year period, and on the external and construct validity of the results. Two different samples from the same person population sorted a similar set of statements after a one- to two-year interval, and a small sample re-

sorted similar statements also after a one- to two-year interval. It was expected that the factors and type descriptions of both studies would show strong similarities, indicating consistent results. It was found that the results were consistent and that some evidence was available to support a case for external and construct validity.

METHOD

To obtain a comparison of Q data, the results of two recent studies, executed by different researchers and based on different samples of the population, were compared. Lewis (1978), in an initial study of private forest landowners in the Missouri Eastern Ozarks, found three types of landowners. Fairweather (1979), in a follow-up study after a one- to two-year interval, also found three types of landowners. Both studies sought to describe the beliefs, values, and attitudes of landowners to their resources and relate these to management strategy. In both cases Q method was used, with minor changes being made in the materials in the second study. These changes included a modification of the 60 Q-sort statements whereby 17 ambiguous statements that originally caused confusion for the subjects were modified to insure clarity. Another change was the number of piles in the distribution of sorted statements and the number of statements in the extreme piles. These latter changes were made in the attempt to decrease the correlation coefficients between factors; no significant changes were found.

Changes in statements are potentially important as a cause of change in results; however, all statements for both studies fitted into a matrix of four management objectives and five management motivations. This insured that modified statements were the same in content although their outward appearance may have been different. The modified statements better represented the categories of objectives and motivations so that a clearer understanding of landowners could be obtained.

In both studies the sampling of forest landowners

from a 14 county area in Missouri followed the Q-methodological principle of searching for diversity of attitude, value, and belief. In addition to selecting different people in the second study, a group of seven subjects from the original study was included in a separate Q sort.

This design provides two ways of assessing reliability and yields data that can be used to assess two types of validity. Reliability can be examined by comparing the results of the Lewis and Fairweather studies. This is a case of reliability over a long period of time, the use of a nearly identical instrument for a different sample of the population, and interpretation of factor arrays by different researchers. Because of these sources of error, and also because of possible attitude change, a high level of reliability is unlikely to be found. The reliability being assessed here most closely approaches the method of test-retest with equivalent forms. This method is generally regarded as the most useful estimate of test reliability (Gronlund, 1971).

It should be noted that discussions of reliability and validity refer to the results of an instrument and not to the instrument itself. Because results are produced for a given set of conditions the statements about reliability and validity must be made including a description of the nature of those conditions. It follows that the results of an instrument are not reliable in general but only for specific conditions.

The comparison of the Lewis and Fairweather results is performed in two ways. First, a raw-score analysis of statements categorized by management objective and management motivation can be performed and the scores compared. Raw scores, based on the number of columns in the distribution of statements and derived from the position of a statement in a factor array, can be assigned to each statement in the 4 X 5 matrix. The scores of groups of statements under either objective or motivation headings can be added to find which objectives and motivations are most important. This method avoids any error that may be derived from the interpretations of the factor analysis data. A se-

cond means of comparison can be obtained by comparing the actual interpretations of the factor arrays.

Reliability can also be assessed by comparing the results obtained from the seven restudy subjects. This method eliminates error derived from using a different sample, as above. It also provides precise information on the extent of attitude change that has occurred in the seven subjects over the one- to two-year period, and limited information about attitude change in the population.

The Q-sort data of the restudy subjects has to be included in the factor analysis of the Fairweather study because a factor analysis of the results from only seven subjects is unworkable. The seven subjects were selected because they loaded highly on their respective factors in the Lewis study. They should be the least ambiguous and also least likely to change their beliefs, values, and attitudes. Therefore, these subjects should be most likely to load highly on the factors of the Fairweather study and not be so different as to form new factors. If this is found to be the case then attitude change is slight and the results are reliable.

