On Validity and Replicability

Steven R. Brown
Kent State University

We are indebted to Professors Brouwer (1992/1993) and Thomas and Baas (1992/1993) for giving sustained attention to the issues of validity and reliability, and their connection to replicability and generalization. William Stephenson bequeathed us few guidelines on these matters. He made scant mention of validity in particular (Stephenson, 1953, index), and distinguished it from verification: "... one would study three or four 'single cases' for verification, but not for validation" (Stephenson, 1983, p. 51). Validity he associated with large sampling, i.e., "only when normative conditions are involved ..., for the sake of generalization as to fact" (p. 54).

In R methodology, validity is an issue because of the annoying fact that there is a respondent between the observer and the trait being measured, just as in medicine there is between physician and disease an interceding body that can interfere with diagnosis: bullet fragments and surgical clips, for instance, can mislead a magnetic resonance imaging scanner, and a gas pocket can deny ultrasound a clear view of the disease (Baron, 1992, pp. 38-39). However, in Q methodology, due to the subjectivity involved, it is the respondent who is doing the measuring, and this is the *only* person who can do so, at least on a first-hand basis. As Stephenson (1972) has remarked in this regard:

> Objective measurements and observations can, in principle, be made by everyone (or by a piece of apparatus), whereas measurements and observations of a person's subjectivity can be made only by himself. (p. 17)

Author's address: Department of Political Science, Kent State University, Kent, OH 44242-0001.

Operant Subjectivity, 1992/1993 (October/January), *16*(1/2), 45-51

This originally appeared in italics, indicating that Stephenson considered it to be important. And this principle, which marks the boundary between Q and R, effectively removes the issue of validity in Q (except, perhaps, for the Q sorter!). That the respondent may be lying or self deceived or simply verbalizing an attitude borrowed from a recent talk show is another matter.

As for reliability: Stephenson has only slightly more to say, most of it technical. As he shows (Stephenson, 1953, pp. 174, 283), specificities can far outweigh factor communality so that the reliability of individual Q sorts could be respectably high at the same time that the factors themselves could be somewhat unreliable (with stable specificities accounting for all the reliability). Or the reverse: factors could be reliable even though individual Q sorts are relatively unreliable. From a purely technical standpoint, therefore, reliability by itself may not provide especially useful information.

This matter comes into sharper focus when we switch from the extensive to the intensive case -- e.g., to a single-case study under multiple conditions of instruction. The studies by Brouwer, Thomas, and Baas may not provide good illustrations of this due to the strong beliefs involved, as will be discussed below, but under other circumstances it is readily evident that most of us have multiple selves and perspectives available to us, any or all of which may be replicable and statistically reliable. A reliability coefficient representing a single Q sort given only once (and then readministered) may reveal only a small portion of the existing terrain, and almost nothing about dynamics -- e.g., when the visible and statistically reliable Jekyll of factor A is suddenly superseded by the Hyde of factor B, which may also be reliable even though rarely seen.

But even in opinion and attitude studies of the above kind, the factors obtained merely provide us with a single vantage-point, which we can only hope gives us a clear view of the attitudinal landscape. Typically, the factors are orthogonal, but even between orthogonal factors there are inevitably a few statements in opposition, and any one of those statements

could provide the basis for yet another probe into the same domain. Assume, for example, that statement *x* has received scores of +4 and -4 in otherwise orthogonal factors A and B. Interviews about *x* with subjects from A and B would likely produce an inherently polarized concourse, and bipolar factors in contrast to the previous orthogonal ones; or perhaps three factors (XYZ) will emerge where before there were only two (AB). None of this implies a lack of replicability necessarily, nor could we claim that XYZ invalidates AB; rather, XYZ is the change of view that is inherited as a consequence of the new vantagepoint.

