Commentary

Looking at Reliability and Validity Through Q-Colored Glasses

Karen E. Dennis
University of Maryland

Insisting on the irrelevance of reliability and validity in Q methodology places Q methodologists in the same kind of trap that ensnares their R methodology counterparts: adhering to the specifications of one methodology while conducting research in the other. Alternatively, what is needed is a different way of looking at reliability and validity from within Q methodology, not outside it. Laudan (1977), a philosopher of science, noted, "A research tradition is a set of general assumptions about the entities and processes in a domain of study, and about the appropriate methods to be used for investigating the problems and constructing the theories in that domain" (p. 81). The appropriate methods to be used in examining reliability and validity in Q methodology require a conceptual view that uses a combination of strengths from both quantitative and qualitative research traditions. Whether termed internal consistency or test-retest reliability in the semantics of R (Nunnally, 1978), repeatability in ethnographic research (Denzin, 1978) or reliable schematics as referenced in Q (Expositor, 1987), reliability in these three different research traditions is a matter of approach, not relevance.

And it is exactly the approach that wreaks such havoc, for even Q methodologists who attempt to demonstrate reliability

Author's address: School of Nursing, University of Maryland-Baltimore, 655 W. Lombard Street, Baltimore, MD 21201-1579.

Operant Subjectivity, 1992/1993 (October/January), *16*(1/2), 37-44

and validity within the parameters of Q all too often continue to cling to the vestiges of R, whether intended or not. One of the earliest of these methodologists was Fairweather (1981), who examined reliability and validity by forcing the sorted items into the heuristic factorial design of the Q set envisioned by the investigator, rather than capitalizing on the subjectivity that created the resultant factor structure. In a modification of the test-retest procedure that is more appropriate to Q methodology, intra-individual correlations are calculated for Q sorts administered to several individuals at two points in time (Dennis, 1988; Brown, 1980). While this somewhat traditional approach uses Q sort data without forcing it into the investigator's preconceived conceptual framework, its consistently high correlation coefficients of 0.80 or higher make this kind of stability a virtual "given" when studying more enduring phenomena that do not change over time.

Looking at Validity

A recent attempt to follow the guidelines of Q methodology in evaluating validity unfortunately speaks more closely of R (Brouwer, 1992/1993). Through subsetting the Q and P sets within a single study, collecting Q sorts, using both Q and R factor analysis, and comparing the results, Brouwer contends that "the extent to which the findings concur [between Q and R] will be considered to constitute a mutual validation of the two systems of analysis." This assertion reflects the R concept of concurrent, construct validity in that multiple measures of the same phenomena should elicit highly similar results. However, this stance does not reflect the Q orientation that the subjective perceptions of individuals are fundamentally and categorically different, not more or less of something on a sliding-scale continuum. This erroneous assumption of comparability between the two research traditions undermines much of the work that follows. After citing Brown's (1980) striking example of differences between R and Q factor analysis when considering measurements of the human body (i.e., R produced unintelligible results while Q provided a clear

one-factor solution), it is quite amazing that Brouwer is disappointed when his "metacorrelations" between Q and R fail to reveal similarities between outcomes for persons and outcomes of items. Moreover, the analysis of unrotated factor structures may lend insight into the statistical processes involved, but its exclusive emphasis ignores the operant subjectivity that the factors represent.

Qualitative Approaches to Validity

That Brouwer summarily dismisses qualitative approaches to strengthening validity in Q methodology marks him as intolerant of the parameters that guide ethnographic research as R methodologists are intolerant of the parameters that guide Q. However, he is far from being alone among Q methodologists in such intolerance. Reaching a conceptual rather than literal understanding of validity may resolve some of the long-standing issues and arguments. Just as the R multitrait, multimethod approach to the determination of construct validity is irrelevant for ethnographic research, so is the R notion of an external criterion for a person's point of view irrelevant for Q. Validity in Q methodology refers more to the ability of individuals to accurately share their perspectives on the subjective phenomenon under investigation, and to the researcher's ability to accurately elucidate and portray the subjectivity expressed. Using Brown's (1980) example of a Q set on "being in love," individuals will not be able to accurately portray their subjectivity regarding this issue if they are confronted with items that deal with the score of "love" in a tennis match or the erstwhile Chevy Luv pick-up truck. Nor, as Whiting (1959) charged decades ago, will individuals accurately model their points of view if they resort to a mechanical sorting of cards as if they were dealing a poker game. Both the relevance of items to the concourse and the mechanical versus conceptual sorting of cards indeed are validity issues.

