

Comments on Professor Howard's Quest-Sort

Steven R. Brown
Kent State University

ABSTRACT: The Quest-Sort modification of Q technique is examined in historical context and is shown to have several predecessors, which are criticized for their overemphasis on technicalities at the expense of methodological considerations. Attention is also drawn to the foundations of measurement in Q method and to Stephenson's replacement of classical psychophysics with Fisher's experimental principles. Questions are raised about various of Professor Howard's assumptions about Q methodology.

Professor Howard has done us the service of demonstrating in considerable detail the extent to which Q-technique results remain stable despite changes in format. As he shows, it matters little whether the data are collected in typical Q-sort fashion or via his recommended Quest-Sort, which consists of a number of traits or short sentences appearing on a single sheet and to which the participant can then assign scores. The advantages are said to be in terms of savings in time, space, and administrative costs, and there may well be situations in which these considerations assume proportions that would justify these measures.

I would hasten to add, however, that the occasions are apt to be rare rather than frequent, and that it would not be wise to assume the suitability of the Quest-Sort format under usual circumstances, despite the virtues claimed for it; indeed, the claimed virtues could, in many instances, be considered liabilities instead. The time required to develop acquaintance with another person's subjectivity, for example, will likely be at odds with efficiencies aimed at getting in and out of the interview as quickly as possible.

Author's address: Department of Political Science, Kent State University, P.O. Box 5190, Kent, OH 44242-0001; sbrown@kent.edu

Retrospective on Q Modifications

It is perhaps of historical interest to note that something akin to the Quest-Sort has been used on prior occasions. Stephenson (1949, pp. 76-77), for example, refers to a 1939 study of his in which children involved in a performance test were scored by psychologists who placed check-marks next to a list of 49 items, such as "impulsive," "lacks persistence," "muddled," "flexible," etc. Especially salient traits received two checks, hence each trait received a score of 0, 1, or 2 checks, and the psychologists' scores were then correlated (Q) and factor analyzed (cf. Stephenson, 1953, pp. 232-234). More directly comparable to Professor Howard's study was one by Brown (1982) in which a participant described her feelings about political objects by placing a score (from +4 to -4) in a blank associated with each of 42 adjectives (such as angry, pleased, warmhearted, etc.). This same strategy was also incorporated in studies by Baas (1979) and McKeown (1984), and the Q-METHOD electronic discussion list was named using a similar procedure, i.e., subscribers assigned scores (from 1 to 5 in this case) to each of the several proposed list names, all of which appeared on the screen at the same time. (In none of the above instances, however, was the procedure ever referred to as other than *Q technique* since a special name, such as Quest-Sort, seemed unnecessary: The usual format was simply altered in light of situational exigencies.) The modified Q-sort procedure developed by Jackson and Bidwell (1959) is almost identical to Professor Howard's, and was developed for essentially the same reasons, i.e., to save time and to facilitate the testing of larger numbers of respondents. The idea of connecting small-sample Q technique to large-sample questionnaires also has a history (e.g., Nitzberg, 1981; Stephenson, 1953; Theiss-Morse *et al.*, 1992). As can be seen, therefore, the Quest-Sort proposal is not without precedent.

I hope that Professor Howard will understand my reticence in welcoming yet another "improvement" in Q technique as stemming in large part from the unremarkable track record accompanying such recommendations in the past. In most cases, their sponsors introduced their innovations in the absence of any real understanding of *Q methodology* and then moved on, never to be heard from again, and it was this as much as anything that was responsible for Q's "faddish" quality (Brown, 1968). With few if any exceptions, suggested modifications have been of technical interest only. The following are exemplary:

- Bolland (1985) introduced his recommendations with much fanfare, but as far as I am aware neither he nor anyone else has ever used the free-distribution and nonmetric analysis procedures which he assured everyone would solve the substantial problem which he had identified.
- Cronbach (1953) introduced a *D* statistic which was presumably superior to the correlation coefficient in Q studies, but the significance claimed for it has never energized a following.
- Edwards (1957) introduced procedures to inoculate against the so-called social desirability malady thought to be congenital to Q-sort studies; having once accomplished this, however, he then turned to other enticements and never returned to this ostensibly urgent matter.
- Garrard and Hausman (1985) decided that using a Q sort for purposes of decision making (as opposed to clinical assessment) was so innovative as to require a new name, hence christened their creation the Priority Sort which, as far as I am aware, has never spawned a follow-up.
- Goodling and Guthrie (1956) advanced a set of statistical procedures—which, to my knowledge, have never been utilized since—designed to serve as criteria for selecting Q statements which have minimum *intrasubject* and maximum *intersubject* variability.
- Jackson and Bidwell's (1959) modification was, as noted previously, much like Professor Howard's, but after separately applying their modification to curricula (Jackson, 1956) and administration (Bidwell, 1957), they abandoned it.
- Livson and Nichols (1956) recommended a rectangular distribution because it would maximize the number of discriminations each Q sorter would have to make, yet I know of no one (including Livson and Nichols) who has ever followed this advice. We are all well aware, however, of the wasted time which subsequent Q methodologists have had to invest in defending against the "free vs. forced distribution" non-issue, as with the social desirability non-issue noted above.
- Miller, Wiley and Wolfe (1986) introduced the "F sort," which they claim follows the format of Q technique, but which leads to subsets of *nominal* groupings rather than a single ranking of all the items. Once again, this proposed change in procedure seems to have been recommended just for the sake of recommending something: So far as I have been able to determine, even its innovators have shown no

interest in it since.

