International Journal of Orofacial Myology and Myofunctional Therapy
Official Journal of the International Association of Orofacial Myology
50th ANNIVERSARY

Volume 4 | Number 2 | pp. 6-14 1978

Tutorial

Experimental designs used in assessing the effectiveness of social programs: A review

Monte D. Smith *(George Peabody College for Teachers)*

Suggested Citation
Smith, M. D. (1978). Experimental designs used in assessing the effectiveness of social programs: A review. *International Journal of Oral Myology, 4(2)*, 6-14.
DOI: https://doi.org/10.52010/ijom.1978.4.2.1

The International Journal of ORAL MYOLOGY
Volume 4, Number 2 April, 1978

Experimental Designs Used in Assessing the Effectiveness of Social Programs:

A Review

Monte D. Smith, Ph.D.
George Peabody College for Teachers

The focus in this review is on pre-experimental, quasi-experimental and "true" experimental designs specifically cited in the literature as suitable for implementation in social action program research endeavors. It will be noted that a number of the designs reviewed also appear in Campbell and Stanley (1966), but no effort is made to enumerate all sixteen of Campbell and Stanley's designs. The designs which are discussed will be divided into classifications of pre-experimental designs, "true" experimental designs, non-equivalent control group designs, and time-series and multiple time-series designs.

The notational system used by Campbell and Stanley will be employed in this review:

"An X will represent the exposure of a group to an experimental variable or event, the effects of which are to be measured; O will refer to some process of observation or measurement; the X's and O's in a given row are applied to the same specific persons. The left-to-right dimension indicates the temporal order, and X's and O's vertical to one another are simultaneous" (p. 6).

The letter R will indicate that subjects were randomly assigned to treatment and comparison groups, while dotted lines separating parallel rows indicate comparison groups were not equated through randomization.

Pre-experimental designs

Although Suchman (1967) cited the Pretest-Posttest Control Group Design with randomized assignment to treatment and control conditions (Campbell & Stanley's [1966] design #4) as the ideal experimental design for evaluation research in social action programs, he did admit (1967, 1972) that a much more frequently employed design was the One-Shot Case Study (Campbell and Stanley's design #1):

$$X \quad O$$

Suchman (1972) referred to the One-Shot Case Study as an "adaptation" of the Pretest-Posttest Control Group Design with random assignment. Another "adaptation" frequently used, Suchman (1972) stated, is the "Survey Design" which appears to be the Static-Group Comparison Design (design #3 in Campbell and Stanley [1966]):

$$\begin{array}{cc} X & O_1 \\ \hdashline & O_2 \end{array}$$

The three pre-experimental designs in Campbell and Stanley (1966) were categorized by Cook and Campbell (Note 1) as "generally uninterpretable". The One-Shot Case Study design was judged incapable of ruling out any threats to internal validity. The One-Group Pretest-Posttest Design was judged only slightly better:

$$O_1 \quad X \quad O_2$$

Campbell and Stanley's (1966) third pre-experimental design, the Static-Group Comparison Design, was renamed by Cook and Campbell the "Posttest-Only Design with Non-Equivalent Groups". The design was judged uninterpretable because posttest differences may be attributable to either treatment effects or to selection differences between the non-equivalent groups. The three Campbell and Stanley (1966) pre-experimental designs also were cited by Wholey, Scanlan, Duffy, Fukomoto, & Vogt (1970) as appropriate for use in social action program evaluations.

"True" experimental designs

The Pretest-Posttest Control Group Design with random assignment to treatment and non-treatment conditions was cited by Suchman (1967, 1972) as the ideal for program evaluation research:

$$\begin{array}{cccc} R & O_1 & X & O_2 \\ R & O_3 & X & O_4 \end{array}$$

Campbell (1957), however, emphasized that the Posttest-Only Control Group Design with random assignment is often preferable in situations where pretesting may be potentailly reactive:

$$\begin{array}{ccc} R & X & O_1 \\ R & X & O_2 \end{array}$$

Nunnally (1975) recently reaffirmed the merit of the Posttest-Only design, calling it the "workhorse" of evaluation research. The posttest-only design is, nevertheless, susceptible to bias in situations where differential experimental mortality may occur. This susceptibility arises because it is not possible to check on the comparability of participants lost from treatment versus non-treatment groups.

Wholey, et al. (1970) included the Solomon Four Group design (Campbell & Stanley, 1966) in their enumeration of designs suited to social action program evaluations:

$$\begin{array}{cccc} R & O_1 & X & O_2 \\ R & O_3 & X & O_4 \\ R & O_2 & X & O_5 \\ R & O_3 & X & O_6 \end{array}$$

Suchman (1967) suggested the following design for testing planned variations in social action programs:

$$\begin{array}{cccc} R & O_1 & X & O_2 \\ R & O_3 & X & O_4 \\ R & O_5 & X & O_6 \\ R & O_7 & X & O_8 \end{array}$$

Although this design appears to be formally indentical to the Pretest-Posttest Control Group Design (Kirk's [1968] completely randomized design), Wholey, et al. (1970) listed the two designs separately and dignified Suchman's formulation with the appellation the Comparison of Alternative Program Strategies Design. The same design is referred to by Huck, Cormier, and Bounds (1974) merely as an extension of the basic Pretest-Posttest Control Group Design.

