

Detecting Scale Recalibration in Survey Research

A LABORATORY INVESTIGATION

M. RONALD BUCKLEY
University of Oklahoma

ACHILLES A. ARMENAKIS
Auburn University

The accurate detection of scale recalibration is an issue that should not be ignored by organizational researchers and practitioners concerned with the accurate assessment of change interventions. In an effort to more effectively operationalize one methodology that is designed to detect scale recalibration (using ideal scales), three research questions were developed and investigated in a laboratory study utilizing videotape technology. It was concluded that ideal scale methodology can be used to accurately detect the presence of scale recalibration in survey research data. The implications of these findings for both researchers and practitioners are discussed.

Scale recalibration, also referred to as beta change (Golembiewski, Billingsley, & Yeager, 1976), occurs in temporal survey research when respondents (1) give different ratings across time for identical behavior or (2) give the same rating across time for different behaviors. In the first case, identical behavior at Time 1 and Time 2 may be rated a 2 and a 3, respectively. That is, a researcher may conclude that behavior has changed (based on the changes in the ratings) when, in fact, behavior has remained unchanged. In the second case, different behaviors at Time 1 and Time 2 may be rated

This research was partially funded by the Auburn Technical Assistance Center of Auburn University. Portions of this research were presented at the National Academy of Management meetings, 1985. Requests for reprints or further information should be sent to M. Ronald Buckley, Division of Management, University of Oklahoma, Norman, OK, 73019.

Group & Organization Studies, Vol. 12 No. 4, December 1987 464-481
© 1987 Sage Publications, Inc.

identically. That is, a researcher may conclude that behavior has not changed (based on the ratings) when, in fact, it has changed.

This concept, analogous to time-order error, has been the subject of research in psychophysics for over 50 years (Guilford & Park, 1931). For organizational researchers, the issue can be traced to the work of Walker, Shack, Egan, Sheridan, and Sheridan (1972). However, it was not until the Golembiewski et al. research that the implications of the issue in diagnosis and evaluation began to be appreciated.

Much of the data collected by organizational consultants are obtained through survey research methodology. If the data are used for diagnostic or evaluation purposes, one should be concerned that conclusions and recommendations are based upon appropriate data. Responses to survey questionnaires in which the scales have been recalibrated can result in misguided change efforts and erroneous conclusions regarding consultation programs. Because of the extensive use of survey methodology, research on the issue of scale recalibration is needed.

The assessment of scale recalibration has been addressed from a number of different perspectives. Randolph (1982) and Schmitt (1982) have proposed statistical approaches. Using a modified analysis of variance (ANOVA) approach, Randolph reasoned that the scale recalibration concept could be measured by examining variance shifts in data across time. Schmitt used LISREL in comparing variance-covariance matrices of data collected across time.

Terborg, Howard, and Maxwell (1980), van de Vliert, Stok, and Huisman (1985), and Zmud and Armenakis (1978) addressed the scale recalibration issue using three distinct procedural approaches. In a series of experiments, Terborg and his colleagues utilized retrospective measures in order to detect scale recalibration. In addition to the traditional pre- and postintervention measures, their procedure requires that after an intervention, respondents be asked to recall their preintervention level on the variable of interest—"how they now perceive themselves to have been just before the intervention was conducted" (p. 112). Scale recalibration has been operationalized as the difference between the preintervention measurement and the then-measure (which has theoretically been changed by the intervention). Terborg et al. have concluded that the retrospective design is an effective methodology in assessing improvements in *self-competencies* when they are the target of change programs.

Van de Vliert et al. (1985) addressed the issue by suggesting that criterion variables should be empirically and theoretically related to any longitudinal shifts on a variable of interest. In order to demonstrate this, a series of dynamic correlations was proposed that are designed to measure the relationship of changes on the criterion variable with changes in the focal variable.

The third procedural approach to assessing scale recalibration incorporates the use of scales to measure perceptions of ideal and actual behavior. By answering a series of questions regarding responses to ideal and actual scales, it is argued that scale recalibration can be identified (see Zmud & Armenakis, 1978). Scale recalibration has been operationalized as changes in ideal responses over repeated measures. This procedure was introduced into the organizational change literature by Zmud and Armenakis but, as will be described below, the logic for using the ideal concept was based on prior research and practice.