- Nahinsky (1967) wished to place Q on a normative footing, and to this end introduced the idea of variance analyzing sorts \times groups \times items: It has been almost 30 years since that idea was stillborn.
- Neff and Cohen (1967), in the same year, made exaggerated claims about the importance of determining that Q statements really belong in the factorial categories to which they are assigned, hence urged that a Q study be preceded by an R factor analysis that served to verify the statement dimensions incorporated into the Q sample: Neither those authors nor anyone else has ever stepped forward to embrace this recommendation.

It is fervently hoped that Professor Howard's innovation will not join the ranks of the above, many of which are of an older vintage mainly dating back to Q's heyday in psychology. As is apparent, these contributions and modifications of Q technique reveal a tendency for critics to focus primarily on technical matters to the exclusion of those deeper methodological considerations which were Stephenson's central concern and which required broad scholarship for their appreciation. Issues more significant than mere Q sorting and its surrogates are involved.

The Psychophysics of Q Method

A word is in order about measurement *per se*, especially with respect to Professor Howard's second study, where Thurstone's method of paired comparison applied to $N=15$ items (105 comparisons) is referred to as the "benchmark" against which the Q-sort and Quest-Sort results are compared for determining *validity*. The assumption seems to be that paired comparisons produce results which are more accurate, or in some sense more true. Stephenson (1953, pp. 60-61, 240-241) went out of his way to indicate how Q differs from both Thurstone's and Likert's procedures as from other psychophysical methods. His primary reference point was Beebe-Center's (1932) forgotten classic *The Psychology of Pleasantness and Unpleasantness*, in which the fundamental distinction was made between the methods of *expression* and those of *impression*. The former reach into autonomic responses and inform on the relations between *hedonic tone* (pleasantness/unpleasantness) and *stimuli*; the methods of impression emphasize the relations between hedonic tone and the individual's *responses*.

Q method is more obviously related to the methods of impression, but there are important differences. The classical methods of impression originated with Fechner, who originally distinguished three—the methods of choice, production, and use—of which the method of choice was further elaborated into the methods of merit (rank-ordering), paired comparison (items compared two-by-two), and single stimuli (each item judged independently). When evaluated in terms of adequacy, accuracy, completeness, and convenience, the method of paired comparison was judged the least effective of the three—the method of single stimuli was judged best—and so can hardly serve as a benchmark against which to appraise either the Q sort or the Quest-Sort. At best, therefore, reliability rather than validity is more likely at issue in Professor Howard's study.

Q sorting has affinities with all three methods of choice. Participants ultimately place statements in an order (merit), for example, but in the initial phases of the sorting they tend to make absolute judgments (single stimuli) by determining that particular statements belong in the positive, negative, or neutral categories. Paired comparisons are implicit throughout, but also occasionally explicit, as when a participant singles out particular statements or statement pairs for more careful consideration. It is a natural feature of Q sorting (and should be a built-in feature of the Quest-Sort as well) that the participant becomes generally familiar with the array of statements before scoring any of them: The Quest-Sorter should therefore be encouraged to read all of the items at the outset so as to form a point of "choice equilibrium" (Stephenson, 1953, p. 60) in relation to which the more detailed item assessments will then be made.

As a side note, Beebe-Center (1932, p. 20) reminds us that Fechner introduced the method of paired comparisons after having observed that his experimental subjects, when presented with a set of stimuli all at once (e.g., all on a single page, as in the Quest-Sort), showed preference for those objects in the middle of a series, regardless of their character: By forcing comparisons pair-by-pair, he sought to overcome this "influence of the mean" (cited in Beebe-Center, 1932, p. 20).

We are further reminded that when items are appraised by the method of single stimuli (e.g., on, say, a 9-point scale in a Likert format), participants show a marked proclivity to use evenly-spaced scoring categories (such as 2, 4, 6, and 8) rather than the entire range, indicating that their responses are affected by the scale as well as the stimuli (Beebe-Center, 1932, p. 26). It is this kind of idiosyncrasy which manifests itself in free-distribution Q sorts, and it is this

idiosyncrasy which Beebe-Center sought to nullify with his "method of percentages of pleasantness" (p. 30) and which Stephenson sought to overcome with his method of forced sorting (Stephenson, 1953, pp. 60-61). The forced-distribution is therefore not a model of the way in which Q sorts appear in empirical reality; rather, it is a model of the Law of Error which is superimposed on the Q sorter so as to overcome the idiosyncratic responses that otherwise occur as part of the measuring process itself and which have little if anything to do with the person's reaction to the subject matter *per se*. I mention this in part since Professor Howard notes that some researchers "permit" a free distribution, as if it were a matter of style rather than a decision based on a body of literature with which some are familiar and some are not.