Non-equivalent control group designs

Suchman (1972) cited the "Panel Design", which is Campbell and Stanley's (1966) Non-Equivalent Control Group Design, as probably the best approximation to the "true" experimental design realistically attainable in evaluation research. This design may be depicted:

$$\begin{array}{ccc} O & X & O \\ \hdashline O & & O \end{array}$$

The dashed line separating the two rows of O's indicates that the groups were not equated, probabalistically, by random assignment prior to treatment.

Cook and Campbell (Note 1) renamed this design the Untreated Control Group Design with Pretest and Posttest, which was classified as a "generally interpretable non-equivalent control group design". Cook and Campbell recommended the design in situations where nothing better can be used, pointing out that its interpretability depends on the particular empirical outcome obtained in a given study. If the alternative hypotheses of selection-maturation, instrumentation (in situations of pretest group mean non-equivalence where one or both means approach an end of the scale), and local history can be ruled implausible alternative explanations, then this design may be considered strong in terms of internal validity.

Cook and Campbell (Note 1) illustrated a special version of the non-equivalent control group design with pre and post measures:

$$\begin{array}{ccc} O_{A1} & X & O_{B2} \\ \hdashline O_{A1} & & O_{B2} \end{array}$$

The A and B indicate non-equivalent measures. Typically, this design is used as a last resort in *ex post facto* situations where the investigator feels compelled to locate a pretest measure that correlates (within groups) with posttest scores. The difficulty in obtaining robust $O_{A1} \bullet O_{B2}$ correlations is a major drawback of this design. Cook and Campbell noted that attempts to equate groups by matching on O_{A1} will only lead to regression artifacts.

Cook and Campbell (Note 1) cited the following design, the Non-Equivalent Dependent Variables Design, as one of the weakest generally interpretable quasi-experimental designs:

$$\begin{array}{ccc} O_{1A} & X & O_{2A} \\ \hdashline O_{1B} & & O_{2B} \end{array}$$

Campbell and Stanley's (1966) notational system is not appropriate for depicting what Cook and Campbell had in mind with this design. The notation above appears to depict two groups, only one of which receives the treatment. These two groups, it would appear, are measured on "non-equivalent dependent variables". In fact, these are not the circumstances which the notational depiction was intended to convey. Instead, there is only one group of participants involved. This single group is tested on two variables (A and B), only one of which is expected to change as a function of introduction of the X. Of course, there could be multiple (Type A) variables expected to change (i.e., reflect the impact of X), and multiple (Type B) variables expected not to change. Extreme caution and theoretical clarity would be necessary in selecting the variables. Differential changes on the two variables would be expected as a function of the independent variable manipulation. Both variables, however, would have to respond similarly to plausible alternative influences other than the experimental manipulation.

Cook and Campbell's (Note 1) Reversed Treatment Non-Equivalent Control Group Design with Pretest and Posttest:

$$\begin{array}{ccc|l} O_1 & X+ & O_2 & \text{(Group X)} \\ \hdashline O_1 & X- & O_2 & \text{(Group Y)} \end{array}$$

represents the utilization of two different treatments. Treatment X+ is expected to produce an effect in one direction, while treatment X− is expected to produce an effect in the conceptually opposite direction (Refer to Figure 1). Although not discussed by Cook and Campbell, it would seem that the addition of another measurement point (O_3) would permit the determination of the relative temporal stability of any obtained effects, should that be of interest. The addition of this feature would, however, transform the design into a variation of the Multiple Time-Series Design.