A major criticism of the ideal scale concept and the use of articulations of ideal behaviors has been suggested by Terborg et al. (1980) and Ito and Srinivas (1981). These researchers hypothesize that there is a ceiling effect that invalidates the use of these articulations in assessing scale recalibration. However, an important observation concerning this criticism is that it has been made in the absence of any empirical evidence to support such a claim. Therefore, one purpose of this research is to contribute to the literature surrounding the research on scale recalibration by reporting the findings of an empirical investigation of the ideal scale concept.

RELATED RESEARCH ON THE IDEAL CONCEPT

The concept of an idealized image can be traced to psychoanalytic theory, leadership research, and organization theory and practice. Horney (1945) posited that everyone possesses an idealized self-concept that is invariably the reference point against which our actual self-concept is viewed and judged. One creates an imagined picture of someone who is the *standard* in beauty, intelligence, kindness, goodness, and so forth. This becomes the kind of person one would like to see as himself or herself. According to Horney, "the idealized self . . . becomes a measuring rod with which to measure his actual being" (p. 110). Regarding the stability of this

idealized mental image, Havinghurst and MacDonald (1955) demonstrated that as one grows older, incorporating different elements or components (such as experience) into this idealized self-image, changes this idealized view.

Further, Rosenberg (1979) points out that although it is meaningful to distinguish between the actual-self and the ideal-self, an individual is not always successful in compartmentalizing these constructs. Therefore, actual and ideal cognitions are probably interdependent.

The implications of this research are that asking persons to articulate ideal ratings of self or others does not require talents that are foreign to an individual's cognitive-behavioral repertoire. Furthermore, over relatively short time periods, persons *not* exposed to organizational interventions could be expected to have relatively stable perceptions of idealized mental images. On the other hand, individuals who are exposed to some form of "effective" intervention may be expected to change their idealized mental image. The last implication of the above research is that actual and ideal perceptions are somewhat related so that one's idealized image may be affected by the existing conditions currently being experienced.

Leadership research that incorporated the ideal concept can be found in Fleishman (1953) and Hemphill, Siegel, and Westie (1951). Hemphill et al. developed the *Ideal LBDQ* that was modified by Fleishman and labeled the *Leadership Opinion Questionnaire*.

The research of Likert (1967) was the forerunner of the Survey of Organizations Questionnaire (SOO; Taylor & Bowers, 1972). The instructions of the Profile of Organizations (Likert, 1967) suggest that the profile can be administered to ascertain what respondents would *like* their employing organization to look like. The SOO incorporates the ideal scale on 21 of the items.

Based on this related research, one may presume that the ideal concept would be a logical option for assessing scale recalibration. As pointed out above, the criticism made against the use of the ideal scale has not been supported by empirical research. Therefore, the current research was designed to address issues pertinent to this criticism, as well as other important and related concerns.

RESEARCH QUESTIONS

The research questions and the research design for this study were formulated to address issues in the use of the ideal scale in assessing scale recalibration of responses to a self-report instrument completed by respondents, who were not, themselves, exposed to an intervention. In other words, the participants would be articulating their perceptions of the behavior of *others*, a common task expected of organizational members participating in personnel and organizational research studies. The rationale for this aspect of the design was to distinguish this research from that of Terborg and his colleagues (which focused on ratings of self-competencies). Furthermore, this rationale was intended to acknowledge the research of Havinghurst and MacDonald (1955) and Rosenberg (1979), cited above, by minimizing the problems associated with isolating the effects of interventions designed to bring about some type of change in *participants*.

Three major research questions were selected for investigation. All three are current to the issue of utilizing ideal scale methodology in order to assess organizational change accurately. First, there exists some doubt concerning the formation of an individual's ideal perceptions. To what extent are they dependent upon prevailing conditions? In other words, are articulations of ideal performance affected by the current circumstances experienced by respondents? This research issue has been formulated as Hypothesis 1: *Prevailing conditions do not affect the way in which respondents articulate their perceptions of ideal behavior.*

Second, there is a need to analyze articulations of ideal performance over a period of time and over different levels of observed performance. Do ideal perceptions change over time? And, most important, do changes in the performance level of the behavior that is observed have an effect upon articulations of ideal performance? This research issue has been formulated as Hypothesis 2: *Although susceptible to slight changes over time, ideal perceptions form a relatively stable construct that does not change over time.*

