By Gordon Welty  
Wright State University  
Dayton, OH 45435

Americans are vitally concerned for the improvement of the quality of education. President Nixon recently rejected the notion of "more dollars for the same old programs without making the urgent new reforms that are needed." He went on to propose "a new and searching look at our American school system."  

A critical aspect of the educational community's ability to meet the President's challenge is its ability to mount a viable program evaluation effort. To date, educational evaluation has been most notable in its failure.  

Speaking of evaluation in general, the American philosopher Clarence Irving Lewis points out that actions could not attain success except that there are evaluations, which are essentially predictions. "Whether the action is performed or not will depend upon evaluations made."  

In terms of general system theory, Lewis is emphasizing the necessity of feedback in any action scheme.  

Feedback of evaluative reports to decision makers is a necessity in the rationally managed school system, if it is viewed as a system. On the revision of ongoing educational programs, for instance, J. T. Hastings has stated that "without such feedback, either the decision to revise or the decision not to revise -- and most certainly the decision of how to revise -- must be based upon feeling tones and the arguments of personal preference."  

It has been frequently maintained that the demand for evaluative feedback is incompatible with the classical experimental design. For example, consider Dean Egon Guba's statement that "the application of conventional experimental design to evaluation situations...conflicts with the principle that evaluation should facilitate the continuous improvement of a program." A requirement of invariance of treatment and control are sufficient on Guba's argument to preclude program change. Thus, these conditions are sufficient to preclude evaluation feedback for managerial decision-making, because "treatments cannot be altered if the data about differences between treatments are to be unequivocal."  

Guba considers this a problem of "evaluation methodology." As such, it is a problem of paramount importance to the development of evaluation theory. If the dissemination of
evaluative findings is not permitted to preserve the experimental design, then evaluation loses its value to the decision-maker. On the other hand, if experimental design is to be abandoned, serious problems await evaluators in the development of alternatives. We find, however, that these contentions are not valid. It is not the case that treatment must remain invariant. Corrective action by program managers, in light of evaluative feedback, can take place concurrent with an evaluation in the framework of an experimental design.

Of crucial importance is what Guba intends by treatment alteration. Suppose that quantitative change is a change of the value of a given variable (whether an intensive or extensive measurement); then it is convenient to let qualitative change be a change of a variable or dimension itself. Could not the latter be Guba's intent? Let us, then, consider as a possible meaning of qualitative change that a variable or dimension is simply added or deleted from the analysis, as the program manager adds or subtracts from the program, and treatment alterations correspond to this.

The Principle of Dimensional Homogeneity states that for a given equation, all the dimensions or variables in the equation can be categorized in terms of a collection of fundamental measures. For example, if volume occurs in an equation, the dimensions of volume are categorized in terms of length. The principle also states that the dimensionality of the variables (the dimensionality of volume for instance is 3) on the right- and left-hand side of the equation, by fundamental measure, must be equal. This is known as the \( \pi \) Theorem. For example, velocity is distance per time. The fundamental measures for velocity are two, length and time. Dimensionality is 1 and -1, respectively. As this also is the case for distance divided by time, the formula is dimensionally homogeneous. If it were otherwise, the introduction or deletion of either fundamental measures or dimensionally across the equation would be a case of ad hoc theorizing, however subtle./10/

So it is not possible, as a methodological point, for there to be qualitative change in the sense of adding or deleting variables. Hence, we consider here only the case of treatment variance which is both a rational response to evaluative feedback and also treated as dimensionally homogeneous, and indicate how this is compatible with experimental design.

My colleague, Alfred Beradino and I have recently provided a proof that an experimental design and the feedback of action research findings are compatible./11/ This provides sufficient conditions for the falsity of Guba's methodological argument.

The question now can be raised whether a precedent exists in the literature for the use of experimental designs in this fashion. In the writings of Sir Ronald Fisher, we find sufficient conditions for the methodological (not practical) arguments for the use of experimental designs in evaluative research.