From a technical standpoint, of course, we implicate reliability each time we compare scores between factors since individual reliability enters into standard error formulas: we cannot know whether factor scores of +4 and +2 refer to an actual difference between two factors unless we have some sense of how reliably the two statements are distinguished by the individuals comprising the factors. (So, we ask them to do it again, just to make sure.) Reliable or not, however, we long ago learned not to take factor scores at face value: consensus statements may mask great differences, and distinguishing statements may rest upon semantical distinctions rather than true differences of opinion. It's even possible for people with different views to end up on the same factor for different reasons, and none of this is necessarily incompatible with high reliability coefficients.

What do the articles under consideration bring to this concourse?

Marten Brouwer is concerned mainly with validity, and he entertains the idea that R and Q might provide mutual validity: This is more than just flirting with R methodology; it's romantic involvement. Brouwer's carefully controlled experiment harkens back to the early controversies between R and Q factor analysis, and to the reciprocity (or not) of their results. His findings support nonreciprocity, as Stephenson expected: factorial differences are bound to arise due to the different normalization which occurs in the rows compared to the columns of a data matrix, even under conditions of univ-

ersality of measuring unit, i.e., when the *same* matrix of data is at issue. But this special situation almost never occurs in reality, and it certainly does not distinguish a science with self reference from one without -- hence, Stephenson's (1953) assertion that "there never was a single matrix of scores to which *both* R and Q apply" (p. 15). Ultimately, Brouwer's comments are less apropos the distinctions between System 1 (R) and System 2 (Q) as between the latter and its transpose (System 4) (Stephenson, 1953, p. 51). Systems 2 and 4 share the same matrix of data; Systems 1 and 2 do not.[1]

Thomas and Baas' "tandem" strategy is also carefully crafted, and raises the issues involved to a more abstract and (one is tempted to say) *hermeneutic* level, i.e., to the level of interpretation as opposed to statistical demonstration. Their approach is reminiscent of Daily's (1973) comparison of his factors with those from an independent study on essentially the same topic: the factorial fit was not as clear in that case, but was discernible nonetheless. Still, one cannot help wondering to what extent subject matter has dictated findings. Daily's study was of a highly controversial trial, and Thomas and Baas' topics are also controversial, as is Brouwer's on ideological cleavages. In each of these instances, relatively strong beliefs and values are at issue, for which we would expect greater reliability and easier verification due to the fixities involved. Would we be as successful were we to study topics less subject to the formative and maintaining influences of social controls?

But to emphasize technical features is to toy at the margins and needlessly grant R the home court advantage. What we really want to know is what is essential to Q methodology *qua* methodology as distinguished not only from R methodology, but also from necessary accommodations arising from special

[1] Q and R subsume such different domains that it's hard to imagine how either could ever validate the other in the usual sense. One is reminded in this regard of Kierkegaard's comment (I have no idea where it is located) regarding science and religion -- that it is one thing to stand on one leg and prove the existence of God, and quite another to fall to your knees and thank Him.

applications of Q technique and method. Were Q employed in a typical before/after experiment, the question of reliability would naturally arise -- we would need to know how much confidence we could place in factor loading changes following the experimental treatment -- but reliability would matter here because of the nature of experiments, not because of something intrinsic to Q methodology.

I agree with the authors that we must avoid the extremes -- that reliability and validity are either "utterly inconsequential" or absolutely essential -- and examine circumstances. In this regard, it is apparent that interest in precision is highest when *assessment* is assumed to be at issue; i.e., assessment of entities which are presumed to exist a priori, as in the assessment of intelligence or blood pressure, or when a P set is interrogated in hopes of detecting a previously identified audience segment. Brouwer, Thomas, and Baas' concerns are of this same kind: to strengthen conviction that some thing exists, *as a matter of fact,* by inducing it to reveal itself again.[2]

However, Q's use as an assessment device is a special application rather than a general feature. Still, Q always seems to perform well in these respects when compared to R (at least when Q methodologists are in charge of the investigation!), and it did so when D'Agostino (1984) sought to demonstrate replicability at the point of theoretical rotation. "But suppose it had been otherwise [Stephenson (1984) said], that the rotation was not replicable: Would all be lost for Q?" (p. 87).

