Implications for Educational Research of the Phenomenon of Experimenter Bias

C. MITCHELL DAYTON

A RECENT series of experiments has opened up to formal study a possible source of serious bias in behavioral science research. The implications of these investigations are of special importance in education in view of the fact that educational research often involves the use of one or more human experimenters who are aware of the probable outcomes of the research project. Specifically, it has been demonstrated that the outcomes of experiments can be dramatically affected by expectations held by the experimenter and that these effects operate despite relatively highly standardized experimental procedures and data gathering techniques.

We have long been aware of the possible biasing effects which participation in an experiment may have on human subjects (the so-called "Hawthorne effect"). Educational researchers are also trained to be aware of certain other conditions which may lead to invalid research conclusions (e.g., bias introduced by non-respondent groups in questionnaire studies). The purpose of this paper is to review some exemplary studies on experimenter bias and to consider their implications for educational research.

Studies on Experimenter Bias

Experimenter bias may be viewed as a construct which was developed in order to describe the effects of experimenter expectations upon the outcomes of experiments. Operationally speaking, differential instructions given to experimenters concerning the probable outcome of an experiment which each will conduct, have a predictable relationship to the actual outcome of the experiment, even though the experimental techniques, materials, etc., are the same for all experimenters. It is at present a moot point whether the mechanism of experimenter bias operates at a conscious or an unconscious level.

Before reviewing any studies, a note on terminology will prove useful. The term "experimenter" refers to individuals who were under study by the investigator reporting the research; "subjects" refers to individuals (or animals) used by the experimenter. Thus, these studies involve experiments on experimenters. The first study reviewed here differs from the remaining ones in that experimental subjects were subhuman.

Cardaro and Ison (1963) used 17 undergraduate psychology students as experimenters and 34 planaria as subjects. The experimenters were familiar with classical conditioning terminology and procedures. Three groups of experimenters were formed. Group I, the high response expectancy group, was told that the planaria with which they were to work had been previously adversely conditioned (by electric shock) to a light stimulus. Group II, the low response expectancy group, was told that the planaria were naive (i.e., had had no previous light/shock conditioning trials). Group III was informed that half of the group of planaria with which they were to work was naive while the other half had received previous light/shock conditioning. *In reality,* all planaria used by the experimenters had received 50 light/shock training trials.

Each experimenter ran 2 planaria simultaneously in 2 blocks of 50 trials each on light/shock conditioning. On each trial, the experimenter recorded whether: (a) no conditioned response occurred; (b) a head-turn response occurred; or (c) a marked body contraction occurred. The table presents the average proportions of responses (b) and (c) recorded by experimenters in each of the groups. Group III data are recorded separately for high and low response expectancy trials.

Highly significant results were obtained from analysis of the 4x2x2 factorial design (where trials were blocked into quarters). The trial blocks comparisons, however, showed non-significant effects. The stronger effect was obtained for the head-turn response (47.4% reported by Group I versus 9.9% reported by Group II). In Group III the contrast resulting from differences in response expectancy is less than in Groups I and II, but is, nevertheless unmistakable.

Results similar to those reported by Cardaro and Ison were obtained with humans as subjects in a series of 3 investigations by Rosenthal and Fode (1963).

(I) Ten undergraduate psychology students served as experimenters and a total of 206 students served as subjects. A "high bias" group of 5 experimenters and a "low bias" group of 5 experimenters were formed. The task carried out by the experimenters was to obtain ratings from the subjects for 10 stimulus cards. Each stimulus card contained a photograph of a face and the subjects were to rate each card on a 21 point success-failure scale. The high bias group of experimenters was told that from previous studies an average rating of +5 was to be expected. The low bias group was told to expect −5. *In reality,* previous scaling of the cards with

Group	*Response Expectancy*	*(b) Head-turns*	*(c) Contractions*
I	High	.474	.180
II	Low	.099	.009
III	High	.300	.154
	Low	.154	.049

undergraduate subjects had yielded an average rating of approximately 0. As additional motivation, the experimenters were promised $1 an hour for conducting the studies, but were told that if they did a good job they would receive $2 an hour.

The procedure followed by the experimenters was relatively structured. All directions were *read* from prepared instruction sheets. The results of this study showed a pronounced positive bias effect (high bias group mean was 4.05) but only a slight effect from negative bias (low bias group mean was —0.95).

(II) The experimenters were 24 advanced undergraduate psychology students and the subjects were 180 introductory psychology students. The task carried out by the experimenters was the same as described above. Four groups of experimenters were formed: Group I was given a —5 bias as in the first study; Group II was given a +5 bias but these experimenters were behind a screen, out of sight of the subjects when they read instructions; Group III was given a +5 bias but distributed printed directions to subjects and did not communicate verbally with subjects; Group IV was given a +5 bias as in the first study. The group mean ratings are summarized below:

Group	*Mean Rating*
I	.524
II	1.315
III	.489
IV	2.117

Analysis of variance indicated that only Groups I and III *did not differ* significantly.

