DOCUMENT RESUME

ED 323 650 EA 022 234

AUTHOR Pollard, John A.

TITLE The Implementation of Randomized Methods of Subject Assignment in Field Settings: Empirical Findings and the Evaluation of the "Here's Looking at You, 2000" Program.

SPONS AGENCY Haynes Foundation.; National Inst. of Mental Health (DHHS), Rockville, MD.; Russell Sage Foundation, New York, N.Y.

PUB DATE Apr 90

CONTRACT NIMH-MH-39958; NIMH-MH-42694

NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).

PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS *Drug Education; Elementary Secondary Education; Evaluation Methods; *Evaluation Utilization; Evaluators; Meta Analysis; Politics of Education; Program Evaluation; Quasiexperimental Design; *Research Design; Research Problems; Research Utilization; Sampling; *Substance Abuse

ABSTRACT

The implementation of research design in studies on substance abuse education is discussed in this meta-analysis. Findings indicate that the quality of research design is constrained by negative feedback from evaluation stakeholders. This finding is applied to an evaluation of the Here's Looking at You, 2000 drug education program currently underway in California. A conclusion is that the evaluation enterprise, particularly when it concerns the method of subject assignment, is influenced by various mutually antagonistic social and political factors that result in methodological compromise. Seemingly contradictory recommendations for researchers are to adhere to more rigorous evaluation methods and to maintain realistic expectations. A goal is the development of a randomization process that is robust against indifference, incompetence, and corruption. Four statistical tables are included. (12 references) (LMI)

THE IMPLEMENTATION OF RANDOMIZED METHODS OF SUBJECT ASSIGNMENT IN FIELD SETTINGS: EMPIRICAL FINDINGS AND THE EVALUATION OF THE *HERE'S LOOKING AT YOU, 2000* PROGRAM

John A. Pollard

Southwest Regional Laboratory
Western Center for Drug Free Schools and Communities

Paper prepared for the American Educational Research Association annual meeting, Boston, MA., April, 1990.

The research reported in this paper was supported by the Haynes Foundation, the National Institute of Mental Health (MH39958 and MH42694) and the Russell Sage Foundation.

Reviewers have recognized that, in the general evaluation literature, the typical program evaluation utilizes poor methodological practices (Lipsey, Crosse, Dunkle, Pollard & Stobart, 1985). In the field of education, methodological criticisms have been substantial. Experimental design traditionally has received the most attention (Cook and Campbell, 1979). This methodological issue dominates most other technical considerations in evaluation textbooks (e.g., Rossi and Freeman, 1972). In general, the quality of experimental design in field settings throughout the social sciences has been criticized (Lipsey, et al., 1985).

This paper discusses the findings of some recent research examining the implementation of research design in the domain of treatments for juvenile delinquency (Pollard, 1990). The major thrust of this research is that the quality of research design appears to depend upon the conflicting and contradictory influences of a variety of evaluation stakeholders. Further, through a process of negative feedback in a manner analogous to a thermostat, the evaluation stakeholders eventually constrain the research design within a narrow range of methodological rigor. The findings of this research are then applied to understand the implementation of an evaluation of the *Here's Looking at You, 2000* program, an evaluation that is currently underway in California. The decision making and negotiation process leading up the the final formulation of the research design in the *Here's Looking at You, 2000* evaluation illustrates the findings of this research.

The Implementation of Research Design

Many evaluators still routinely employ quasi-experimental designs. Lipsey, et al. (1985), in a random sample of 175 published program evaluations conducted between 1978 and 1980, found that 48% of all evaluations utilized weak quasi-experimental designs and 16% utilized pre-experimental designs that did not incorporate any form of a comparison or control group. While circumstances may limit design options, Boruch and Wothke (1985) have pointed out that fully randomized experiments have been successfully implemented in quite diverse field environments. Unfortunately, program evaluators still most often rely upon the

weaker designs. There are probably numerous factors that limit the implementation of randomized experimental designs. One significant factor may be that the evaluation literature has not provided well defined procedures for how to implement a randomized design in a field treatment setting. Boruch and Wothke (1985) suggest that this weakness leads to a failure by most evaluators to seriously consider implementation of a randomized design. Instead, there is a too easy acceptance of alternative methods. For example, Boruch and Wothke (1985) remarked that many research design textbooks (e.g, Kirk, 1972; Winer, 1971) merely state that randomization should take place, but never give concrete procedures for how randomization can actually be conducted.

There are also substantial political and ethical questions that must be answered for evaluation stakeholders in any evaluation setting that is employing randomization. Individuals, not well versed in the logic of research design, can view randomization as arbitrary, unfair and unethical because it seems to be a denial of services to people in need. These questions, if not fully answered, will often frustrate the attempt to randomize subject assignment. In what is becoming a classic example of this, Sherman and Berk (1984, 1985) invested an extensive amount of time and effort convincing local authorities that the randomization of arrest procedures by police when called to domestic violence situations was both practical, ethical, and beneficial to all parties. They encountered substantial resistance, both from the police and other parties, but eventually were able to implement their plans. It was only because of their extensive efforts and persistence that this study was implemented as it was.