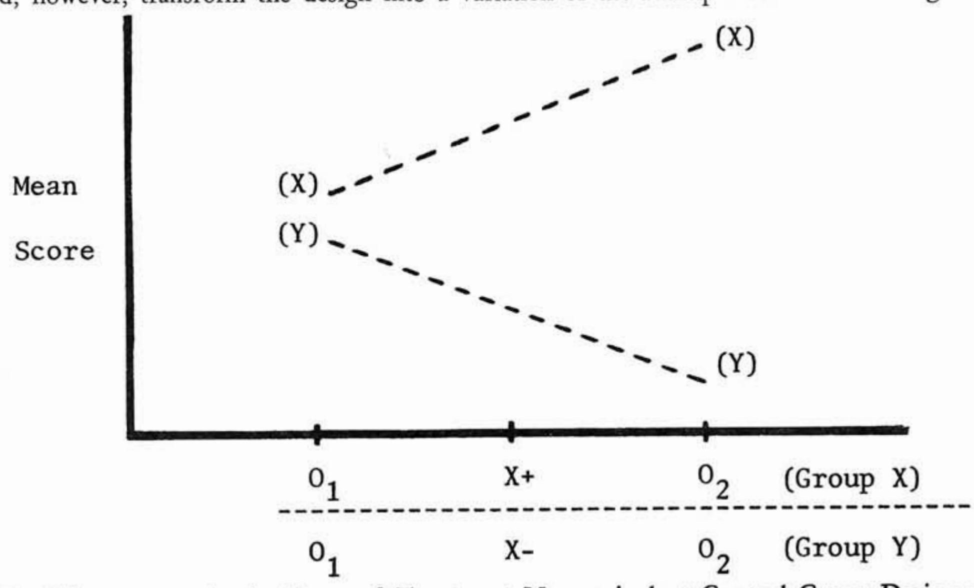

Figure 1. Possible outcome in the Reversed Treatment Nonequivalent Control Group Design with Pretest and Posttest.

Time-series and multiple time-series designs

The addition of a third observation point to Cook and Campbell's Reversed Treatment Non-Equivalent Control Group Design with Pretest and Posttest would produce this variation of the Multiple Time-Series Design:

O_1 X+ O_2 O_3 (Group X)
- - - - - - - - - - - - - - - -
O_1 X− O_2 O_3 (Group Y)

Figure 2 depicts a pattern of results indicating that obtained effects are temporally stable, and a pattern of transient treatment effects is depicted in Figure 3. A reversal of the conceptually opposite treatments subsequent to $Time_3$ should provide an efficient replication in situations where the latter results are obtained. Figure 4 graphically depicts this possibility.

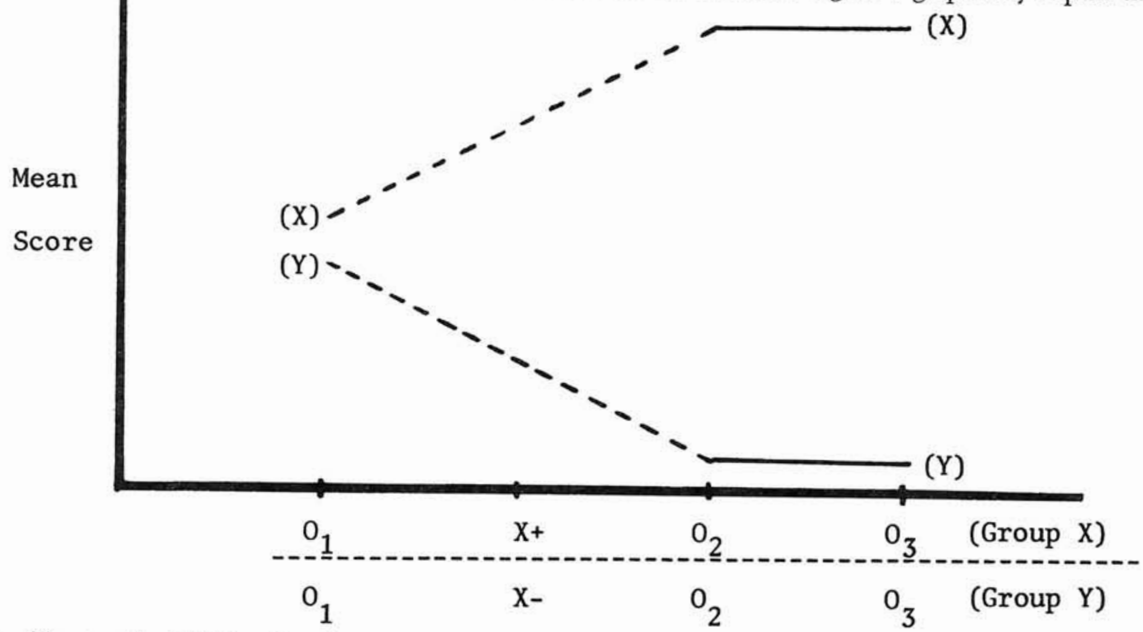

Figure 2. Temporal stability of opposite-direction results obtained with a variation of the Reversed Treatment Nonequivalent Control Group Design with Pretest and Posttest.