Third, there has been some controversy in the literature concerning ideal perceptions. One argument is that ideal perceptions are contaminated by a ceiling effect and are therefore unusable in assessing scale recalibration. In other words do articulations of ideal performance, which are collected during a pretest, cluster around

the maximum scale point with minimal variability? And, further, does there exist any upward scale space in which these same articulations can increase when collected during a posttest? This research issue has been formulated as Hypothesis 3: *A ceiling effect does not contaminate and, therefore, invalidate the use of ideal perceptions in organizational change research.*

METHOD

To determine the impact of conditions affecting temporal survey responses, it is important to control variations in stimuli that are to be rated. Simply stated, without knowing whether or not a target stimulus has or has not changed, a researcher cannot offer definite explanations of observed response variations. For example, in a field study, a researcher cannot manipulate the amount or direction of behavioral change that occurs in an organization. Furthermore, in a field setting, behavioral changes can be attributed to (1) changes in focal stimuli, (2) changes in the raters observing focal stimuli, (3) changes in the instrument used to measure the focal stimuli over time, and (4) a combination of the above. To achieve the necessary control in our study, laboratory methodology and videotaped vignettes of performance were utilized. The use of videotape technology enabled the controlled manipulation of the amount, as well as the direction, of behavioral change observed by the respondents. The vignettes employed were selected from eight scripts, each depicting a personnel manager interviewing a disgruntled employee (Borman, Hough, & Dunnette, 1976). Respondents were asked to evaluate seven dimensions of interviewer performance as displayed on different vignettes following a predetermined scheme. The dimensions that served as the dependent variables were as follows: (1) interview structure and control, (2) establishment and maintenance of rapport, (3) reaction to stress, (4) information acquisition, (5) conflict resolution, (6) subordinate development, and (7) subordinate motivation. Similar research methodology has been developed and used in recent research investigations concerned with survey research measurement issues (Armenakis, Buckley, & Bedeian, 1986).

A 43-item survey instrument, *Dimensions of Interviewer Performance* (DIP), was developed from the manual accompanying the vignettes (Borman et al., 1976). Each item was to be evaluated using

a 5-point Likert scale. In order to minimize artificial inflation of reliability estimates, item placement was systematically varied throughout the questionnaire across all seven dimensions of interviewer performance. As indicated, respondents were requested to evaluate the extent to which a tape-specific activity was performed by the interviewer (personnel manager) in question.

PROCEDURE

Undergraduate students enrolled in introductory courses in psychology, biology, business statistics, and organizational behavior were offered an opportunity to participate in this research study for extra course credit. These particular courses were deemed an appropriate source for participants, given that none was concerned with interviewing. All students were informed of the study at the beginning of the school term. They were told that they would be expected to view a videotape twice during the term and respond to a series of corresponding questions. Sessions were estimated to last about 30 minutes.

During the data collection sessions, a booklet containing (1) a consent form, (2) a brief vignette introduction, and (3) the survey instrument was distributed to all participants. Participants were then instructed to observe a video monitor on which the research directions, taped earlier, were played. This assured uniformity of directions across all of the experimental groups.

The videotape instructed the participants to read the consent form to themselves as it was played on the videotape. After signing the consent form, respondents were directed to read the vignette introduction silently as it was played on the videotape. The vignette of the manager interviewing a subordinate followed.

For our study, low-, average-, and high-performance vignettes were selected as stimuli. Judgments of interviewer performance were obtained from the manual accompanying the tape and were based on performance scores obtained from expert judges (see Borman et al., 1976) across the seven dimensions of performance. An overall performance score was obtained by summing the expert judgments over the seven dimensions of performance for each of the eight vignettes. The low-performance vignette had received a rating of 13.39, the average-performance vignette had received a rating of 25.43, and the high-performance vignette had received a rating of 37.46.

After viewing one of the three selected vignettes, participants were directed by the videotape to form in their minds an image of a manager who would have performed the job of the interviewer in a "perfect or model manner." They were then further instructed to complete the Ideal DIP survey based on the ideal interviewer they had just conceptualized.

