EFFECTS OF A PERSUASIVE MESSAGE UPON ATTITUDES: A METHODOLOGICAL COMPARISON OF AN OFFSET BEFORE-AFTER DESIGN WITH A PRETEST-POSTTEST DESIGN

WILLIAM D. BROOKS

BACKGROUND

It was in response to the need to control extraneous variables that different experimental designs have evolved. The most acceptable design for many years has been the traditional *pretest–posttest with control group* design, which controls the main or primary effects of history (events occurring during the time span from the pretest to the posttest), maturation (effects systematic with the passage of time such as growing older, hungrier, or more tired) and testing. However, in recent years this design has come under increasing attack because it has a serious weakness when used with certain measuring instruments. That weakness is the interaction effect of the pretest with the experimental variable. When measuring instruments of a kind which sensitize the subject in one way or another toward the experimental variable are used in the pretest, the effect operates to contaminate the experimental results.

In contrast to the main effects of testing, history, etc., which manifest themselves in mean differences independently of the presence of other variables, interaction effects are joint effects, and may occur even when no main effects are present. Campbell has said: "Applied to the testing variable, the interaction effect might involve not a shift due solely or directly to the measurement process, but rather a sensitization of respondents to the experimental variable so that when X was preceded by O there would be a change, whereas both X and O would be without effect of occurring alone." [2 p. 302] It should be noted that interaction effects are not limited to the creation of positive effects, i.e., of *producing* a change, but the sensitization may also be of such a nature as to *prevent* a change that would have occurred otherwise. Hovland reported a finding in his Army studies in which the pretest reduced the effects of the experimental variable because of the elicitation of a commitment by the

subject to a given position. [5] Similarly, Solomon found that a spelling pretest of children reduced the effects of a training period. [6] In like manner, when direct attitude scales are used, apparently the verbalization of an attitude or of commitment to a position creates a tendency for subjects to defend that commitment or to move more slowly from it than if the verbalization or commitment had not been made. Campbell has said that in opinion and attitude research our well-developed interview and attitude tests must be rated as highly reactive. [2 p. 299] Brembeck and Howell report the results of experimentation and observation concerning the reluctance of subjects to *admit* to altering a previously expressed opinion. They say, "After our opinions are registered publicly, we are reluctant to change them. Then, too, the expression of an opinion tends to establish the attitude more clearly and firmly in our minds rendering it less subject to change." [1 p. 102] Thus, the use of a direct attitude scale as a measuring instrument in the pretest is both reactive and conducive to the creation of an interaction effect.

The problem of reactive measuring instruments could be solved in either of two ways: (1) by using non-reactive measuring instruments, or (2) by eliminating the pretest so that the interaction with the experimental variable is prevented. Attempts have been made to develop subjective and indirect attitude measurement instruments, but often such instruments are not usable in the "live" audience or field-experiment situation. Despite the demonstration of the pretest weakness in the traditional design by Star and Hughes, [7] Hovland, [5] Crespi, [3] and Solomon, [6] many of our experimental studies are still conducted within the context of the traditional design while using measuring instruments which sensitize the subject and create interaction effects. The following studies desired to test a newer design as a possible solution to the problem.

Experiment # 1

Procedure

This study desired to compare the offset before–after design

$$\begin{bmatrix} O_1 & & \\ & X & O_2 \end{bmatrix}$$

to the traditional pretest–posttest design to discover whether or not the offset before–after design could control the variables or primary effects as well as does the traditional design, and whether or not the linear type attitude scale, when used in a pretest and a posttest on the same individuals, was a reactive measuring instrument. The hypothese were: (1) that the before–after design can control the primary effects of extraneous variables as well as does the traditional design; and (2) that the offset before–after design will allow for more accurate measurement of the persuasive effect of a speech by controlling the secondary effect of a reactive measuring instrument.

The design worked out for this experiment was

$$\begin{bmatrix} O_1 & X & O_2 \\ O_3 & & \\ & X & O_4 \end{bmatrix}$$

in which O_1 was the pretest of one group, X was the speech, and O_2 was the posttest of that same group. O_3 was the pretest for a second group, a control group. O_4 was the posttest for yet a third group that did not receive a pretest, but simply heard the speech and took the voting attitude posttest. Thus, the design contained the traditional design in $\begin{bmatrix} O_1 & X & O_2 \\ O_3 & & (O_3) \end{bmatrix}$ and the offset before–after design in $\begin{bmatrix} O_3 & & \\ & X & O_4 \end{bmatrix}$ so that the two designs could be compared. This necessitated three groups:

$\begin{bmatrix} O_1 & X & O_2 \\ O_3 & & \\ & X & O_4 \end{bmatrix}$ designated group C,
designated group B, and
designated group A.

