

IN MEMORIAM

Warren T. Norman (1930–1998): An Appreciation

Lewis R. Goldberg

University of Oregon and Oregon Research Institute

A noted social psychologist, who was carrying out a type of meta-analysis of past studies on personality structure, remarked to me about the wealth of relevant information in each of the reports of studies conducted by a particular personality psychologist, in contrast to the frustrating lack of such information in the reports of his colleagues in social psychology. He attributed the difference to failures in the quantitative training of social psychologists. I did not disabuse him, but inwardly chortled over the happy selection of Warren Norman as his representative of personality psychology.

Warren Norman will not be remembered for the quantity of his scientific publications. He will long be remembered for their *extraordinary* quality. During his 37-year career at Michigan, in addition to a number of book reviews and test reviews, he published about a baker's dozen scientific reports, each a sparkling jewel of creativity and rigor. He helped set the research agenda for the field of personality assessment, and his contributions have been among the most widely appreciated of any of his scientific peers.

In his first major publication (Norman, 1959), he demonstrated the magnificent craftsmanship that would characterize all his later ones. In but four short pages, he (a) provided a compelling rationale for the need to assess the test–retest stability and interjudge consensus of judgments of affective meaning derived from Osgood's Semantic Differential technique; (b) devised a novel procedure for assessing such stability and clearly conveyed its rationale; (c) showed the expected value of the new procedure when responses are random; (d) used both his new metric, *plus* the more usual correlational metric, to demonstrate:

Invited Address at the Departmental Symposium Celebrating his Distinguished Career, Department of Psychology, University of Michigan, Ann Arbor, MI May 19, 1994.

Correspondence may be addressed to Lewis R. Goldberg, Oregon Research Institute, 1715 Franklin Blvd., Eugene, OR 97403-1983. E-mail: lewg@ori.org.

- (1) the average stability of the ratings by individual subjects;
- (2) the extent of individual differences in such stability values;
- (3) the extent of differences among the concepts, among each of the individual rating scales, and among the factors derived from those scales; and
- (4) the extent of gender differences at each of these levels of analysis.

In addition, he presented findings from analyses of the degree of consensus of the total semantic spaces of individuals across all concepts; and, finally, he repeated all of these analyses at the level of groups means, thus providing evidence on how large a difference actually makes a difference in practice. This tiny article, written when Warren was only one year out of graduate school, can still be used as a model for teaching precision in scientific writing.

Warren received his B.S. in Mathematics and Natural Science (1952), his M.A. in Statistics and Educational Psychology (1955), and his Ph.D. in Psychology (1957), all from the University of Minnesota. This “staying power” persisted at the University of Michigan, his only permanent faculty home. During his long tenure at Michigan, he served as Chairman of its University Senate, and Head of its Department of Psychology. In addition, Warren served as the Chairman of the Board of Scientific Affairs of the American Psychological Association, and he was elected President of the Society of Multivariate Experimental Psychology; in both cases, he was the youngest person to have been placed in those roles. He has served on the Editorial Board of such publications as the *Annual Review of Psychology*, *Psychological Bulletin*, the *Journal of Applied Psychology*, and *Organizational Behavior and Human Performance*. He also served at various times as a panelist for the National Science Foundation, the National Institute of Mental Health, and the U.S. Office of Education, as well as a member of the Research Committee for the Graduate Record Examination.

For the young Dr. Norman, only six years out of graduate school, 1963 was a *very* good year. Three of his publications appeared that year, each in a different APA journal, all of seminal importance to the field, and one of which was to become a citation classic. In one of these articles, Warren proposed a novel procedure both for detecting impression management in self reports and for constructing scales that would be relatively impervious to such faking response sets. In the second of the trio, he demolished what was then called the “Deviation Hypothesis,” namely that subtle test content was equally as useful as obvious content in constructing self-report scales. Warren was the first to demonstrate the strong relation, later confirmed by Ashton and Goldberg and by Jackson, between item face validity and empirical cross-validity. And, in his citation classic, entitled “Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomi-

nation personality ratings," he provided the most salient and scientifically compelling early example of what has now come to be known as the Big-Five factor structure.