Validity can be examined in two ways, each yielding information on a different type of validity. External validity, or the concern whether the results are generalizable to the population, can be assessed by consideration of the results of the two independent samples. If similar results are found, and reliability obtains, then there is some evidence to indicate that the two samples are in fact representative of the population of private forest landowners. Similarity of types of landowners would indicate that the nonrandom, snowball sampling technique characteristic of Q method is capable of discovering the variety of beliefs, values, and attitudes that exists in the population.

A different type of validity, namely construct validity, can also be examined. This issue concerns the inference of certain psychological or sociological traits or qualities which are said to exist in order to account for differences in behavior. If external validity is found, this is necessary (although not

sufficient) evidence for construct validity. The constructs themselves, in this case the type descriptions of private forest landowners, must be examined and hypotheses developed for independent testing. Demographic data are examined to see if they fit the pattern of constructs, although this "testing" procedure is outside the realm of Q methodology.

With respect to the relationship between reliability and validity, it should be noted that reliability is a necessary but not sufficient condition for validity. Hence, the evidence for reliability is examined first.

RESULTS AND DISCUSSION

Reliability--Different Subjects

Table 1 shows the raw-score sums for each set of three statements within an Objective X Motivation combination. These scores are derived from the position of each statement in the factor array and are based on a scoring schedule of +5 to -5. In the first column are the results of the Fairweather study, in the second column the results of the Lewis study.

A comparison of the two sets of results shows that they are very similar, the management objectives showing an almost identical distribution. Three out of the five motivation subtotals show the same type giving the highest score.

The rank ordering of objectives can also be compared. As shown in the column subtotals of Table 1, the Lewis type 1 gives priority to the preservation objective, followed by wildlife-recreation and timber production with approximately equal scores, and gives a negative score to grazing. This is the same order of objectives as the Concerned Ecologist (Fairweather type 1), assuming that the scores for wildlife-recreation and timber production are approximately equal. The Lewis type 2 gives priority to grazing, preservation, and wildlife-recreation and a lower score to timber production. This differs a little from the Dedicated Farmer (Fairweather type 2) who ranks timber production highest followed by recreation-wildlife

TABLE 1
Comparison of Raw Scores by Type
(* indicates highest score)

MOTIVATIONS[b]	type	OBJECTIVES[a] Timber Production		Wildlife/Recreation		Grazing		Preservation		Subtotals	
FR	1	8	7	5	5	-4	-1	9*	7*	18	18*
	2	13*	6	3	4	5*	-1	1	4	22*	13
	3	11	10*	6*	4	-4	0*	2	1	15	15
I	1	8	6	3	3*	-5	-3	5	2	11	8
	2	9	2	8*	2	5*	5*	3	2	25*	11
	3	12*	7*	-2	2	-1	0	6*	3*	15	12*
A	1	3	1	9*	6*	-2	-1	13*	17*	23*	23*
	2	6	3	4	5	5*	6*	2	3	17	17
	3	9*	3	5	4	-4	-1	9	6	19	12
R	1	1	-1	3	2	1	-1	1*	0	6	0
	2	5	1	3	3*	7*	4*	-3	6*	12*	14*
	3	5	1	2	2	0	-1	-2	-1	5	1
SR	1	1	-1	-1	-1	-5	-3	9*	4*	4	-1
	2	-1	1	2	1	-3*	1*	4	2	2	5
	3	3*	3*	3*	1	-6	-1	7	2	7*	5
Subtotal	1	21	12	19	15	-15	-9	37*	30*		
	2	32	13	20*	15	19*	15*	7	17		
	3	40*	24*	14	13	-15	-3	22	11		

[a]Col. 1, Fairweather study; Col. 2, Lewis study

[b]FR=Financial Return, I=Investment, A=Aesthetics, R=Residence, SR=Social Responsibility

and grazing with approximately equal scores.

Lastly, the Lewis type 3 gives a negative score to grazing and the highest score to timber production, as does the Practical Forester (Fairweather type 3). Preservation and wildlife-recreation have similar middling scores, their order reversed when compared with the Practical Forester.

