©Academy of Management Review 1982, Vol. 7, No. 2, 296-299

On the Measurement and Control of Beta Change: Reply to Terborg, Maxwell, and Howard

ACHILLES A. ARMENAKIS
ARTHUR G. BEDEIAN
Auburn University

In response to an earlier article in which the present authors discuss scale recalibration and concept redefinition, Terborg, Maxwell, and Howard voice certain criticisms. This reply comments on each of the criticisms and cites evidence that suggests that the retrospective procedure is subject to respondent rhetoric—the omission, juxtaposition, and embellishment of recollections—and thus is of limited value.

The subjects of scale recalibration (beta change) and concept redefinition (gamma change) were introduced to management researchers by Golembiewski, Billingsley, and Yeager (1976). Researchers in other fields, especially psychophysics, however, have been tussling with the problem of instability of temporal measurement for some 50 years. In this respect, the research of theoreticians from a variety of disciplines has been an attempt to provide some guidance in this area of measurement. Because no totally acceptable solution has been presented, the present authors share Terborg, Maxwell, and Howard's hope that this type of exchange will contribute to further understanding in this yet developing area.

Thus, the purpose of this reply is not to "defend" the authors' own procedure, nor is it to "attack" the procedure set forth by Terborg, Howard, and Maxwell (1980). The intent is to review the basics of both procedures and hence permit researchers to consider their relative benefits and limitations.

Retrospective-Then Designs

In contrast to what one might surmise from reading Terborg, Maxwell, and Howard's preceding remarks, the use of retrospective-then designs has been widely studied. Furthermore, the benefits and limitations of retrospective designs have been thoroughly documented. See, for example, Moss and Goldstein (1979). Cherry and Rodgers draw two important conclusions that are especially pertinent for those interested in using a retrospective design. First, "the best that a retrospective enquiry can hope to achieve is an approximation of the event or condition as experienced by the participant" (1979, p. 40). Second, "the influence of [intervening] events on retrospective attitude data is now so well recognized that studies based on information of this sort alone are generally found to be of little value" (1979, p. 40). Although accepted in practice, research (and experience) thus suggests that retrospective designs, not unlike other methodologies, are acceptable only if one is willing to acknowledge their limitations in interpreting the findings that they yield. Two recent studies serve to illustrate the consequences of using retrospective designs.

Green and Wright (1979) employed a retrospective design in an attempt to investigate specifically the reliability of the data that it produced. Although their study was conducted within a school situation, its findings are directly applicable to other settings. The study involved children between the ages of 5 and 14 years and their respective teachers. For purposes of analysis, the teachers

were divided into three categories, based on years of professional experience: (1) 6 or fewer years, (2) 7 to 12 years, and (3) 13 or more years. The design of the study required that the teachers recall the negative and positive behaviors of selected children immediately following a 30-minute class period. The children were requested to do the same. In order to establish a baseline for comparing the results from the two groups, five trained observers were used to determine the actual frequency of both positive and negative behaviors. The interrater reliability of the observers was .96.

An analysis of the resulting data revealed that both the children and their teachers significantly underestimated the frequency of negative behaviors. Further, the data were such that no conclusions could be drawn regarding the number of positive behaviors. In discussing these findings and the use of retrospective designs, Green and Wright concluded that the inability of the children to provide accurate data on their own behaviors was related to developmental and maturational factors, whereas the inability of teachers to provide accurate recall data was related to level of professional experience. Putting the findings related to the children aside and focusing on the teachers, certain organizational parallels are obvious. The Green and Wright study indicates that teachers in a classroom setting are unable to recall the incidence of student behaviors after a brief 30-minute period. In comparison, the confidence that one would be led to place in employees or their supervisors more reliably recalling events and feelings after a lapse of perhaps several weeks certainly would have to be limited. Employees and supervisors, just as teachers, also have various levels of experience. Moreover, the volatile conditions of most industrial settings (compared to most classrooms) would seem likely to generate an even greater decay in recall accuracy.