(III) The experimenters were 12

graduate students and subjects were 58 introductory psychology students. The study was a 2x2 factorial arrangement with one dimension defined as "ego-involvement" ($5 an hour for an experimenter versus $2 an hour for an experimenter) and the second dimension involved paying subjects 50¢ for participating versus not paying subjects. Prior to the study, each experimenter predicted the mean performance of his group. From the two-way analysis of variance, no main effect or interaction effect was significant. However, the experimenter-predicted mean ratings correlated .56 with obtained group means.

In overviewing these three studies, it appears that experimenter bias can be manipulated when subjects are human (Study I); verbal cues without visual cues are sufficient to mediate the mechanisms of experimenter bias (Group II, Study II), however, the effect is attenuated relative to the operation of both visual and verbal cues (compare Groups II and IV, Study II); the range of ego-involvement studied did not show this to be a differentiating factor (Study III). It should be noted that experimenters in Groups II and IV, Study II used identical words with their subjects. This implies that other aspects of the verbal communication (e.g., intonation) are critical in mediating experimenter bias. The enhanced effect in Group V, Study II may be attributed to the operation of both verbal and visual (e.g., gestural) cues.

Discussion

It should be noted that the studies of experimenter bias cannot be classified as examples of errors of observation. While in the first study reviewed (Cardaro and Ison, 1963) the experimenters' failures to observe responses in an accurate manner may be loosely called errors of observation, this classification is not consistent with the apparent functional relationships with direction of induced bias. Errors of observation are assumed to obey a random process, whereas the phenomenon of experimenter bias is decidedly non-random. Similarly, experimenter bias in the studies with human subjects cannot be attributed simply to experimenter dishonesty induced by the promised monetary reward. In these studies experimenters could not tamper with subjects' responses and the only verbal communication between experimenter and subjects was the experimenter reading printed instructions.

It is apparent that the phenomenon of experimenter bias must be taken largely at face value; that is, the experimenter's preconceived notions concerning the outcomes of the experiment operate to bias the experiment in an expected direction. The mechanism by which experimenter bias operates can be partly understood on the basis of available evidence. Evidently, relatively subtle verbal and gestural mannerisms on the part of the experimenter can induce bias in college level subjects. The cues include, most likely, differential emphases and intonation while reading instructions, accompanied by postural and other physical movements which, somehow, alert subjects to the experimenter's biases.[1]

The implications of these findings, and similar results from a number of studies not reported here, for research in educational settings are all too clear. In many educational projects, especially thesis studies carried out by graduate students, the principal experimenter carries out most details of his study himself. Since these studies typically involve human subjects and since most experimenters have expectations concerning their experimental outcomes, the conclusion that experimenter bias *may* be a prime explanation for much of the material in our research literature can hardly be avoided.

[1] The possibility of a second-order experimenter bias effect operating in these experiments can be (at least facetiously) entertained. That is, if experimenter bias is really a potent effect it seems possible that the authors of the studies induced bias (other than that experimentally planned) in their experimenters which led to the expected evidence for experimenter bias. If we pursue this argument, however, we are assuming what we are challenging, namely the operation of experimenter bias.

Replication of an experimental outcome cannot be considered as an adequate antidote for experimenter bias. During the replication study the experimenter is aware of the outcome of the experiment he is repeating. Hence, this situation may be even more susceptible to experimenter bias than when an original study is conducted. That is, experimenter expectations with respect to an original study may well be of a lesser degree of certainty than during a replication study. If degree of experimenter bias is related to the felt certainty of obtaining a given experimental outcome, then a strong bias can be expected in replication studies.

Suggested Areas for Research on Experimenter Bias

Of the many possible directions which research on experimenter bias might take, the most basic questions concern the mechanism by which this phenomenon operates. Manipulation of instructional and related contingencies and closer recording of experimenters' behavior during interaction with subjects seem to be necessary steps in this direction.

While a start has been made toward identifying personality variables which covary with the degree of experimenter bias, this area should be pursued in more detail by including a wider variety and greater complexity of personality scales. Similarly, studies of subject characteristics which are associated with receptivity to experimenter bias would prove fruitful.

From the point of view of educational research, the major need is to systematically investigate the operation of experimenter bias in educational settings. The studies reviewed in this paper were under psychological laboratory conditions. The extent to which these laboratory findings will generalize to less controlled experimental situations is a question amenable to research. It seems both feasible and profitable to perform essential replications of the laboratory experiments with classroom groups.

In summary, the purpose of this paper was to review some exemplary research on the topic of experimenter bias and to present some implications of these studies for educational research. Based on this review, it appears that experimenter bias is a phenomenon which may be exercising an adverse effect on the validity of a great deal of research in education.

References

L. Cardaro and J. R. Ison. "Psychology of the Scientist: X. Observer Bias in Classical Conditioning of the Planarian." *Psychological Reports* 13: 787-89; 1963.

R. Rosenthal and K. L. Fode. "Psychology of the Scientist: V. Three Experiments in Experimenter Bias." *Psychological Reports* 12: 491-511; 1963.

—C. Mitchell Dayton, *Assistant Professor of Education, University of Maryland, College Park.*