After completing the survey instrument based on the conceptualized ideal interviewer, participants were directed to complete the second survey instrument, that is, the Actual DIP, based on the performance of the interviewer that they had actually observed on the videotape. After the participants completed the second survey, the initial experimental session ended.

Ratings of an ideal interviewer were requested before the ratings of the interviewer that had been observed by the participants, based on research by Mitroff and Kilmann (1975) reporting that respondents can more easily assess their organization's characteristics when such ratings are preceded by an opportunity to conceptualize ideal organizational characteristics. Mitroff and Kilmann attributed this to the notion that conceptualizing ideal performance gave the respondents a frame of reference against which to rate the quality of the observed performance.

The second experimental session was held 3 weeks later. The procedure during the second experimental session was an exact reproduction of the procedure during the first experimental session except that specified participants viewed different vignettes.

RESEARCH GROUPS

A total of 270 subjects, composing nine groups, participated in the study. Each of the nine groups contained 30 subjects. All possible combinations of the three selected vignettes were randomly distributed among the research groups. For example, at Time 1 Groups 1, 5, and 9 viewed the low-, average-, and high-performance vignettes, respectively. At Time 2 each group viewed the vignette viewed previously. The remaining six groups viewed the remaining combinations of vignettes, that is, Group 2 viewed the low-performance vignette at Time 1 and the average-performance vignette at Time 2; Group 3 viewed the low-performance vignette at Time 1 and the high-performance vignette at Time 2. This systematic assignment of vignettes was made for all the groups (see Table 1).

TABLE 1
 Group Numbers Assigned to Combinations of
 Interviewer Performance Observed at
 Time 1 and Time 2

		<i>Time 2</i>		
		<i>Low</i>	<i>Average</i>	<i>High</i>
Time 1	high	7	8	9
	average	4	5	6
	low	1	2	3

RESULTS

Seven items common to both the Actual and Ideal DIP questionnaires were eliminated because their deletion improved the coefficient alphas of the remaining items (Cronbach, 1951). For both administrations at Time 1 and Time 2, the alphas for the Actual DIP ranged from .74 to .94, whereas the alphas for the Ideal DIP ranged from .54 to .74 (see Table 2). These reliability estimates were considered acceptable.

Hypothesis 1 stated that prevailing conditions do not affect the way in which respondents articulate their perceptions of ideal behavior. In order to test this hypothesis, ideal responses at Time 1 were grouped as follows: All respondents who observed a low-performance vignette (Groups 1, 2, and 3) were combined into a single Group A; all respondents who observed an average-performance vignette (Groups 4, 5, and 6) were combined into a single Group B; and all respondents who observed a high-performance vignette (Groups 7, 8, and 9) were combined into a single Group C. A Group \times Composite Dimension multivariate analysis of variance (MANOVA; 3×7) was performed on Time 1 *ideal* interviewer responses. The MANOVA was significant, $F(14, 524) = 8.45, p < .001$, thus indicating that there were differences between the groups over the seven composite dimensions. A Bonferroni test on each of the composite dimensions revealed that Group C (those who observed the high-performance vignette) consistently articulated ideal manager responses that were significantly higher than those given by Groups A and B. In conjunction with this test, a similarly formulated Group \times Composite Dimension MANOVA (3×7) was performed on Time 1 *actual* performance responses. Similarly, the

TABLE 2
 Dimensions of Interviewer Performance,
 Number of Items per Dimension, and
 Internal Consistency Measures for
 Actual (A) and Ideal (I) Responses
 at Time 1 (T1) and Time 2 (T2)

<i>Dimension</i>	<i>Number of Items</i>	<i>T1</i>		<i>T2</i>	
		<i>A</i>	<i>I</i>	<i>A</i>	<i>I</i>
Structure	6	.85	.58	.83	.69
Rapport	6	.92	.54	.92	.64
Stress	3	.91	.57	.93	.69
Motivate	6	.93	.71	.94	.72
Information acquisition	6	.92	.64	.91	.72
Conflict	4	.74	.60	.85	.66
Develop	5	.89	.69	.91	.74

MANOVA was significant, $F(14, 524) = 50.03, p < .001$, thus indicating that there were significant differences between the groups over the seven composite dimensions. A Bonferroni test on each of the composite dimensions indicated that Group C responses were significantly higher than Group B responses, which were significantly higher than Group A responses, on all seven composite dimensions. Apparently, respondents perceived the differences in actual performance and these differences in actual performance significantly affected their perceptions of how an ideal manager would have performed the focal tasks. Therefore, ideal perceptions are formed based upon (and are dependent upon) the prevailing conditions that are being experienced by the respondents. Hypothesis 1 was not supported.