Ultimately, however, Stephenson considered the differences between Q and the psychophysical methods to be fundamental. Intrinsic to both the methods of expression and impression was the need to eliminate error, mainly through resort to averaging and large numbers: A person's impression of the pleasantness of a stimulus, for example, might be obtained several times for purposes of achieving accuracy. Instead, Stephenson (1953, p. 61) relied on the newer experimental methods of Fisher and the *estimation* of error rather than its *elimination*.

If I have gone into unnecessary detail concerning measurement, it is as a reminder of the psychometric issues which Stephenson faced as these were emerging in the 1920s and '30s, many of which have slipped from view in the intervening years. Critics who seek to modify Q technique are often unaware of what has come before, hence often end up reinventing the wheel and mistaking the disregard which greets their modification for methodological orthodoxy.

Concluding Remarks

Recommending a change in practice usually comes after a certain level of familiarity has been achieved, but several of Professor Howard's comments raise doubts in this regard: He refers to Q statements as "variables," for instance, and to a nine-point Q-sort scoring scale as "typical" for large Q samples; he relies on eigenvalues and the scree test in factor extraction; in one of his studies, he administers the Q sort to more than 200 respondents; he justifies the use of a large person sample under the assumption that subjective meaning in the aggregate requires it; and he relies on varimax for rotational purposes, despite the fact that this solution has apparently produced some non-comparable

results among some of his factors (Table 1). Many of these points, so obvious in other contexts, are problematic within Q method and would normally require justification. I would therefore encourage Professor Howard to reexamine his assumptions with a "not" in each of them before plying his obvious technical skills to that subject matter for which Q methodology was invented.

References

- Baas, L.R. (1979). The Constitution as symbol: The interpersonal sources of meaning of a secondary symbol. *American Journal of Political Science*, 23, 101-120.
- Beebe-Center, J.G. (1932). *The psychology of pleasantness and unpleasantness*. New York: Van Nostrand.
- Bidwell, C.E. (1957). Some effects of administrative behavior: A study in role theory. *Administrative Science Quarterly*, 2, 162-181.
- Bolland, J.M. (1985a). The search for structure: An alternative to the forced Q-sort technique. *Political Methodology*, 11, 91-107.
- Brown, S.R. (1968). Bibliography on Q technique and its methodology. *Perceptual and Motor Skills*, 26, 587-613.
- Brown, S.R. (1982). Imagery, mood, and the public expression of opinion. *Micropolitics*, 2, 153-173.
- Cronbach, L.J. (1953). Correlations between persons as a research tool. In O.H. Mowrer (Ed.), *Psychotherapy* (pp. 376-388). New York: Ronald.
- Edwards, A.L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden.
- Garrard, J., & Hausman, W. (1985). The Priority Sort: An empirical approach to program planning and evaluation. *American Journal of Social Psychiatry*, 5(5), 29-36.
- Goodling, R.A., & Guthrie, G.M. (1956). Some practical considerations in Q-sort item selection. *Journal of Counseling Psychology*, 3, 70-72.
- Jackson, D.M. (1956). Development of a measure of orientation toward core and subject curriculum theories. *School Review*, 64, 250-255.
- Jackson, D.M., & Bidwell, C.E. (1959). A modification of Q-technique. *Educational and Psychological Measurement*, 19, 221-232.
- Livson, N.H., & Nichols, T.F. (1956). Discrimination and reliability in Q-sort personality descriptions. *Journal of Abnormal and Social Psychology*, 52, 159-165.
- McKeown, B.F. (1984). Q methodology in political psychology: Theory and technique in psychoanalytic applications. *Political Psychology*, 5, 415-436.
- Miller, D.M., Wiley, D.E., & Wolfe, R.G. (1986). Categorization methodology: An approach to the collection and analysis of certain classes of qualitative information. *Multivariate Behavioral Research*, 21, 135-167.
- Neff, W.S., & Cohen, J. (1967). A method for the analysis of the structure

- and internal consistency of Q-sort arrays. *Psychological Bulletin*, 68, 361-368.
- Nitzberg, M.C. (1981). Development of Likert scale to measure salience of need for interpersonal relationships with parents and friends. *Adolescence*, 15, 871-877.
- Stephenson, W. (1949). *Testing school children: An essay in educational and social psychology*. London: Longmans, Green.
- Stephenson, W. (1953). *The study of behavior*. Chicago: University of Chicago Press.
- Theiss-Morse, E., Fried, A., Sullivan, J.L., & Dietz, M. (1992). Mixing methods: A multistage strategy for studying patriotism and citizen participation. In J. Stimson (Ed.), *Political analysis* (Vol. 3, pp. 89-121). Ann Arbor: University of Michigan Press.