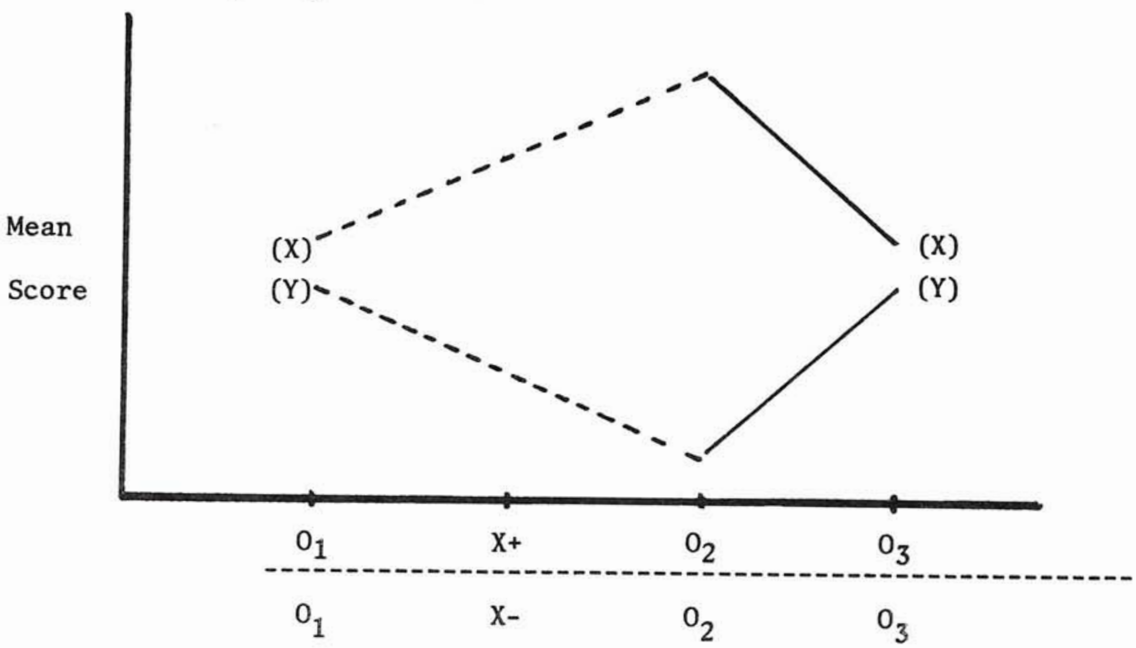

Figure 3. Transient treatment effects with a variation of the Reversed Treatment Nonequivalent Control Group Design with Pretest and Posttest.

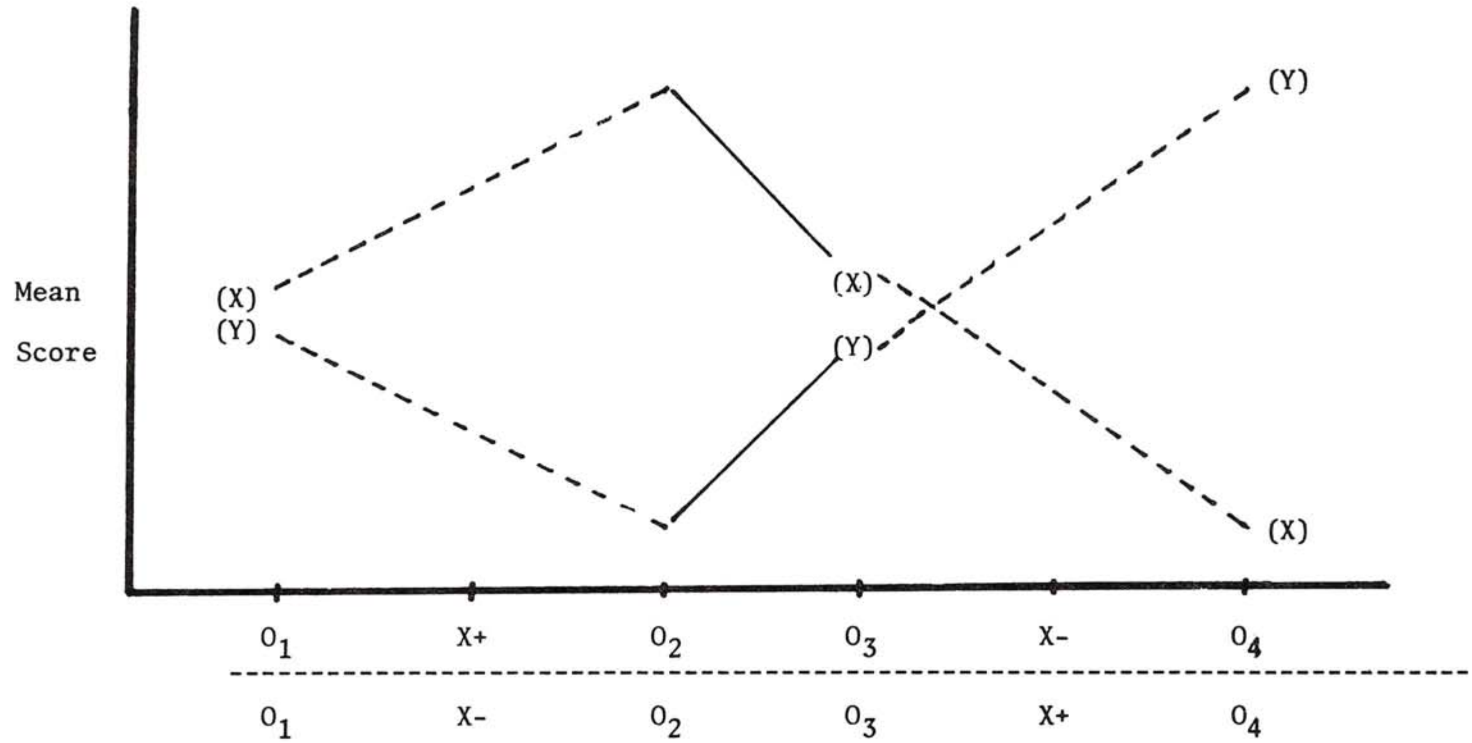

Figure 4. Switched treatment replication of transient treatment effects in a variation of the Reversed Treatment Nonequivalent Control Group Design with Pretest and Posttest.