Hypothesis 2 stated that although susceptible to slight changes over time, ideal perceptions form a relatively stable construct that does not change over time. In order to test Hypothesis 2, a series of multivariate analyses of covariance (MANCOVA) were performed. This analysis was chosen due to the assertion by Huck and McLean (1975) that the use of a repeated measures ANOVA is less powerful and more confusing than analysis of covariance in most situations. Following the suggestions of Huck and McLean, the responses that were obtained at Time 1 were covaried out of the responses that were obtained at Time 2, which were then compared between groups. In order to perform this analysis, the reported respondent

articulations of the *ideal* interviewer performance were split into three categories. The first category comprised the three groups of respondents who observed the low-performance vignette at Time 1 (Groups 1, 2, and 3) and was designated MANCOVA-LOWIDEAL. The second category comprised the three groups of respondents who observed the average-performance vignette at Time 1 (Groups 4, 5, and 6) and was designated MANCOVA-AVERAGEIDEAL. The third category comprised the three groups of respondents who observed the high-performance vignette at Time 1 (Groups 7, 8, and 9) and was designated MANCOVA-HIGHIDEAL.

The results of the computations were as follows:

- (1) MANCOVA-LOWIDEAL: $F(2, 609) = .91, p > .05$
- (2) MANCOVA-AVERAGEIDEAL: $F(2, 609) = .02, p > .05$
- (3) MANCOVA-HIGHIDEAL: $F(2, 609) = 2.35, p > .05$

These analyses indicate that over the three categories, there is no significant change in ideal manager perceptions over time. Therefore, Hypothesis 2 was supported. In order to further investigate this issue of scale recalibration additional analyses were necessary. A series of multivariate analyses of covariance (MANCOVA) were performed on the reported judgments of *actual* interviewer performance that were grouped into three categories. The first category comprised the three groups of respondents who observed the low-performance vignette at Time 1 (Groups 1, 2, and 3) and was designated MANCOVA-LOWACTUAL. The second category comprised the three groups of respondents who observed the average-performance vignette at Time 1 (Groups 4, 5, and 6) and was designated MANCOVA-AVERAGEACTUAL. The third category comprised the three groups of respondents who observed the high-performance vignette at Time 1 (Groups 7, 8, and 9) and was designated MANCOVA-HIGHACTUAL.

The results of the computations were as follows:

- (1) MANCOVA-LOWACTUAL: $F(2, 609) = 36.17, p < .05$
- (2) MANCOVA-AVERAGEACTUAL: $F(2, 609) = 97.84, p < .05$
- (3) MANCOVA-HIGHACTUAL: $F(2, 609) = 103.44, p < .05$

These analyses indicate that the actual perceptions within all three categories were significantly different at Time 2 after the corresponding Time 1 responses were covaried out. This finding provides evidence that the respondents were able to *correctly* articulate

differences in interviewer performance level. That is, they correctly articulated that interviewer performance had changed—either improved or worsened—in the respective vignettes. It is noteworthy that a series of dependent t-tests on the responses between Time 1 and Time 2 for the three groups that viewed the same performance vignette at both time periods revealed no significant differences between these ratings. In other words, this finding provides evidence that the respondents did not experience any scale recalibration because interviewer behavior across time remained constant. So, in summary, ideal articulations did not significantly change over time and the level of ideal responses at Time 1 depended, to a relatively large extent, upon the performance level that was initially observed, at Time 1, by the respondent.

Hypothesis 3 stated that a ceiling effect does not contaminate and, therefore, invalidate the use of ideal perceptions in organizational change research. A severe restriction of range would invalidate statistical analyses on the obtained data. A ceiling effect is a relative term and there has, to date, not been adequate criteria developed to assess the existence of ceiling effects. They must, therefore, be assessed in an indirect fashion. The presence of a ceiling effect can be indirectly assessed from two different perspectives. One perspective is to analyze the standard deviation, range, and skewness in *absolute* terms in order to determine whether these statistics are within acceptable limits. The other perspective is a *relative* analysis that is accomplished by comparing the standard deviation, range, and skewness for the actual responses with those of the ideal responses that were obtained at the same time period. These statistics are contained in Table 3.