Group C received and marked a direct linear type attitude scale having seven alternatives ranging from strongly disagree at one extreme to strongly agree at the other extreme; listened to a persuasive speech approximately 10 minutes in length, played on a tape recorder, and proposing that the voting age be lowered to 18 years; and again received and marked the attitude scale. Group B simply received and marked the attitude scale. They did not take the posttest, but it was assumed

that their posttest scores would not differ from the pretest, since the time lapse was only 12 minutes and one can assume the attitudes would not change by maturation in that short a time. Group A listened to the speech and then marked the attitude scale.

English Language 101 classes at the University of Colorado were used for the experiment. The population with which we were concerned were college freshmen. These students were required to take the course and were assigned at random to a section of the course. It was our intention, thereby to have a random sampling. The *t*-test was used to test for the significance of mean differences. The *t*-test formula used may be found in Edwards. [4]

There were approximately 20 students in each class. Eighteen classes selected at random were used, six for group A, six for group B, and six for group C. Three sections from each of the hours (8:00, 9:00, 10:00, 11:00, 1:10, and 3:10) were used so that all three treatments could be given during each hour. The order of the treatments varied from hour to hour (during the 8:00 hour the order was A, B, C; 9:00 order was B, C, A; 10:00 order was C, A, B, etc.). Of the 372 subjects, 188 were females and 184 were males; 194 were 18 years of age or younger, 111 were 19 years of age, and 67 were 20 years of age or older.

Results

The first hypothesis was that both designs could control the primary effects. Subjects in O_2 (posttest of the traditional design) were divided on the basis of sex. The mean voting attitude for the males was 3.80 and the for the females it was 3.83. Similarly, subjects in O_4 (posttest of the offset before–after design) were divided on the basis of sex. The mean for the males was 4.98 while the mean for the females was 4.57. Both designs showed that the difference between the means was not significant. Thus, both designs showed that sex was not a factor in this experiment insofar as attitude change was concerned.

Similar analyses showed that in both designs neither age nor time of day had any effect on the influence of the speech as regards voting attitude.

The second hypothesis was that there would be no difference between the two designs in their ability to control the reactivity of the measuring instrument. The mean of O_1 was compared to the mean of O_3 and there was no significant difference, a result that was anticipated in view of the random assignment of subjects to classes and classes to experimental groups. We are justified in assuming a similar initial attitude for Group A, since subjects were assigned at the same time in the same manner to all three groups.

We may then conclude that the two groups did not differ in initial attitude. Both groups heard exactly the same speech on the same tape recorder. However, when the posttest mean attitude of the traditional design subjects was compared to the pretest mean there was no difference. According to results obtained by this design, the speech had not changed the subjects' attitude toward 18 year olds voting. On the other hand, comparison of the posttest mean, O_4 with the pretest mean, O_3, revealed a significant change in attitude: according to results obtained by the offset design, the speech had changed the subjects' attitudes significantly. Moreover, when the posttest means of the two designs were compared (O_2 and O_4), they were found to differ by more than one full step on the seven point scale—a difference significant at the 0.01 level.

Conclusions

1. According to these data, the offset before–after design controlled the primary effects of extraneous variables, such as sex, age, and time of day as well as did the traditional design. Of course, the assumption of random sampling is necessary to the conclusion that the offset design controlled for these primary effects, but the same assumption must also be satisfied for the traditional design.

2. The direct attitude scale used in this experiment was a reactive measuring instrument. Campbell [2 p. 299] and Crespi [3 p. 99] have so indicated, and this study agrees with their conclusions concerning the reactivity of the measuring instrument. The attitude scale used in this study was shown to be reactive in that the offset before–after subjects changed attitude significantly, ($t = 6.83$) while the pretest–posttest subjects did

TABLE 1
Means and Standard Deviations of Experiment #1

	Pretest	Posttest
GROUP C	$\overline{X}_{O_1} = 3.18$ $\sigma = 2.06$	$\overline{X}_{O_2} = 3.56$ $\sigma = 2.03$
GROUP B	$\overline{X}_{O_3} = 3.13$ $\sigma = 1.86$	
GROUP A		$\overline{X}_{O_4} = 4.77$ $\sigma = 1.74$

not change attitude ($t = 1.52$). After the speech, the difference between the two groups was significant at the 0.01 level of confidence ($t = 5.04$).

3. The offset before–after design allowed for a more sensitive assessment of the subjects' changes of attitude than did the traditional pretest–posttest design; i.e., it registered a change in attitude that was not detected by the traditional design. In the traditional design the subjects' attitudes (the phenomena being measured) appear to have been affected by the measuring instrument.

EXPERIMENT # 2

In order to re-test the hypotheses studied in 1960 and reported as Experiment # 1 in this paper, a second experiment was conducted in 1963 at McPherson College, McPherson, Kansas. The same speech that had been used in Experiment # 1 was used for three Fundamentals of Speech classes. Students were assigned to the classes at random. The three classes were designated Class A, Class B, and Class C. Class A took the attitude test in a pretest, listened to the speech from the tape recorder and took the attitude test again. Class B took the pretest but did not listen to the speech. Class C listened to the speech and then took the attitude test.