Campbell (Note 2) predicted increased utilization of the Time-Series Design in assessing social action program effects, especially as work on analysis proceeds. The Time-Series Design (Campbell & Stanley, 1966) is depicted:

$$O_1 \quad O_2 \quad O_3 \quad O_4 \quad X \quad O_5 \quad O_6 \quad O_7 \quad O_8$$

The multiple time-series design gains in certainty of interpretation by adding a second, control group. "The experimental effect is in a sense twice demonstrated, once against the control and once against the pre X values in its own series". (Campbell & Stanley, 1966, p. 55)

$$\begin{array}{ccccccccc} O_1 & O_2 & O_3 & O_4 & X & O_1 & O_2 & O_3 & O_4 \\ \hdashline O_1 & O_2 & O_3 & O_4 & X & O_1 & O_2 & O_3 & O_4 \end{array}$$

The multiple time-series was judged by Campbell and Stanley (1966) as probably the best of the quasi-experimental designs. It controls for all nine threats to internal validity enumerated by Campbell and Stanley.

Campbell (1975) and Ross, Campbell and Glass (1970) demonstrated how the time-series and multiple time-series designs could be utilized to assess the impact of legal reforms. In both cases the legal reform was a highway vehicular speeding crackdown and the dependent variable was the traffic fatality rate. Both studies, one in Connecticut and the other in the United Kingdom, entailed abrupt, well-publicized reforms. In analyzing the Connecticut speeding crackdown, Campbell (1975) ingeniously monitored unintended side effects and used these measures as adjuncts in interpreting the dependent variable of primary interest, the traffic fatality rate.

Neither the Connecticut speeding crackdown (Campbell, 1975) nor the British "Breathalyzer" study (Ross, Campbell & Glass, 1970) properly qualify as field experiments, since the investigator clearly was not responsible for the independent variable manipulations (Kerlinger, 1964, p. 382; French, 1953, p. 101; Stanley, 1967) nor, apparently, was he "instrumental in effecting the manipulation" (Katz, 1953, p. 95). The two speeding crackdown studies are more properly classified as "natural experiments", a type of field study where "a social change takes place without any action by the social researcher. It just happens to be an interesting change for him to measure." (Katz, 1953, p. 95). The two studies are instructive, however, in that the procedures involved are precisely those that would be employed in cases where the social change of interest was engineered by the investigator.

It should also be noted, moreover, that the definition of experiment advanced by Cook and Campbell (Note 1) does not include the stipulation that the investigator must be instrumental in effecting the manipulation:

> "By 'experiment' we mean any experimenter-controlled or naturally-occurring event (a treatment) which intervenes in the lives of respondents so that its consequences can be empirically assessed and a cause-effect relationship corroborated." (p. 1)

Cook and Campbell distinguished between experimental and quasi-experimental designs on the basis of whether treatment and control groups were formed by assigning respondents to treatments in a random or non-random fashion. "The former designs are called *'true' experimental* and the latter *'quasi-experimental'.*" (p. 1, italics original) Cook and Campbell thus considerably played down the importance of whether or not the investigator controls the independent variable manipulation.

Cook and Campbell (Note 1) discussed two variations of the multiple time-series design. The first of these, called the "Multiple Time-Series with Non-Equivalent Dependent Variables" appears to be a time-series extension of the Non-Equivalent Dependent Variables design:

$$\begin{array}{ccccccc} O_{1A} & O_{2A} & O_{3A} & X & O_{4A} & O_{5A} & O_{6A} \\ O_{1B} & O_{2B} & O_{3B} & & O_{4B} & O_{5B} & O_{6B} \end{array}$$

Apparently, there is only one group involved. In the example given by Cook and Campbell, the British Breathalyzer Crackdown (Ross, Campbell, & Glass, 1970), one behavior (O_A : serious traffic accidents while the pubs are open) would be expected to show an effect due to the treatment X (the Crackdown). The other dependent variable (O_B : serious accidents during commuting hours when pubs are closed) would not, however, be expected to show an effect.

Cook and Campbell's (Note 1) second variation on the multiple time-series, the Interrupted Time-Series with Switching Replications Design, uses two non-equivalent groups which receive the same treatment, but at different points in time. Group A serves as a control when group B received the treatment, and vice versa:

$$\begin{array}{cccccccccccccc} O_1 & O_2 & O_3 & O_4 & & O_5 & O_6 & O_7 & O_8 & X & O_9 & O_{10} & O_{11} & O_{12} \\ \hdashline O_1 & O_2 & O_3 & O_4 & X & O_5 & O_6 & O_7 & O_8 & & O_9 & O_{10} & O_{11} & O_{12} \end{array}$$

Two of the designs which Cook and Campbell (Note 1) classified as "generally interpretable non-equivalent control group designs" actually do not incorporate control groups. Since these designs were suggested for use in situations where the investigator has access to only one research population, and since they bear much similarity to the basic time-series design, they are listed in this section.