If a ceiling effect existed, the standard deviation and range of responses would be near zero and the skewness would be high. These statistics for the ideal responses at Time 1 are as follows: standard deviation varies from .40 to .65, a moderate standard deviation, which indicates a moderate amount of variability (Chou, 1969); range varies from 1.8 to 3.5; and skewness varies from .11 to .96, a small degree of skewness (see Chou, 1969, pp. 108-109). The same statistics for the ideal responses at Time 2 are as follows: standard deviation varies from .41 to .64, a moderate standard deviation; range varies from 1.7 to 3.0; and skewness varies from .01 to 1.03, a small degree of skewness (see Table 3).

A comparison of the standard deviation, range, and skewness for

TABLE 3
Standard Deviation, Range, and Skewness
at Time 1 and Time 2 for
Actual (A) and Ideal (I) Perceptions
($n = 270$)

<i>Dimension</i>	<i>Standard Deviation</i>		<i>Range</i>		<i>Skewness</i>	
	<i>A</i>	<i>I</i>	<i>A</i>	<i>I</i>	<i>A</i>	<i>I</i>
	Time 1					
Structure	.81	.46	3.5	2.7	-.18	-.74
Rapport	.90	.40	3.8	2.0	-.42	-.72
Stress	1.09	.52	4.0	2.7	-.87	-.96
Motivate	1.11	.56	4.0	2.7	-.02	-.34
Information acquisition	.76	.43	3.8	1.8	-.32	-.63
Conflict	.90	.65	4.0	3.5	.21	-.11
Develop	.98	.59	4.0	3.0	-.16	-.26
	Time 2					
Structure	.84	.43	3.7	1.8	-.34	-.50
Rapport	.98	.41	4.0	1.7	-.48	-.93
Stress	1.21	.50	4.0	2.3	-.63	-1.03
Motivate	1.13	.57	4.0	2.3	-.18	-.26
Information acquisition	.86	.44	4.0	2.0	-.46	-1.01
Conflict	1.00	.64	4.0	3.0	.11	-.01
Develop	1.08	.59	4.0	2.8	-.11	-.19

the actual and ideal scales reveals that the ideal responses are relatively more restricted than the actual responses. However, this finding is to be expected because of the nature of ideal perceptions. Respondents were requested to articulate their perceptions of how the ideal interviewer that they had conceptualized would have performed. Their perceptions should have approximated the "perfect, or model, interviewer" and, thus, should be somewhat negatively skewed. In spite of this, it can be concluded that there is significant upward scale space in which it is possible for these ideal articulations to increase during a future repeated measure of data collection. Nevertheless, it is significant that these statistics are not more extreme.

To determine the effect of these restrictions on the use of the ideal scale, Pearson product-moment correlations (r_s) were com-

puted between the Time 1 and Time 2 responses, within each group, to the ideal scales. In order for these restrictions to be considered severe, these r s would all be equal to, or very close to, 1.00. About 20% of the correlations were below .40. Approximately 80% varied between .40 and .80, a moderate to high range. Therefore, the use of the ideal scale does not appear to be severely limited by a ceiling effect and the commensurate restriction of range problems that typically accompany ceiling effects. Therefore, Hypothesis 3 was supported.

DISCUSSION

The issue of scale recalibration is important to organizational consultants who diagnose the need for change and evaluate interventions using survey research methodology. The potential problems arising from misguiding change programs and from making erroneous inferences from recalibrated data are sufficient to warrant the necessary support for research on this topic. The use of ideal scale is one alternative from a set of procedural options, composed of the retrospective procedure (Terborg et al., 1980) and the criterion approach (van de Vliert et al., 1985). Organizational consultants must make informed decisions regarding the various methods for detecting scale recalibration when designing consultation efforts. This research has contributed to the knowledge base necessary for consultants to make these decisions.