A second study casts even further question on the use of retrospective designs. Following the conclusion of an 8-week course, Rippey, Geller, and King (1978) asked 34 medical students to recall their answers to 46 questions on a precourse examination. Of the 34 students, 14 incorrectly recalled more than 50 percent of their responses. The remaining 20 students fared slightly better, correctly recalling somewhere between 50 percent and 78 percent of their answers. In this connection, Terborg et al. contend

> there is no reason to suspect Then scores except in situations where it is to the participants advantage to give false Then responses, where participants are confused as to the instructions, or where participants in a no-treatment control group are asked to give Post and Then ratings within a few hours or days of the Pre-ratings (1980, p. 114).

The results of the Rippey et al. study, however, would seem to cast serious doubt on this contention. The two studies cited, as well as other research, indicate that retrospective designs almost invariably are subject to respondent rhetoric—the omission, juxtaposition, and embellishment of recollections. Despite Terborg, Maxwell, and Howard's claims to the contrary, questioned here is their assumption that respondents are capable of accurately recollecting and articulating past events. The bulk of the available evidence simply does not support this conclusion.

Specific Criticisms

One point will be clear from the following comments. Neither the present authors nor anyone of whom they are aware is capable of providing a definitive response to several basic questions involved in the measurement of change. Part of this inability clearly is due to the limitations of existing statistical methods. Additional problems may be attributed to the quality of data typically encountered in the social sciences. Again, however, it is hoped that this type of dialogue will contribute to the resolution of the difficulties that yet exist.

Terborg, Maxwell, and Howard raise seven major issues concerning the present authors' procedure for measuring and controlling beta change. Each will be discussed.

Terborg, Maxwell, and Howard's first two criticisms deal with the steps proposed for detecting gamma change and the definition of alpha change. Despite Terborg, Maxwell, and Howard's comments to the contrary, it still is felt that the approach presented for detecting gamma change, and the definition of alpha change are consistent with the Golembiewski et al. (1976) change typology and are appropriate in conjunction with the use of an ideal scale. At the same time, it is felt that the approach proposed by Terborg, Maxwell, and Howard for detecting gamma change and their definition of alpha change also are consistent with the Golembiewski et al. typology and are appropriate

for use in conjunction with a retrospective design. The first point, however, still remains. Data, and the conclusions that result, are only as good as the methods that produce them. In this respect, the shortcomings of the retrospective-then design must be acknowledged.

The third criticism raised by Terborg, Maxwell, and Howard deal with the use and meaning of ideal scores. The present authors have previously defined "ideal" as "respondent perceptions of perfect or model conditions (i.e., conditions existing in the form of an idealized mental image)" (Bedeian, Armenakis, & Gibson, 1980b, p. 201). The intention was to include this definition in the *Academy of Management Review* paper (Bedeian, et al., 1980a) with which Terborg, Maxwell, and Howard take exception. The above quoted definition, however, along with an accompanying paragraph, were deleted by the editor after the acceptance of the paper. In any case the paper does contain a sample item that clearly indicates a request to respondents to indicate "this is how I'd *like* it to be." Terborg, Maxwell, and Howard seem to feel that the request could be interpreted as meaning either "this is what I *need*" or "this is what I *expect.*" The present authors of that paper disagree. On the occasions that an ideal scale has been used by the present authors, neither questionnaire pretesting nor respondent debriefing indicated any misunderstanding in this regard. In any case, if such a reconceptualization had occurred, it would have been articulated in subject responses and therefore would have been reflected in the correlation of ideal scores across time.

Terborg, Maxwell, and Howard's fourth criticism is that the use of an ideal scale likely will result in a restriction of the range of responses. On the surface this may seem plausible, but it is felt that sufficient evidence exists to suggest that this is not the case. See, for example, Armenakis, Bedeian, & Niebuhr (1979) and Armenakis and Zmud (1979). Furthermore, if differences in responses from pretest to posttest do create the effects illustrated in the example provided in the preceding remarks by Terborg, Maxwell, and Howard, then use of an ideal scale seems supported. The drop in r from .65 (i.e., the r of the pretest with posttest A) to .20 (i.e., the r of the pretest with posttest B) satisfies the requirements for gamma change. Terborg, Maxwell and Howard seem distressed that this drop in r was the result of changes in responses to 2 items on a 10-item questionnaire. What they do not mention is that a change in two responses means that 20 percent of the total responses from pretest to posttest B are different. The question therefore seems to be whether or not a change in 20 percent of the responses on a questionnaire incorporating an ideal scale indicates gamma change.