The Removed Treatment Design with Pre and Post is one of those two designs. Only one group is involved. X represents the treatment and X is the removal of the treatment:

$$O_1 \;\; X \;\; O_2 \;\; \Big| \;\; O_3 \;\; X \;\; O_4$$

An effective treatment would be indicated using this design in cases where the (O_1 - O_2) difference is opposite in direction to the (O_3 - O_4) difference. (See Figure 5)

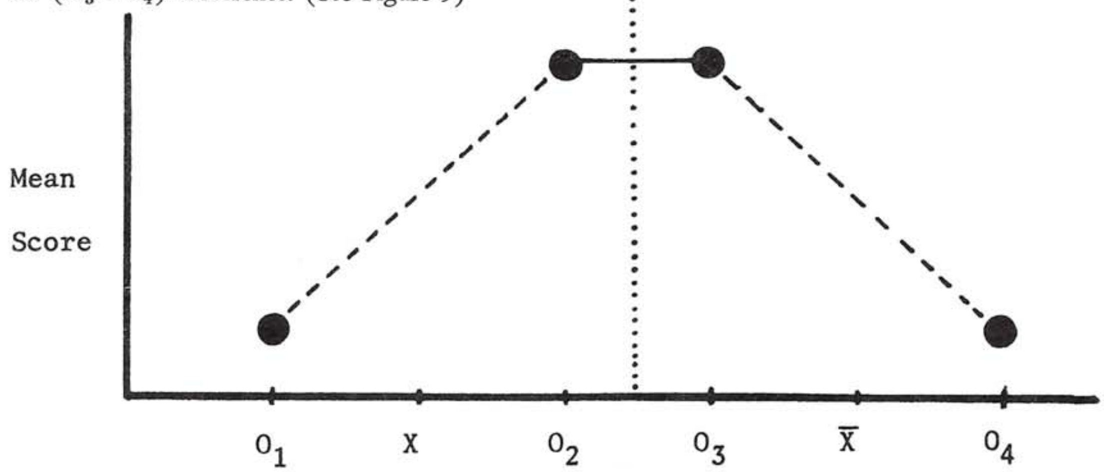

Figure 5. Successful treatment outcome as assessed by the Removed Treatments Design with Pre and Post.

An implicit assumption in this design is that changes induced by the treatment do not persist after treatment is withdrawn.

The second time-series variation design is the Repeated Treatment Design. A baseline measure is taken, the treatment is introduced, faded out, and then reintroduced while additional measurements are taken:

$$O_1 \quad X \quad O_2 \quad O_3 \quad X \quad O_4$$

As with the Removed Treatments Design with Pre and Post, this design assumes that the effect of X, the treatment, is transient. That is, X is assumed to exert an effect only when it is present. (See Figure 6) The Repeated Treatment Design is similar to Campbell and Stanley's (1966) Equivalent Time Samples Design,

$$X_1O \qquad X_0O \qquad X_1O \qquad X_0O \quad \text{etc.}$$

where the treatment (X_1) is alternatively introduced and removed (X_0).

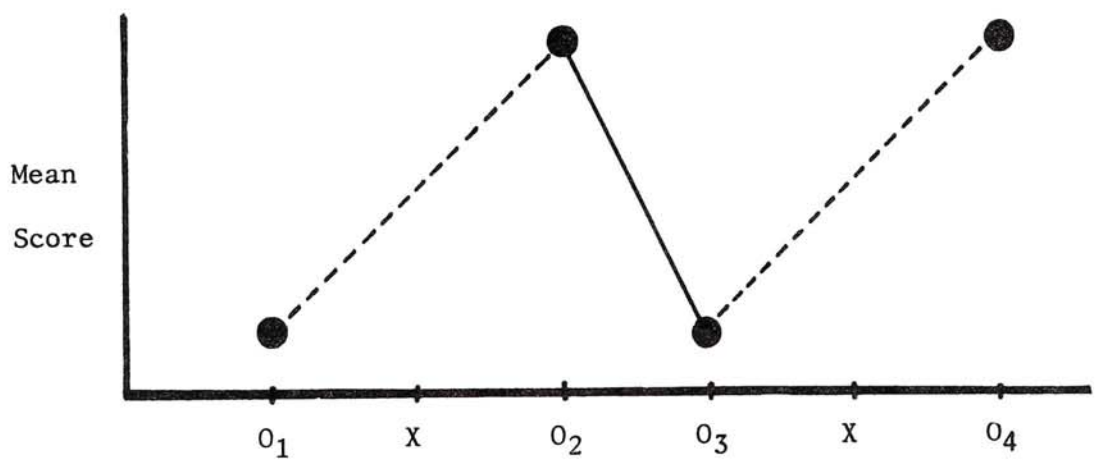

Figure 6. Possible outcome with the Repeated Treatment Design.