The results and interpretation of the data were essentially the same as for the first experiment. The pretest means of Classes A and B were compared by means of the *t*-test and there was no significant difference ($t = 0.20$). However, the posttest

TABLE 2
Means and Standard Deviations of Experiment #1 for Sex, Age, and Time of Day

		Males	Females
SEX	O_2	$\bar{X} = 3.80$ $\sigma = 1.94$	$\bar{X} = 3.83$ $\sigma = 1.99$
	O_4	$\bar{X} = 4.98$ $\sigma = 1.75$	$\bar{X} = 4.57$ $\sigma = 1.72$
		18 year olds	20+ year olds
AGE	O_2	$\bar{X} = 4.03$ $\sigma = 1.86$	$\bar{X} = 3.81$ $\sigma = 2.27$
	O_4	$\bar{X} = 4.75$ $\sigma = 1.84$	$\bar{X} = 4.30$ $\sigma = 1.89$
		8:00	3:00
TIME	O_2	$\bar{X} = 3.95$ $\sigma = 2.18$	$\bar{X} = 4.05$ $\sigma = 2.12$
	O_4	$\bar{X} = 4.64$ $\sigma = 1.70$	$\bar{X} = 4.94$ $\sigma = 1.44$

means of Classes A and C were significantly different ($t = 2.73$) at the 0.05 level of confidence. Therefore, the conclusions drawn from the first experiment were substantiated in the second study.

SUMMARY

It is often contended that negative results are as valuable as positive results, but this is true only when the experimenter has some reasonable assurance that his statement "I measured no difference" is equal to the statement "the experimental variable caused no difference." If the pretest–posttest design reduces the likelihood of producing a difference, its value is correspondingly reduced, leading to an important advantage of the offset before–after design.

There is further harm in using the traditional design with reactive measuring instruments in that it tends to stop or shut off research. This is, perhaps, the more serious of the two errors that can be made in hypothesis testing.

TABLE 3
Means and Standard Deviations of Experiment #2

	Pretest	Posttest
	$\bar{X}_{O_1}$	$\bar{X}_{O_2}$
GROUP A	$\bar{X} = 3.30$	$\bar{X} = 3.63$
	$\sigma = 1.48$	$\sigma = 1.50$
	$\bar{X}_{O_3}$	
GROUP B	$\bar{X} = 3.21$	
	$\sigma = 1.50$	
		$\bar{X}_{O_4}$
GROUP C		$\bar{X} = 4.94$
		$\sigma = 1.51$

One should also remember that the effect of reactive measuring instruments is not always a negative effect, but with some instruments upon some independent variables (such as I.Q. tests in measuring intelligence) the reactive effect might act in a positive manner to exaggerate a difference or even to create a difference when in fact none existed.

It would appear that the weaknesses of the traditional design when used with reactive instruments are serious enough to warrant careful consideration by social science experimenters.

One alternative solution to the problem of the interactive effect of testing is to perfect a non-reactive test or measuring instrument. However, even if a totally non-reactive attitude test could be perfected it would take time and expenditure of effort that are in many instances prohibitive to its use. For example, the announcement by press or other media that a speech of importance will be given at a certain place and time does not always allow the necessary time to perfect and test a non-reactive measuring instrument. Furthermore, the typical "live audience" speech situation is one in which it is difficult for an experimenter to control the subjects so as to guarantee pre- and posttests on the same individuals. Finally, the live audience situation is one that prevents the use of a set of cards for a Q-sort or the use of a sheet of paper with numerous lengthy statements to be read and evaluated. In short, indirect tests seem difficult to use in certain field situations. One can use a direct attitude scale instrument with confidence by resorting to the offset before–after design.

REFERENCES

1. Brembeck, Winston Lamont, and William Shirley Howell. PERSUASION: A FORM OF SOCIAL CONTROL. New York: Prentice-Hall, 1952.
2. Campbell, Donald T. "Factors Relevant to the Validity of Experiments in Social Settings," PSYCHOLOGICAL BULLETIN, 54 (July, 1954) 297–312.
3. Crespi, L. P. "The Interview Effect on Polling," PUBLIC OPINION QUARTERLY, 12 (1948) 99–111.
4. Edwards, Allen L. STATISTICAL ANALYSIS. New York: Rinehart and Company, Inc., 1946, 177.
5. Hovland, Carl I., A. A. Lumsdaine, and F. D. Sheffield. EXPERIMENTS IN MASS COMMUNICATION. Princeton: Princeton University Press, 1949.
6. Solomon, R. L. "An Extension of Control Group Design," PSYCHOLOGICAL BULLETIN, 46 (1949) 137–150.
7. Star, S. A. and H. M. Hughes. "Report on an Educational Campaign: the Cincinnati Plan for the United Nations," AMERICAN JOURNAL OF SOCIOLOGY 55 (1945–1950) 389.