The Big-Five structure was based on what is called the lexical hypothesis, of which Warren's descriptions have become the most widely cited: "Attempts to construct taxonomies of personality characteristics have ordinarily taken as an initial data base some set of perceptible variations in performance and appearance between persons or within individuals over time and varying situations. By far the most general efforts to specify the domain of phenomena on which to base such a system have proceeded from an examination of the natural language" (Norman, 1963, p. 574). Later, he argued that a truly comprehensive (or in his words, "exhaustive") taxonomy of personality attributes must take as its fundamental data base ". . . the set of all perceptible variations in performance and appearance between persons or within individuals over time and varying situations that are **of sufficient social significance, of sufficiently widespread occurrence, and of sufficient distinctiveness to have been encoded and retained as a subset of descriptive predicates in the natural language during the course of its development, growth, and refinement**" (Norman, 1967, p. 2).

Ironically, Warren's most famous publication pales in significance when compared to two of his other ones that are rarely, if ever, cited today. Page for page, and pound for pound, the two most outstanding contributions to personality assessment in the decades of the 1960s and 1970s were an article and a book chapter, both written by the same author, namely Warren Norman. The first of this magnificent duo was an article published in *Psychological Bulletin*, entitled "On estimating psychological relationships: Social desirability and self-report" (Norman, 1967). The article begins: "'How highly related are variables X and Y ?' is a fundamental and ubiquitous question in empirical science. It is also, despite its seeming simplicity and directness, a highly complicated one. A major tenet of the following article is, in fact, that this apparently singular and straightforward query, as it arises in typical research settings, actually entails an entire set or multiplicity of separate and logically independent forms of the question. It is further argued that the kinds of data ordinarily collected in even the most rudimentary empirical studies of such relationships provide a basis for computing separate and appropriate answers to each of the possible variants of the question, and that, in general, these several answers will not be the same" (p. 273).

In this monumental article, Warren provided in exquisite detail and in both algebraic and verbal forms an analysis-of-variance type of decomposition of the variances, covariances, and correlations that can be used to express facets of the relations among variables, subjects, and the critical interaction between the two—the latter being proposed for the first time as the actual relation in which most scientists are interested but which they virtually never

actually measure. This seminal bit of original thinking foreshadowed the classic contribution of "Generalizability Theory" by Cronbach and Gleser, as well as some of the applications of today's great fad, structural equation (or "causal") modeling. Had personality psychologists only understood the import of this article, and used it to guide their theoretical and empirical analyses during the "cross-situational consistency controversy" of the 1970s, a decade of fruitless debate could have easily been avoided, and far more penetrating conclusions about trait consistency could have been drawn.

As magnificent a contribution as was this *Psychological Bulletin* article, it was still not Warren Norman at his very best. That honor must go to a short chapter entitled "Psychometric considerations for a revision of the MMPI" (Norman, 1972). In my judgment, there is more psychometric savvy in that publication than most of us produce over a lifetime. Warren was invited to deliver an address at a conference on future revisions of the MMPI, and this chapter was an extended version of his conference talk. In it, he lamented "I come not to bury the Mult *nor* to praise it. The first would certainly be premature, and the second unnecessary" (p. 60); later, he described the currently popular Deviation Hypothesis as "empiricism gone mad as well as blind." In this gem of distilled psychometric wisdom, Warren made many cogent points, among which were the following:

- (1) ". . . *it is unreasonable*, given what we already know of the complexity of human behavior, the personological determinants of responses to inventory items, the effects of situational conditions on test performances, the ambiguities of verbal items both between and within various language communities, and the host of other things we have learned over the past several decades *to expect any single, fixed format, verbal stimulus and (quasi-) verbal response inventory to be a sufficient, or even a generally efficient, means for assessing human personality for all purposes*. The day is past when we can reasonably expect to get by with a jackknife, even a big and elaborate one, for all the things we have to do" (p. 62).
- (2) "From a strictly diagnostic viewpoint, the Multiphasic is a mess! Its original clinical criteria are anachronistic; its basic clinical scales are inefficient, redundant, and largely irrelevant for their present purposes; its administrative format and the repertoire of responses elicited are, respectively, inflexible and impoverished; and its methods for combining scale scores and for profile interpretation are unconscionably cumbersome and obtuse" (p. 64).
- (3) "First, the individual scales in most cases are highly heterogeneous in content. Scores are, accordingly, difficult to characterize simply or to describe concisely and unambiguously. Although this lack of a clear *internal* basis for interpretation may have been tolerable when one

could appeal to a well-established diagnostic type or syndrome as an alternative (at least for “high” scores), the gradual blurring and eventual abandonment of this external framework and the further complications introduced by the need to consider various configurations of two or more scales simultaneously made it virtually necessary to evolve an entirely new typal system to which to appeal for meaning. . . . And this, of course, was just what was done in the development of profile coding methods and in the construction of the Atlas. But if one were going to set out to build a new taxonomy of psychopathology on the basis of test behavior on the MMPI, what an extremely awkward and self-restricting way to go about it!” (pp. 64–65).