There are several other examples of randomization in turbulent field settings (see Boruch and Wothke, 1985). The common features of these examples are a strong commitment to the value of randomization by all parties, a strong and continuing presence of the evaluator throughout the life of the evaluation, and a willingness to educate those individuals who will actually implement the evaluation as to the rationale for randomization.

In summary, the details of how evaluators select and implement specific research designs are not well known. While the literature on this topic is increasing, no systematic

studies have previously been completed on this topic. What is available suggests that numerous political and social factors influence design choice.

The Present Research

Recent research, based on the author's dissertation (Pollard, 1990), provides some insight into many of these questions about how research design is implemented in evaluation settings. This research was conducted as part of a large NIMH sponsored meta-analysis of the juvenile delinquency treatment evaluation literature (Lipsey, 1990). In the dissertation research, 269 evaluation studies of the efficacy of treatments for juvenile delinquency were included in a meta-analysis. The primary goals of the dissertation meta-analysis were: 1) to determine how the social and political context of the evaluation affected several important methodological variables in the evaluation, including the method of subject assignment; and 2) to determine how the methodological characteristics of the evaluations, as they were eventually implemented, eventually impacted the reported treatment effectiveness. This latter analysis is not reported here but is available from the author (Pollard, 1990).

In the following results, the degree to which the method of subject assignment could be predicted by variables coding for the political and social context of the evaluation (called the Study Context variables) was examined. (The Study Context variable are listed in Table 1.) Looking at the results on Table 1 for the variable Method of Subject Assignment, about 46% of the studies used randomized methods of subject assignment, while 26.8% of the studies used strong quasi-experimental designs (e.g., matched subject designs, equated groupwise), while the remainder (24.9%) used weaker quasi-experimental designs (convenience samples of volunteers, intact groups, etc.). Studies with pre-post designs were not included in the research . This analysis investigated the extent the evaluator's choice of method of subject assignment was predictable from other characteristics of the evaluation that presumably predate the evaluator's choice of design. In other words, do

characteristics of the early evaluation context in some way constrict or influence the researcher's choice of method of subject assignment?

In Table 2, the summary results of the regression analysis predicting Method of Subject Assignment from several Study Context variables, one missingness variable and one coder confidence variable are presented. In Table 2 the criterion variable is listed at the top with the number of valid cases utilized in this analysis. The meaning of the direction of the correlations is also provided at the top of the table. The predictor variables, and their order of entry, are provided in the next section. Included in this section are the correlation coefficients (r) between the individual predictor variables and the criterion, as well as the final beta values and their associated t-tests. If the beta value was not significant at least to the $p \leq .15$ level, the t-test value was not entered into the table. In the final section of the table, in the Summary of Results section, the R^2_{Change} and R^2_{Total} values are provided, along with F ratios and significance levels for each set of variables.

A missingness variable was included in the analysis with its corresponding coded variable if more than 10% of the citations reported missing data on the coded variable. In the case of the analysis presented in Table 2, only one variable, coding for whether the study was conducted at a Criminal Justice Site, met this criterion.

The first set of variables entered into the regression equation were the variables: Year of Publication and Academic Affiliation of Senior Author. The second set of variables entered were the remainder of the Study Context variables. In the third and final step of this regression, only a single variable was entered, Coder Confidence in Method of Subject Assignment. This variable represented how confident the person coding the evaluation report for the meta-analysis project felt in their judgement about the Method of Subject Assignment. In terms of the causal logic outlined above, each successive set of variables was presumed to be potentially able

to influence the criterion variable and the predictor variables that entered later into the regression equation, but not the variables that have already entered the equation.

In Table 2, after entry of the first two variables, $R^2 = .029$, p =021. This means that approximately three percent of the variance of the criterion variable Method of Subject Assignment could be accounted for by the two variables Year of Publication and Academic Author Affiliation.

After the introduction in Step 2 of the remainder of the Study Context variables, $R^2_{Total} = .194$, ($p<.001$). This represents the total amount of variance accounted for by all of the variables entered into the analysis to that point. The amount of variance accounted for by the second variable set was $R^2_{Change}=.164$ (p = . 0001). This is calculated simply by taking the Total R^2 at that point, and subtracting the Total R^2 calculated at the previous step. In this case R^2_{Change} = .194 - .030 = .164.

Finally, in Step 3, the Coders' Confidence in Method of Subject Assignment was entered into the regression. This variable, entered alone, made a significant contribution to the analysis ($R^2_{Change} = .040$, $p <001$). The Total R^2 after the entry of this final variable was .233, $p<.001$.

At the individual variable level, six variables were significant predictors of the criterion variable, as assessed by their individual beta weights and associated t-test statistic. Academic Author Affiliation, Evaluator Involvement with the Treatment, whether the evaluation was conducted with a Demonstration Program , Criminal Justice Site, and Coder Confidence in Method of Subject Assignment were all positively correlated with Method of Subject Assignment. This means that the author reporting an academic affiliation, the evaluator being heavily involved with the treatment, the treatment being described as a demonstration program, the treatment being conducted at a criminal justice site, and the coders' confidence in the method of subject assignment were all associated with the presence of a randomized method of

subject assignment. Conversely, Mandatory Treatment Participation was negatively correlated with Method of Subject Assignment, meaning that the juveniles being forced to participate in the experimental treatment was associated with weaker methods of subject assignment. The Interaction between Evaluator Involvement and Demonstration Program showed a negative, non-significant relationship with Method of Subject Assignment.