Stanley (1965) suggested a specific quasi-experimental design strategy for assessing the effects of educational innovations. The suggested design is a variation of the multiple time-series (Campbell & Stanley, 1966):

Previous Year		Current Year			
O_1	O_2	O_3	X	O_4	(Group A)
		O_5	X	O_6	(Group B)
O_7	O_8	O_9	X	O_{10}	(Group C)

requiring the collection of measures over parts of at least two academic years in this instance. The groups represented by the top two rows are drawn from the same school, although only one group receives the innovative treatment. The third row of observations is taken from a group in another, control school. $(O_2 - O_1) - (O_4 - O_3)$ would indicate the effect of history within the experimental school. This could be compared with $(O_8 - O_7) - (O_{10} - O_9)$ to determine if the innovative approach affected non-involved students in the experimental school. O_3 should equal O_5 and O_5 should equal O_9. The basic comparison is $(O_6 - O_5) - (O_4 - O_3)$. If the innovative treatment was effective, and non-treated subjects within the experimental school were not "contaminated", an expected pattern of results might look like those graphically depicted in Figure 7. If group A, the non-treated group in the experimental school, were "contaminated" by the successful innovation, they would be expected to fall somewhere between points C and B at T_4.

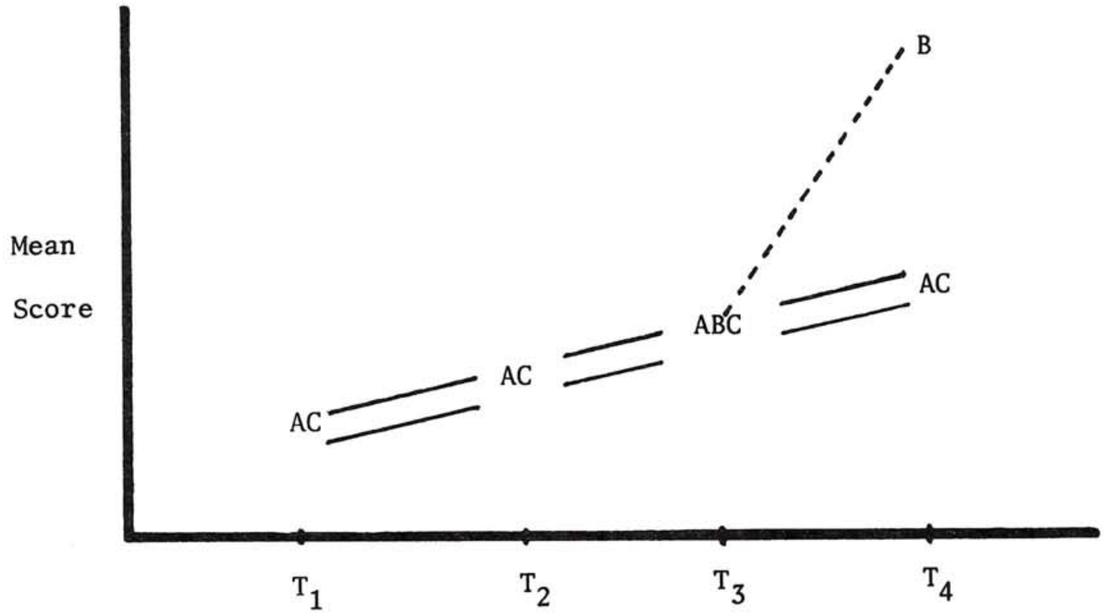

Figure 7. Variation of the Multiple Time Series potentially useful in assessing the effects of educational innovations.

This short review of Time-Series and Multiple Time-Series designs should make clear that a host of possibilities exist in this area. Donald Campbell has predicted the increased utilization of this "family" of designs as work on analysis proceeds. A recent book by Glass and his associates (Glass, Wilson & Gottman, 1975), entitled *Design and Analysis of Time-Series Experiments* is recommended highly for anyone wishing to read more on the topic. Huck, Cormier, and Bounds (1974) provide a somewhat more basic introduction (but still informative) to the use and interpretation of Time-Series designs.

Summary

Specific experimental designs discussed in the program evaluation literature as appropriate for social action program investigations range from the rigorous Pretest-Posttest Control Group Design with randomized assignment of subjects to conditions, Suchman's (1967) postulated "ideal" for evaluattion research, to the methodologically weak pre-experimental designs discussed (and condemned) in Campbell and Stanley (1966). While Suchman (1967) cited the Pretest-Posttest Control Group Design with randomization as the "ideal" social action program research design, he admitted that in most situations the best that could be hoped for was probably the Pretest-Posttest Non-Equivalent Control Group Design. On a continuum of interpretability the Non-Equivalent Control Group design is certainly more ambiguous than the Pretest-Posttest with randomization. The latter, a "true" experimental design, ordinarily rules out all threats to internal validity. The former, a quasi-experimental design, is susceptible to the alternative interpreations of selection-maturation and regression (Campbell & Stanley, 1966) and under some circumstances instrumentation and local history (Cook & Campbell, Note 1), although local history would probably pose as great a threat to the Pretest-Posttest Control Group Design with randomization as to its non-equivalent counterpart under similar experimental conditions. That is, local history might spuriously influence the experimental outcome of either design if treatment and non-treatment groups were intact aggregations, only one of which experienced an extraneous impinging event (a fire across the street, for example).