This comparison of types shows that they are similar, although some minor differences exist, especially for type 2. Some differences are to be expected because in each study different statements were used, rendering impossible a perfect match of results. This raw-score comparison can be supplemented by an examination of interpretations.

The interpretations of the three factors found in both studies show similar patterns of emphases and overall orientation of landowners to their woodland. For each type of landowner, Lewis provided a summary of essential characteristics which convey the important attributes of the types. For the Lewis type 1, the Timber Aesthetist, these are: emphasis on the beauty of forest land, the compatibility of timber harvesting, interest in wildlife and faith in landowners, strong opposition to forest grazing, and strong concern for long-term forest preservation. These attributes are similar to the Fairweather type 1, the Concerned Ecologist, who is concerned about the preservation and conservation of forest land and has a long time perspective regarding its future. He particularly enjoys the non-economic benefits and the aesthetic return of owning land. To him the beauty, the pleasure of walking through the woods, and the joy of seeing wildlife are key facets of his ownership. To this end he is likely to encourage wildlife by growing food plots and discouraging hunters. He harvests timber although this is not his primary objective. His concern for broader issues like the future of forest land and the conservation of wildlife illustrate an altruistic moral perspective perhaps derived from his more financially secure position, not requiring immediate return on his investment. His urban experience may also foster interest in conservation and general moral issues. Frequently this type plants food plots and trees to encourage the wildlife population.

In summary comparison: Except for the fourth statement of the Lewis type 1, which shows "strong opposition to forest grazing," both type 1s are very similar. Like the Concerned Ecologist, the Timber Aestheticist is most concerned with the beauty of

woodland, of which timber harvesting is a part but not the most important aspect of his ownership. As well as harvesting, both types are interested in wildlife and preserving forest land. The Concerned Ecologist is less adamant about forest grazing compared to the Timber Aestheticist, showing a more neutral reaction to this issue. This is the only point of difference.

The Lewis type 2, the Range-Wildlife type, has the following characteristics: Emphasizes grazing operations, has strong interest in wildlife, regards timber harvesting as of secondary importance, finds forest grazing acceptable, and has only slight concern for long-term forest preservation. These characteristics are similar to the Fairweather type 2, the Dedicated Farmer. This type is characterized by an ever-present business attitude combined with a powerful interest in cattle farming. He has strong feelings about the need to produce and manage his resources accordingly. Awareness of the fact that excessive conversion jeopardizes the future of woodland is demonstrated; however, he could be tempted to convert some woodland to pasture if it were profitable. He enjoys wildlife and appreciates it in his woods for hunting enjoyment rather than aesthetic appreciation.

Both types can be seen to be similar because of the emphasis given to grazing and wildlife. Timber harvesting for both types is important, but not as important as grazing, the distinction being strongly drawn for the Range-Wildlife type who gives secondary interest to timber production. This is not precisely the orientation of the Dedicated Farmer who does see timber as an important crop. For both these grazing-oriented types, forest grazing is an important part of their farm operations, and the feasibility and practicality of this strategy are emphasized. Both types are also concerned about the future of woodland, giving some consideration to the issue of preservation.

Finally, the Lewis type 3, the Timber-Wildlife type, has the following characteristics: Emphasizes timber production and forest productivity, has a

fairly high interest in wildlife aesthetics, is dissatisfied with forest grazing, has interest in long-term timber investments, and has faith in long-term forest preservation. Again, this type is similar to the Fairweather type 3, the Practical Forester, who is essentially concerned with the primary goal of continued production and harvesting of timber. He harvests mature timber and carefully manages his woodlot to ensure future supplies, and because of this orientation he is concerned about the future of woodland. Grazing cattle is incompatible with his forestry operation. He has a modest interest in wildlife because he associates its welfare with the welfare of forest land. The enjoyment this person experiences from owning a wood resource derives from an attraction to the forest, the security of growing a valuable product, and the challenge of managing his resource properly.

The type 3 person in both studies is opposed to forest grazing because this is perceived as incompatible with the primary goal of timber production. This interest is almost single-minded, but in both cases is mollified by an interest in wildlife and aesthetic enjoyment.