Previous investigations of the scale recalibration issue have been conducted primarily in field settings. Therefore, the degree of control exercised by investigators has been limited. Although this research was conducted in a laboratory, the research scenario was similar to most organizational change assessment strategies, that is, respondents experienced no intervention but were expected to observe and assess the behavior of a manager whose behavior, in six of the nine research groups, had either improved or worsened over a period of time.

The first research question was concerned with the formation of ideal perceptions and whether they were dependent upon or independent of prevailing conditions. The preceding analysis revealed that, in general, respondents accurately perceived actual performance levels and that their perceptions of actual performance affected their perceptions of how an ideal manager would have

performed the focal tasks. This finding, in a laboratory experiment, is consistent with that of Armenakis et al. (1979), which was conducted in a field setting. Those who observed a high-performance vignette articulated conceptualizations of an ideal manager that were significantly higher than those who observed an average-performance vignette and those who observed a low-performance vignette. Prevailing conditions affect the way in which respondents articulate their perceptions of ideal behavior.

The nature of ideal perceptions is probably closer to those notions suggested by Rosenberg (1979). Although it may be useful to distinguish between these two concepts (actual versus ideal perceptions), individual respondents do not successfully compartmentalize these constructs. Actual perceptions and ideal perceptions are not independent of each other and, as has been shown, there is probably a direct relationship between the level of actual performance observed and the level of ideal perceptions.

The second research question was concerned with analyzing articulations of ideal perceptions. Of primary interest was whether these perceptions changed over time or over different levels of observed performance. Over time, although respondents observed different vignettes of performance, actual perceptions were significantly different at Time 2 after Time 1 responses were covaried out. Ideal perceptions, even in groups in which significant differences were detected in articulations of actual managerial performance at Time 2 after the corresponding Time 1 responses were covaried out, were found to be stable and relatively unchanged over time.

The collective findings concerning these two research questions yield an interesting implication for the measurement of change using ideal scales. The implication is that those who assess the effectiveness of change interventions in organizations will probably articulate an ideal concept that is dependent upon their existing knowledge of the organization in question. Once this ideal is formed, it is relatively stable over time. In order to ensure that ideal conceptualizations of behavior (which, according to Mitroff and Kilmann, 1975, are used as a basis for comparing differences in actual performance) are similar across the respondents, organizations would be well advised to expose respondents, perhaps via some "training," to all factors related to the potential change that they are requested to assess. Therefore, respondents could be expected to form ideal concepts that approach similarity. Then,

presumably, changes would be assessed from the same reference point, over all assessors.

The third research question was concerned with the notion of whether ceiling effects sufficiently restricted the range of the obtained responses. The analysis revealed that although ideal perceptions were slightly skewed, the use of ideal scales does not appear to be severely limited by a ceiling effect and the commensurate restriction of range problems.

The major implication of this research is that ideal scales can be used by individuals who have experienced no intervention but are expected to assess the behavior of those people who have been exposed to an intervention. Apparently, respondents are fairly adept at articulating ideal performance. Further, this adeptness is relatively stable across the time period employed in this investigation. Apparently, requesting respondents to articulate their perceptions of ideal behavior and requiring that they observe a videotaped vignette of performance was not responsible for creating ideal perceptions. Obviously, the respondents' ideal cognitions existed before this research study and will continue to exist and be adjusted after this research study (Havinghurst & MacDonald, 1955).

Finally, the findings from the analysis of responses to actual scales served as a way to verify these ideal scale findings. That is, the manipulation of three levels of interviewer performance (i.e., low, average, and high) at Time 1 and Time 2 required the respondents to assess nine combinations of interviewer performance. The fact that these respondents *correctly* assessed actual behavior, in all nine combinations, suggests that ideal scales were effective in correctly demonstrating that scale recalibration did not occur. In fact, in this study, significant changes in articulations of actual performance resulted in negligible changes in ideal conceptualizations of performance. That is, ratings of ideal behavior at Time 1 closely approximated ratings of ideal behavior at Time 2. In other words, a researcher using ideal scales would not be led into incorrectly concluding that scale recalibration had occurred. It is therefore recommended that researchers who are engaged in the assessment of behavioral interventions also assess the presence of scale recalibration in their obtained responses using ideal scale methodology. Failure to do so may result in inaccurate assessments of intervention effectiveness.