In his classic *Design of Experiments*, Fisher proposes to "examine the physical conditions of the experimental technique." After mentioning that matching of conditions across treatment levels in the experimental design is a formal condition for minimizing errors,
Fisher argues it is impossible to realize this condition in fact, since "uncontrolled causes which may influence the result are always strictly innumerable."/12/ With regards matching of conditions, the assumption that "refinements constitute improvements to the experiment" is dismissed on the basis of cost considerations. Since matching is a sufficient but not a necessary condition, control of errors in the experiment can and must be realized by other means. The cost of complete matching across treatment levels would be (quite strictly) infinite, and since "an essential characteristic of experimentation is that it is carried out with limited resources," Fisher proposes randomization as an alternative. This is a procedure by which the experiment "may be guaranteed against corruption by the causes of disturbance which have not been eliminated."/13/ More precisely, random assignment of subjects to treatment levels permits a precise estimate of error. Thus there are two and only two sufficient conditions for experimental control; hence, one of the two is always necessary. Irrelevant variables are eliminated in effect either by matching of conditions, that is, "eliminated in the field," or by randomization. Fisher emphasizes the sufficiency of the latter technique when he argues that "it is apparent that the random choice of the objects to be treated in different ways would be a complete guarantee of the validity of the test of significance, if these treatments were the last in time of the stages in the physical history of the objects which might affect their experimental reaction."/14/ This is to say that randomization is sufficient in the absence of treatment variation, to which Guba would undoubtedly agree.

In evaluation or action research, a new aspect is added. Because of the various institutional contingencies, it is usually an unacceptable policy to randomly choose subjects for treatment. It would be possible, for instance, to take the lower two-fifths of the students, as ranked by a standardized achievement test. This group could then have remedial treatment provided, by random assignment, to one half, which would amount to one-fifth of the total population. However, it is usually policy to take the lowest fifth, and administer treatment to them as a group. Thus no "control group" is available. This is, however, an institutional contingency, hence not a methodological problem per se. Randomization is still a possibility, hence Fisher's discussion of randomization is relevant to the methodology of evaluation.

Fisher generalizes his argument at this point by emphasizing that variance in treatment subsequent to randomization presents no "practical inconvenience." He states "subsequent causes of differentiation, if under the experimenter's control .. can either be predetermined before the treatments have been randomized, or, if this has not been done, can be randomized on their own account."/15/ The first alternative here is merely the recognition that the rational decision maker's response to evaluative feedback is program change. The second alternative is excluded from our discussion, as randomly distributed response by a program manager is not conducive to systematic pursuit of policy.

At this point, we can discuss three possible sources of error: first, consequences of differences already randomized: these are accounted for by the initial randomization; second, natural consequences of the difference in treatment levels: since the null hypothesis argues there will be no treatment effect, there can be no consequences of this effect; and third, effects supervening by chance, independent of treatment levels: because
of random assignment, estimates of deviance from a specified distribution across all treatment levels for these effects can be given. Any systematic variance will have been eliminated by the initial randomization.

The dissemination of evaluation findings to the rational program manager/16/ will produce program change. As a corollary of the Principle of Management by Exception, we know that if a defect in the program is noted and reported, given adequate program resources, the defect will be corrected by the rational manager. Thus both the corrective action of the manager, and the adequacy of resources are determinate. As such, the variance of treatment as a function of evaluative feedback can in Fisher's terms, be "predetermined."

Hence, we see that, contrary to the contentions of Guba, it is possible to implement a rigorous experimental design, and also provide feedback for managerial decision-making, in the context of action research. Whether practical concerns, such as the competence of the researcher, or the resources and administrative support available to him, do in fact militate against his ability to implement a rigorous design, is not a methodological issue, and not under consideration here. On the other hand, if the manager is oblivious to feedback, or responds to feedback with random and affective behavior rather than systematic and rational action, this is a psychological issue, and not under consideration here. But the methodological "problem," posed by Guba, can be considered ill-conceived and non-existent.

Notes

This is a revised version of a paper presented in Division D, AERA 1969 Annual Meeting, Los Angeles, California. (February 7, 1969)


5. Ibid, Chap. XII, esp. p. 367f.


9. Guba, *ibid*.


15. Ibid.