Another qualitative approach to insuring validity involves the determination of equivalence in meaning between the re-

searcher and the participants (LeCompte & Goetz, 1982). This is the tactic that ethnographic researchers use when they bring their conceptual interpretation of the data back to their informants and ascertain whether the informants can corroborate the meaning and identify themselves in the categories that emerged from the data. Q methodologists likewise use this technique during follow-up interviews when they elicit in-depth comments from participants about some of the specifics of their Q sort, and later when they ascertain whether these individuals can identify with one of the factor arrays.

Reliability in Q as Reliable Schematics

Drawing from the tenets of several research traditions, Thomas and Baas (1992/1993) overtly make the case for repeatability in Q methodology as "reliable schematics" and covertly make the case for validity as generalization. These authors cogently note that Q methodology uses generalization in a more qualitative sense, with substantive inference about a phenomenon rather than to a population. These Q methodologists, by taking a conceptual approach to reliability and working within the parameters of the research tradition, provide strong evidence and a convincing argument for replicability (i.e., reliability) in Q methodological inquiry.

Reliable Schematics: One Phenomenon, One Q Set

In the two studies that elicit appraisals of Ronald Reagan, the authors repeatedly claim that the conceptual interpretations of the two factors emerging from each study are "virtually indistinguishable." A close examination of the items from the respective factor arrays provided in the tables substantiates their claim despite the fact that none of the statements drawn from the conceptually equivalent factors R1.1/R2.1 or R2.1/R2.2 are the same. Abduction in the interpretation of factor arrays, however, overrides concern for this lack of identity and emphasizes the entire context of the factor array in the interpretation. Moreover, as LeCompte and Goetz

(1982) noted in their treatise on reliability and validity in ethnographic research, the dynamic nature of human behavior and circumstances virtually precludes the exact replication of studies, regardless of the method employed. On the other hand, it also would be important to examine the items that significantly differentiate the "illusory" and "authenticity" factors in each study to ascertain whether these particular items are identical in wording or similar in feeling. Without the context of the entire factor array to support or refute the assertions of similarity, it is difficult to discern whether the strength of the statements on each factor make a difference. That is, the items that the authors provide to represent the conceptual interpretation of R1.1 are strongly like the opinions of the individuals on that factor (+5 and +4), but the statements of R2.1 are somewhat weaker in their representation of the belief structure (+4, +3, and +2). This same type of pattern holds for the items of R2.1 and R2.2. The comparability of the items that significantly distinguish the factors would make an even stronger case for replicability than the authors already demonstrate.

Relying solely on a conceptual interpretation to make their case for a reliable schematics, Thomas and Baas do not report a Q methodology type of statistical approach to strengthen their claim. Submitting the first-order factor arrays to a second-order factor analysis and finding that R1.1 and R2.1 load together on the same second-order factor would lend statistical credence to their conceptual comparability. The lack of identity from factor-to-factor in the items that are presented in the table would argue that a second-order factor analysis might not bear out the replicability. However, the contribution of the items that are not shown in the table may provide a counterbalance. Since the authors did not conduct this analysis, or did not report it, the outcome is merely speculation. Using the factor arrays in a second-order factor analysis stays within the realm of Q methodology and does not reduce the process to a correlation coefficient as these investigators claimed. As an example, in three separate studies of patients' control during hospitalization, the same Q set and P set, but different in-

dividuals, elicited conceptually similar factors from study to study that loaded significantly together on the same second-order factors (Dennis, 1987, 1990). Thus, the similarities found in the factor arrays and their abductive interpretation received further support through second-order factor analysis.

Reliable Schematics: One Phenomenon, Two Q Sets

Even more methodologically intriguing and equally as supportive of reliable schematics as the studies of Ronald Reagan are the studies that sought subjective appraisals of George Bush. Despite different conceptual frameworks (Alford vs. Hargrove and Nelson), heuristic factorial designs for deriving the Q sets (4×3, $N=60$ vs. $4\times3+2$, $N=50$), and thus completely different items, the findings from the two studies of Bush essentially are as comparable to one another as those from the studies of Reagan. Thomas and Baas contend that reliable schematics are at issue in the feeling states or states of mind that make the factors operant, not in the interpretation per se. Both, however, are germane. Since the same thoughts and feelings may be expressed in different words, it is the abductive interpretation that transcends the linguistics of what specifically is said to portray the subjectivity that undergirds the factor structures. Throughout the discussion of the Bush studies, Thomas and Baas repeatedly attest that there is a resemblance of sentiments between the factors (e.g., B1.2 and B2.2) despite linguistic differences. Such concerns clearly are founded in R. Qualitative researchers have the perspicacity and the necessity to look beyond the specific words that are used by their informants to understand and illuminate the central themes that the words convey.