This comparison of interpretations for the three factors, derived from the two different studies, shows basic similarities of attitude to forest land, despite using different samples of people, slightly different statements, and being compared over a long time interval. There are only slight variations in emphasis.

Restudy Reliability--Same Subjects

This aspect of the method required seven people from the original study to sort the statements of the latter study, and to include their Q-sort data in the factor analysis. The addition of seven people changed the factor solution producing an extra, residual factor. However, the first three factors of the factor solution appeared to have similar factor arrays as those produced when 66 people were used. Before examining the factor loadings of the retest subjects,

TABLE 2
Type Classification and Factor Loadings
for Restudy Subjects

LEWIS 1978, N=66			FAIRWEATHER 1979, N=63			
subject number	type	loading	type	loading	type	loading
40	3	0.629	3	0.528	2	0.519
68	3	0.554	1	0.675	3	0.405
69	1	0.553	1	0.462		
70	2	0.680	2	0.713	3	0.424
71	1	0.679	1	0.666		
72	1	0.831	2	0.556	1	0.423
73	2	0.713	2	0.556		

this apparent similarity has to be demonstrated.

To demonstrate that the factor arrays of the first three factors are the same as the Concerned Ecologist, Dedicated Farmer and Practical Forester, a Spearman's rank-order correlation coefficient was calculated to compare the order of statements between the factor arrays thought to be the same. In all three cases the correlation coefficient was greater than 0.90 with two coefficients equal to 0.99. This indicates that a detailed interpretation would yield the same description because the arrangement of statements is similar. It can safely be concluded that the first three factors of the retest solution are the same as those originally found.

The factors on which each of the restudy respondents loaded, as well as the factor loadings for each individual, are shown in Table 2. These results show that generally the restudy subjects loaded on the same factors although some multiple loadings occurred between the factors in the restudy. All restudy subjects, except numbers 68 and 72, loaded most highly on the same factor they were on in the Lewis study. The two cases which did not had significant secondary loadings on the factor they were on in the Lewis study.

These restudy results are evidence that the factors and their corresponding type descriptions, for

both studies, are similar. They are not identical because the loadings of some of the restudy respondents were not so high on the same factor and in some cases multiple loadings between factors occurred. A perfect match would be impossible because of possible changes in attitude, changes in statements, and the similarity between types. However, despite these influences a close affinity between factors has appeared.

The different-subject reliability method used to compare the types found in the two studies, separated by at least a year, produced results showing that the essential characteristics of the Lewis types have been found again in the Concerned Ecologist, Dedicated Farmer and Practical Forester of the Fairweather study. This indicates that the results are consistent over time and are therefore reliable. Further, the restudy results show consistency when the same subjects are studied over a one- to two-year interval indicating that attitude change is minimal. These indications of reliability provide a basis for assessing the validity of results produced by Q method.

External Validity

The reliability found for the two major studies using different subjects provides a basis for determining external validity. Lack of representativeness is a problem often seen in the small sample size of all Q studies. Currently, social science methodology prefers samples large enough to obtain valid representation based on statistical principles. However, some objections to this latter position have been made by sociologists. For example, Willer (1967) makes a case for scope sampling, a process of finding a number of natural cases fitting the conditions of the theory model and which are ranged along the major dimensions of the theory. Willer argues against the preoccupation of sociologists with random samples in an attempt to gain certified knowledge at the expense of gaining knowledge based on a wide range of subjects.

The theoretical case for small, exploratory samples made by Stephenson (1953) and Willer (1967) needs some empirical support. The comparison of the results for the two separate Q studies shows that small samples can produce similar results representative of the population. This claim of representativeness can only be made for the beliefs, values, and attitudes found in the sample as being a good indicator of the patterns of beliefs, values, and attitudes in the population--i.e., qualitative inferences to the population can be made, but not quantitative inferences about the number of the different types of landowner.