Stephenson's answer was no, and, on this one main trail which he left regarding these matters, he based his assertion

---

[2]Medicine, too, is concerned with establishing facts for practical purposes, as noted above, and of controlling conditions affecting facts (pulse after exercising is different than before); and this need for precision in medical assessment, fortified by ever-present liability risks, is perhaps what has spilled over into an overscrupulous concern for precision in Q, as in Larson's (1987) CARE-Q Caring Assessment Instrument and in the practice of explicitly writing out Q-sort conditions of instruction so as to reduce presumed perils to reliability and validity (Dennis, 1988, p. 415).

on a distinction introduced by McKeon (1967) between statements of *facts* and statements of *problems.* Science is normally thought of in terms of the former -- of falling apples, perturbations in Mercury's perihelion, the undular or dart-like behavior of light, and (we might add) of the replicability or not of Q factors: all of these are matters of truth or falsity and controversies over meaning. No one of course questions the importance of facts to the growth of science, but science is also a history of problems which "are subject to discussion and controversy to explore the range of their meanings and the variety of facts to which they apply" (McKeon, 1967, p. 26) -- e.g., of gravitation, which implicates not only apples, but pendulums and the tides; of relativity, which applies to electrons as well as galaxies; of ludenic behavior, that brings observations about charismatic leadership into the same discussion with newsreading and shopping behavior; of caring, which may reveal itself not only on the nursing floor, but also in the pruning of a bonsai tree, the preparation of a meal, or the selection of just the right tie to complement a shirt. These too are matters of generalization: not as to facts, but to more abstract concepts.

Valid and replicable observations are important, and we are grateful to Marten Brouwer, Dan Thomas, and Larry Baas for showing how Q data can stand on their own in these respects: this will make critics pause and reconsider. At the same time, we mustn't lose sight of the limited role that facts play in science. Galileo wouldn't have gotten far had he been content to document the reliable performance of metal balls rolling down inclined planes, however precious these observations obviously were for the theories which subsequently rested upon them. The accomplishment which lifted him from the endnotes to the main text in the history of science was to penetrate these surface impressions and to reveal the deeper invariances beyond. And whereas facts are important to Q methodology in the same way as in physics and medicine, we too must look beyond and recognize that "the special possibility for Q is in the other direction, a quest for concepts of importance" (Stephenson, 1984, p. 90).

References

Baron, R.J. (1992) Why aren't more doctors phenomenologists? In D. Leder (Ed.), *The body in medical thought and practice* (pp. 37-47). Dordrecht, The Netherlands: Kluwer.

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

D'Agostino, B. (1984) Replicability of results with theoretical rotation. *Operant Subjectivity, 7,* 81-87.

Daily, J.H. (1973) *Dimensions of political attitudes: A Q technique study of public reactions to the Calley verdict.* Doctoral dissertation, Kent State University.

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.

Larson, P.J. (1987) Comparison of cancer patients' and professional nurses' perceptions of important nurse caring behaviors. *Heart & Lung, 16,* 187-193.

McKeon, R. (1967) Scientific and philosophic revolutions. In F.J. Crosson (Ed.), *Science and contemporary society* (pp. 23-56). Notre Dame, IN: University of Notre Dame Press.

Stephenson, W. (1953) *The study of behavior.* Chicago: University of Chicago Press.

Stephenson, W. (1972) Applications of communication theory: I. The substructure of science. *Psychological Record, 22,* 17-36.

Stephenson, W. (1983) Comment [on Thompson, Frankiewicz & Ward (1983)]. *Operant Subjectivity, 6,* 50-55.

Stephenson, W. (1984) Comment [on D'Agostino (1984)]. *Operant Subjectivity, 7,* 87-91.

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

Thompson, B., R.G. Frankiewicz & G.R. Ward (1983) Cross-technique validation of attitude measures. *Operant Subjectivity, 6,* 37-50.