When viewed as a composite, the methodological studies addressing issues of reliability/generalizability that use a strictly conceptual approach (Thomas & Baas, 1992/1993), or that add to it the examination of raw factor structures (Expositor, 1987) or second-order factor analysis (Dennis, 1987, 1990), consistently demonstrate a reliable schematics between studies of the same phenomenon. Nevertheless, caution is

warranted when comparing factor structures and factor arrays from one study to another. Similar factors may not be expressed between studies of the same phenomenon if the P set differs in known or unknown ways. In a study of nurses' perceptions of the real and ideal hospital work environments, there were three real and three ideal factors for a P set of nurses working in a professional practice model and for a P set of nurses working under a traditional care delivery system (Dennis, 1991). Two of the real and two of the ideal factors were the same between P sets: "interpersonal" and "self-focused" in the real environment, "patient care professional" and "self-focused" in the ideal environment. For the remaining factors, nurses in the professional practice model revealed a "constrained" factor in the real and an "assertive/-controlling" factor in ideal hospital work environment that simply did not exist for the nurses in the traditional care model. The two factors that did not replicate should not be construed as undermining support for the reliable schematics that was so soundly substantiated by the other four factors. Investigators must realize that vagaries between the factor arrays in similar studies may result from a disparity in the P sets, a concept well known to R methodologists who not only are aware that outcomes may vary across populations, but actively search for the nature of the variation.

Summary

Not relieving Thomas and Baas from the tandem studies they decry, R methodologists would argue that along with reliability of the methodology, reliability of each measure (i.e., Q set) must be demonstrated as well. Nevertheless, the former's point is well taken that the convincing results for a reliable schematic may be best targeted to the provisional skeptics rather than "the scholars uncongenial to Q's model of science" who either cannot or will not see beyond the dictums of their own research tradition. For Q methodologists, issues of reliability and validity, such as generalization about a phenomenon versus to a population as well as reliable schematics, may

be realized as worthy of consideration if viewed conceptually through Q-colored glasses.

References

Brouwer, M. (1992/1993) Validity: Q vs. R. *Operant Subjectivity, 16,* this issue.

Brown, S.R. (1980) *Political subjectivity: Applications of Q methodology in political science.* New Haven, CT: Yale University Press.

Dennis, K.E. (1987) Dimensions of client control. *Nursing Research, 36,* 151-156.

Dennis, K.E. (1988) Q-methodology: New perspectives on estimating reliability and validity. In O.L. Strickland & C.F. Waltz (Eds.), *Measurement of nursing outcomes: Vol. 2. Measuring nursing performance* (pp. 409-419). New York: Springer.

Dennis, K.E. (1990) Patients' control and the information imperative: Clarification and confirmation. *Nursing Research, 39,* 162-166.

Dennis, K.E. (1991) Real and ideal hospital work environments in juxtaposition. *Western Journal of Nursing Research, 13,* 761-765.

Denzin, N.K. (1978) The logic of naturalistic inquiry. In N.K. Denzin (Ed.), *Sociological methods: A source book* (pp. 6-29). New York: McGraw-Hill.

Expositor (1987) Reliable schematics. *Operant Subjectivity, 10,* 81-86.

Fairweather, J.R. (1981) Reliability and validity of Q-method results. *Operant Subjectivity, 5,* 2-16.

Laudan, L. (1977) *Progress and its problems: Towards a theory of scientific growth.* Berkeley: University of California Press.

LeCompte, M.D. & J.P. Goetz (1982) Problems of reliability and validity in ethnographic research. *Review of Educational Research, 52,* 31-60.

Nunnally, J.C. (1978) *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Thomas, D.B. & L.R. Baas (1992/1993) The issue of generalization in Q methodology: "Reliable schematics" revisited. *Operant Subjectivity, 16,* this issue.

Whiting, J.F. (1959) Needs, values, perceptions and the nurse-patient relationship. *Journal of Clinical Psychology, 15,* 146-150.