REFERENCES

- Armenakis, A. A., Bedeian, A., & Niebuhr, R. (1979). Planning for organizational intervention: The importance of existing sociopsychological situations in organizational interventions. *Group & Organization Studies*, 4, 59-70.
- Armenakis, A. A., Buckley, M. R., & Bedeian, A. G. (1986). Survey research measurement issues in evaluating change: A laboratory investigation. *Applied Psychological Measurement*, 10, 147-157.
- Borman, W., Hough, L., & Dunnette, M. (1976). *Performance rating: An investigation of reliability, accuracy, and relationships between individual differences and rater error* (Vol. 1). Minneapolis: Personnel Decisions, Inc.
- Chou, Y. (1969). *Statistical analysis with business and economic applications*. New York: Holt, Rinehart & Winston.
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Fleishman, E. (1953). The measurement of leadership attitudes in industry. *Journal of Applied Psychology*, 37, 153-158.
- Golembiewski, R., Billingsley, K., & Yeager, S. (1976). Measuring change and persistence in human affairs: Types of change generated by OD designs. *Journal of Applied Behavioral Science*, 12, 133-157.
- Guilford, J., & Park, D. (1931). The effect of interpolated weights upon comparative judgments. *American Journal of Psychology*, 43, 589-599.
- Havinghurst, R., & MacDonald, D. (1955). Development of the ideal self in New Zealand and American children. *Journal Educational Research*, 49, 263-273.
- Hemphill, J., Siegel, A., & Westie, C. (1951). *An exploratory study of relations between perceptions of leader behavior, group characteristics, and expectations concerning the behavior of ideal leaders*. Unpublished manuscript, Ohio State University, Personnel Research Board, Columbus.
- Horney, K. (1945). *Our inner conflicts*. New York: Norton.
- Huck, S. W., & McLean, R. A. (1975). Using a repeated measures ANOVA to analyze the data from a pretest-posttest design: A potentially confusing task. *Psychological Bulletin*, 82, 511-518.
- Ito, J., & Srinivas, K. (1981). The use of impact and self-impact correlations in causal analysis. *Academy of Management Review*, 6, 301-308.
- Likert, R. (1967). *The human organization*. New York: McGraw-Hill.
- Mitroff, I., & Kilmann, R. (1975). Stories managers tell: A new tool for organizational problem solving. *Management Review*, 64, 18-28.
- Randolph, W. (1982). Planned organizational change and its measurement. *Personnel Psychology*, 35, 117-139.
- Rosenberg, S. (1979). New approaches to the analysis of personal constructs in person perception. *Nebraska Symposium on Motivation* (pp. 179-242). Lincoln: University of Nebraska Press.
- Schmitt, N. (1982). The use and analysis of variance and covariance structures to assess beta and gamma change. *Multivariate Behavioral Research*, 17, 343-358.
- Taylor, J., & Bowers, D. (1972). *Survey of organizations*. Ann Arbor: Center for Research on Utilization of Scientific Knowledge, Institute for Social Research, University of Michigan.

- Terborg, J., Howard, G., & Maxwell, S. (1980). Evaluating planned organizational change: A method for assessing alpha, beta, and gamma change. *Academy of Management Review*, 5, 109-121.
- van de Vliert, E., Stok, J., & Huismans, J. (1985). The criterion approach to unraveling beta and alpha change. *Academy of Management Review*, 10, 269-274.
- Walker, R., Shack, J., Egan, G., Sheridan, J., & Sheridan, E. (1972). Change in self-disclosure after group experience. *Journal of Applied Behavioral Science*, 8, 248-251.
- Zmud, R., & Armenakis, A. (1978). Understanding the measurement of change. *Academy of Management Review*, 3, 661-669.

M. Ronald Buckley is Assistant Professor of Management at the University of Oklahoma. He received his Ph.D. from Auburn University in industrial/organizational psychology. His primary areas of interest are the accurate assessment of behavioral interventions, construct validation, the employment interview, performance appraisal in organizations, and subordinate perceptions of the feedback process.

Achilles A. Armenakis is the Associate Dean for External Affairs in the College of Business at Auburn University. He received his D.B.A. from Mississippi State University. He has published extensively in the areas of the accurate assessment of behavioral interventions, management consulting, using research methodologies in organizations, microorganizational issues, and organization development.