The fifth criticism advanced by Terborg, Maxwell, and Howard deals with the statistical assumption of independence. This, in fact, is a valid criticism common to both of the suggested procedures. (See Terborg, Howard, & Maxwell, 1980, p. 115.) Although these authors did not emphasize a solution, they accept Terborg, Howard, and Maxwell's suggestion that the statistics involved be interpreted descriptively rather than inferentially.

A sixth criticism deals with questionnaire length. It is Terborg, Maxwell, and Howard's belief that use of an ideal scale unnecessarily lengthens a questionnaire. By comparison, however, their approach seems to offer little advantage. In a time series design, say, with four observations, 0_1, 0_2, 0_3, 0_4, using their proposed retrospective-then design requires responses to "now" scales at all observations and responses to "then" scales at three observations (0_2, 0_3, 0_4). Using an ideal scale, one also must respond to ideal and now scales at each administration. One advantage of the present authors' procedure, however, is its simplicity. In contrast, use of the procedure proposed by Terborg, Maxwell, and Howard results in a rather complex situation. At 0_2, 0_3, and 0_4, for instance, a respondent would be required not only to recall a given state of affairs, but to do so at *three* different points in time. Again, given the earlier cited research, one can only wonder about the reliability of the data that would result.

The final criticism advanced by Terborg, Maxwell, and Howard "concerns the ambiguity over whether beta change has occurred." Their criticism here seems to be with the computation of the magnitude of the values for the slope and the intercept of the regression function described in the proposed procedure. It is agreed that slope may seldom equal exactly 1 and the intercept may seldom equal exactly 0. Again, it seems practical, however, to accept Terborg, Maxwell, and Howard's earlier suggestion that the statistics involved be interpreted descriptively rather than inferentially.

Conclusion

It is hoped that this dialogue will be helpful in emphasizing the advantages and disadvantages of both the Terborg, Howard, and Maxwell and the Bedeian, Armenakis, and Gibson procedures. As has been pointed out, neither procedure is free from criticism. There is agreement that for those interested in the measurement of change, a careful weighing of the benefits and limitations of each procedure must be performed before a decision can be made as to which technique is preferred. Terborg, Maxwell, and Howard's interest, and the opportunity it has provided to comment further on the relevant issues involved, are appreciated.

References

Armenakis, A., & Zmud, R. Interpreting the measurement of change in organizational research. *Personnel Psychology,* 1979, 32, 709-723.

Armenakis, A., Bedeian, A. & Niebuhr, R. Planning for organizational intervention: The importance of existing socio-psychological situations in organizational diagnosis. *Group & Organization Studies,* 1979, 4, 59-70.

Bedeian, A., Armenakis, A., & Gibson, R. The measurement and control of beta change. *Academy of Management Review,* 1980a, 5, 561-566.

Bedeian, A., Armenakis, A., & Gibson, R. The measurement and control of beta change. *Proceedings of the Annual Meeting of the Academy of Management.* 1980b, 200-204.

Cherry, N., & Rodgers, B. Using a longitudianl study to assess the quality of retrospective data. In L. Moss & H. Goldstein (Eds.), *The recall method in social surveys.* London: University of London, Institute of Education, 1979, 31-47.

Golembiewski, R., Billingsley, K., & Yeager, S. Measuring change and persistence in human affairs: Types of change generated by OD designs. *Journal of Applied Behavioral Science,* 1976, 12, 133-157.

Green, G., & Wright, J. The retrospective approach to collecting baseline data. *Social Work Research & Abstracts,* 1979, 15 (3), 25-30.

Moss, L., & Goldstein, H. (Eds.). *The recall method in social surveys.* London: University of London, Institute of Education, 1979.

Rippey, R., Geller, L., & King, D. Retrospective pretesting in the cognitive domain. *Evaluation Quarterly,* 1978, 2, 481-491.

Terborg, J., Howard, G, & Maxwell, S. Evaluating planned organizational change: A method for assessing alpha, beta and gamma change. *Academy of Management Review,* 1980, 5, 109-121.

Achilles A. Armenakis is Associate Professor of Management and Director of the Auburn Technical Assistance Center, Auburn University.

Arthur G. Bedeian is Edward L. Lowder Professor of Management, Auburn University.