- (4) “. . . diagnosis, if it is to be done efficiently and, in some cases, if it is to be done at all, is inherently a *multistage, sequential, and branching* sort of enterprise in which discriminations that one must make at each stage depend upon what [one] has already learned, including both what [one] already knows to be relevant as well as what [one] knows to be impertinent” (p. 69).
- (5) And, finally, he concluded: “. . . if we are ever to come to a full measure of maturity in this field, our methods, procedures, and instruments are going to have to be far more sophisticated than even the most complex of those currently available. I am optimistic enough to believe we are moving in that direction even though the pace of that progress often seems to me unnecessarily and agonizingly slow” (p. 82).

Warren Norman may have best summed up his major scientific tenets at the end of an article entitled “‘To see ourselves as others see us!’: Relations among self-perceptions, peer-perceptions, and expected peer-perceptions of personality attributes” (Norman, 1969):

“. . . the day has passed when anything of much value is going to be learned by ‘one-shot’ studies using just any conveniently available inventory, sample, and data-analysis procedure. The need at the present is for systematic investigations that are extensive in scope, that are both multivariate and multi-method in design, and that span an appreciable range of variation on developmental, demographic, and situational facets. . . . It seems unlikely also that conventional, automatic, and “canned” analysis programs are going to suffice for processing the kinds of data that these more elaborate studies will yield. Such methods were, by and large, designed for other and simpler models, and they often serve effectively for such uses. But their suitability for sorting out the diverse kinds of information implicit in complex batteries administered to stratified samples under multiple task conditions is, in many cases, highly questionable. . . . [T]here is nothing simple, routine, or automatic about research on personality assessment these days; either in

the design of investigations or in the analysis and interpretation of results. If we are to make the progress that now seems possible in this area we will have to have the courage and perseverance to undertake large and complex investigations on a programmatic basis and will also have to have the ingenuity to adapt or invent methods of analysis to the demands that such research designs present" (pp. 441–442).

Personally, Warren took neither himself nor his achievements too seriously, and he would have insisted on casting his contributions in more humble terms. On the few occasions when I have seen him out of sorts, it was typically because someone referred in print to "The Norman Five," a faux pas that would make him shake his head in wonder at the ignorance of those who did not understand that this factor structure should far more accurately be called "The Tupes and Christal Five." None of us who knew him will ever forget him.

A BAKER'S DOZEN OF WARREN NORMAN'S BEST

- Norman, W. T. (1959). Stability-characteristics of the semantic differential. *American Journal of Psychology*, **72**, 581–584.
- Norman, W. T. (1960). A spatial analysis of an interest domain. *Educational and Psychological Measurement*, **20**, 347–361.
- Norman, W. T. (1963). Personality measurement, faking, and detection: An assessment method for use in personnel selection. *Journal of Applied Psychology*, **47**, 225–241.
- Norman, W. T. (1963). Relative importance of test item content. *Journal of Consulting Psychology*, **27**, 166–174.
- Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, **66**, 564–583.
- Norman, W. T. (1965). Double-split cross-validation: An extension of Mosier's design, two undesirable alternatives, and some enigmatic results. *Journal of Applied Psychology*, **49**, 348–357.
- Norman, W. T., & Harshbarger, T. R. (1965). Matching components of self-report and peer-nomination personality measures. *Psychometrika*, **30**, 481–490.
- Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology*, **4**, 681–691.
- Passini, F. T., & Norman, W. T. (1966). A universal conception of personality structure? *Journal of Personality and Social Psychology*, **4**, 44–49.
- Norman, W. T. (1967). On estimating psychological relationships: Social desirability and self-report. *Psychological Bulletin*, **67**, 273–293.
- Norman, W. T. (1969). "To see ourselves as others see us!": Relations among self-perceptions, peer-perceptions, and expected peer-perceptions of personality attributes. *Multivariate Behavioral Research*, **4**, 417–443.
- Passini, F. T., & Norman, W. T. (1969). Ratee relevance in peer nominations. *Journal of Applied Psychology*, **53**, 185–187.
- Norman, W. T. (1972). Psychometric considerations for a revision of the MMPI. In J. N. Butcher (Ed.), *Objective personality assessment: Changing perspectives* (pp. 59–83). New York: Academic Press.