Construct Validity

This form of validity is the most difficult to assess. However, some limited demographic data is available which corroborates the interpretations and provides evidence that the constructs describe patterns of beliefs, values, and attitudes that are producing the different types. Independent evidence for construct validity is necessary.

The essence of construct validation is finding out whether the beliefs, values, and attitudes are in fact the real basis for the discovery of three different types of landowner. This issue can be resolved by deriving hypotheses from the theory underlying the construct. In this case it is predicted that occupation and place of residence will be strongly associated with different belief systems and that the different types of landowner will correlate with these variables.

The limited amount of available demographic data shows that each type has a consistent pattern of occupation and residence characteristics, and very few subjects do not fit the pattern of the type to which they belong. The Concerned Ecologist type is made up of city-dwellers and retired people and does not include fulltime farmers. Of the 26 people composing the Dedicated Farmer type, 16 are farmers and 3 are urban residents who own a farm. The Practical Forester type displays a less consistent pattern of people, but does contain many fulltime workers who

neither live in the city nor are farmers.

The consistency of the demographic data as it fits the type descriptions is taken to indicate that landowner beliefs, values, and attitudes are the underlying basis for the formation of the three types. To some extent an intuitive case for construct validity can be made. Given the nature of the statement sorting task, it seems most likely that only differences in individual beliefs, values, and attitudes caused people to order the statements in a given way.

CONCLUSION

The results of this study provide some evidence to support the case that the results produced by Q method are both reliable and valid. Given the similarities of the types found by the two separate studies, and the many potential sources of error, the data make a strong case for reliability. The restudy design suffers from a small sample size, although consistent results were found. The data from this part of the study indicated that attitude change over the one- to two-year period was slight. Both analyses of reliability suffer from a lack of quantification.

A good case was made for the external validity of the results of Q method, and a weaker case for its construct validity. The importance of the results derives from the summation of all the evidence rather than a consideration of any one piece. Taken as a whole the results suggest that Q method produces results that are both reliable and valid. No conclusive case can yet be made for this assessment. This paper has attempted to bring empirical evidence to bear on these issues and indicates that more evidence needs to be considered before a definitive conclusion is reached.

John R. Fairweather, Department of Sociology, University of Missouri, Columbia, MO 65211

REFERENCES

Brenner, D.J. Dynamics of public opinion on the Vietnam War. In S.R. Brown & D.J. Brenner (Eds.), *Science, psychology, and communication.* New York: Teachers College Press, 1972. Pp. 345-380.

Cronbach, L.J. & G.C. Gleser. Review of *The Study of Behavior*, by William Stephenson. *Psychometrika*, 1954, *19*, 327-330.

Fairweather, J.R. Types of forest landowners in the Missouri Eastern Ozarks. Master's thesis, University of Missouri, 1979.

Gronlund, N.E. *Measurement and evaluation of teaching.* New York: Macmillan, 1971.

Kahle, R.R. & L. Lee. *A Q-methodological study of attitudes toward water resources and implications for using mass media in dissemination of water research results.* (Completion report, U.S. Department of the Interior Project No. X-127-MO) Columbia, MO: Missouri Water Resources Research Center, University of Missouri, 1974.

Kerlinger, F.N. *Foundations of behavioral research.* 2nd ed. New York: Holt, Rinehart and Winston, 1973.

Lewis, B.J. Establishment of typologies for private non-industrial forest landowners in the Eastern Missouri Ozarks: a Q-methodology approach. Unpublished paper, University of Missouri, Columbia.

Neff, W.S. & J. Cohen. A method for the analysis of the structure and internal consistency of Q-sort arrays. *Psychological Bulletin*, 1967, *68*, 361-368.

Stephenson, W. *The study of behavior: Q-technique and its methodology.* Chicago: University of Chicago Press, 1953.

Willer, D. *Scientific sociology: theory and method.* Englewood Cliffs, NJ: Prentice-Hall, 1967.

Without reliability, argument and deliberation cannot proceed. (A. James Gregor)