Both Cook and Campbell (Note 1) and Porter (Note 3) have questioned the interpretability of the Non-Equivalent Pretest-Posttest Control Group Design where pretest and posttest variables are not factorially similar. Porter (Note 4) urged that evaluation strategies be developed prior to program implementation in order to obviate the need to resort to *ex post facto* designs. Campbell (Note 5) unequivocally rejected *ex post facto* designs, and urged that the methodological community rule out all *ex post facto* studies.

Campbell (Note 2) expressed belief that the time-series and multiple time-series designs will be utilized more frequently in the future in social action program research, especially as work on analysis strategies for these two designs proceeds. The time-series and multiple time-series seem particularly amenable to implementation in social action program research on three counts: 1) the time-series design may be utilized where only one research population is available; 2) the multiple time-series design, where a control group is used, does not require random assignment of respondents to the treatment and non-treatment groups; 3) both the time-series and the multiple time-series entail data collection over a period of time. Since social action program participants are typically affiliated with the program over some period of time anyway, these two designs may represent a more practical procedure for assessing program impact than any pretest-posttest design can, due to the latter's specification of only two data collection times.

Reference Notes

1. Cook, T. C., & Campbell, D. T. The design and conduct of quasi-experiments and true experiments in field settings. Prepublication manuscript obtained from Northwestern University, 1973. Subsequently published in M. D. Dunnette (Ed.), *Handbook of Industrial and Organizational Research.* Chicago: Rand-McNally, 1975, 223-326.
2. Campbell, D. T. *Critical problems in the evaluation of social programs.* Unpublished manuscript. Northwestern University, 1972.
3. Porter, A. C. *Analysis strategies for some common evaluation paradigms.* Unpublished manuscript. Michigan State University, 1973.
4. Porter, A. C. *Comments on some current strategies to evaluate the effectiveness of compensatory education programs.* Unpublished manuscript. Michigan State University, 1969.
5. Campbell, D. T. *Methods for the experimenting society.* Unpublished manuscript. Northwestern University, 1971.

References

Campbell, D. T. Factors relevant to the validity of experiments in social settings. *Psychological Bulletin,* 1957, 54, 297-312.

Campbell, D. T. Reforms as experiments. In E. L. Struening and M. Guttentag (Eds.), *Handbook of Evaluation Research* (Vol. 1). Beverly Hills: Sage, 1975, 71-100.

Campbell, D. T, & Stanley, J. C. *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally, 1966.

French, J. Experiments in field settings. In L. Festinger and D. Katz (Eds.), *Research methods in the behavioral sciences.* New York: Dryden, 1953, 98-135.

Glass, G. V., Wilson, V. L., & Gottman, J. M. *Design and analysis of time-series experiments.* Boulder: Colorado Associated University Press, 1975.

Huck, S. W., Cormier, W. H., & Bounds, W. G., Jr. *Reading statistics and research.* New York: Harper & Row, 1974.

Katz, D. Field studies. In L. Festinger and D. Katz (Eds.), *Research methods in the behavioral sciences.* New York: Dryden, 1953, 56-97.

Kerlinger, F. N. *Foundations of behavioral research.* Holt, 1964.

Kirk, R. E. *Experimental designs Procedures for the behavioral sciences.* Belmont, California: Brooks/Cole, 1968.

Nunnally, J. C. The study of change in evaluation research: Principles concerning measurement, experimental design, and analysis. In E. L. Struening & M. Gottentag (Eds.), *Handbook of evaluation research* (Vol. 1). Beverly Hills: Sage, 1975, 101-137.

Ross, H. L., Campbell, D. T., & Glass, G. V. Determining the social effects of a legal reform: The British "Breath-alyser" Crackdown of 1967. *American Behavioral Scientist,* 1970, 13, 493-509.

Stanley, J. C. Quasi-experimentation. *School Review,* 1965, 73, 197-205.

Stanley, J. C. On improving certain aspects of educational experimentation. In J. C. Stanley (Ed.), *Improving experimental design and statistical analysis.* Chicago: Rand McNally, 1967, 1-27.

Suchman, E. A. *Evaluative research.* New York: Russell Sage Foundation, 1967.

Suchman, E. A. Action for what? A critique of evaluative research. In C. H. Weiss (Ed.), *Evaluating action programs.* Boston: Allyn & Bacon, 1972, 52-84.

Wholey, J. S., Scanlan, J. W., Duffy, H. G., Fukomoto, J. S., & Vogt, L. M. *Federal evaluation policy.* Washington, D.C.: The Urban Institute, 1970.