At the individual variable level, most of the above effects initially appear straight forward in their explanation. Academic evaluators are likely to adopt a more rigorous method of subject assignment because of their training and research activities. The control offered the evaluator by conducting the study at a CJ site would make the adoption and implementation of a randomized procedure logistically more feasible. Also, naturally enough, evaluators interested enough in the program/treatment delivery may be more willing to adopt a careful approach to evaluation. They also are more likely to be in a position to approve of and enforce a good randomization procedure. Consistent with this explanation, it may be that evaluators who are independent of the program do not have the political "clout" to generate the approval or implementation of a randomized procedure, which is typically more difficult to get program stakeholders to approve. It is not clear is why Mandatory Treatment Participation, which again should offer the evaluator greater influence and control in the evaluation setting, was negatively correlated with randomized methods of subject assignment. It is possible that, because of the seriousness of offenses associated with juveniles in institutional settings, more outside stakeholders were involved in the formation of the treatment group (e.g., law enforcement officials, judges, institutional officials). These stakeholders may be less willing to allow randomization to occur in settings where strict control over the juveniles and their services is maintained. (The next analysis, looking at control group activity, provides results supporting this interpretation.)

Finally, coder confidence regarding the method of subject assignment utilized in the study made its own significant contribution to the predictability of Method of Subject Assignment. One possible interpretation of this findings is that randomization was associated with better quality (or more enthusiastic) evaluation reporting. Perhaps evaluation authors are more willing to be explicit on this variable when they have utilized a randomized procedure. Thus, the coders were simply more likely to be more confident of the design when the evaluation report was more explicit.

However, interpretation of the individual variable effects does not provide a complete view of these results. It is important in this analysis to compare the final beta weights of the individual variables with their associated correlation coefficients (r). Generally speaking, the final beta weights are smaller than the original correlation coefficients between the predictor variables and the criterion. This is because the variance accounted for by the correlation coefficient includes some criterion variable variance that is not uniquely associated with the predictor variable. The beta weight, on the other hand, represents only that criterion variable variance that is uniquely associated with predictor variable. Thus, the beta value is typically smaller than the correlation coefficient.

The final beta weights reported in Table 2 were not consistent with this pattern. Several predictor variables showed beta values either of opposite sign, or larger than, their correlation coefficients. When this pattern occurs, this indicates that the measurement of the direct relationship between a criterion and a predictor variable (the correlation coefficient) is being "suppressed" by the effects of other predictor variables. That is, the effect of the predictor variable (X1) upon the criterion (Y) is obscured by the effects of other predictor variables (X2, X3, ..., Xn), that are themselves correlated with the first predictor and the criterion.

Because of the role that suppression will play in this and two later analyses, it is worthwhile to elaborate on the concept of suppression. For an hypothetical

example of suppression, suppose that the predictor variable Level of Taxation (X1) is positively associated with the criterion variable Government Revenue (Y), and that the predictor variable Honesty in Tax Returns (X2) also is positively associated with Government Revenue (Y). However, suppose that the Level of Taxation (X1) is negatively associated with Honesty in Tax Returns (X2), because as the burden of taxation increases, people are more likely to find "deductions" to reduce their tax liability. Further, as honesty declines, the government might be inclined to raise the taxation level to make up the lost revenue. The relationship between these three variables is presented in Figure 1.

In the hypothetical case represented by Figure 1, the beta weights for the predictor variables might well show a moderate positive relationship with the criterion, while the correlation coefficients between the predictors and the criterion would be smaller than the beta weights. For example, the indirect effect of Level of Taxation (X1), acting to reduce Honesty in Tax Returns (X2), would have the effect of lowering government revenue. Thus, as the Level of Taxation (X1) increases, Government Revenue (Y) decreases because of the indirect effect Level of Taxation (X1) has through Honesty in Tax Returns (X2), but increases because of the direct link between the two variables. In effect, the direct and indirect effects of Level of Taxation (X1) counteract each other. Thus, without statistically controlling for the value of Honesty in Tax Returns (X2), the correlation coefficient between Level of Taxation (X1) and Government Revenue (Y) appears small. Of course, the reverse could also occur, where the indirect effect and direct effects of Honesty in Tax Returns (X2) would act in a similar, antagonistic way upon Government Revenue (Y). Given actual values for the correlations between the variables, an equilibrium point could be calculated which would maximize Government Revenue (Y) through the choice of optimal values for Level of Taxation (X1) and Honesty in Tax Returns (X2).

This example of suppression has been presented in some detail because it appears that some of the predictor variables influencing Method of Subject Assignment act in a similar fashion. Returning to the analysis of Method of Subject Assignment, one example of suppression occurs with Criminal Justice Site and Academic Author Affiliation. The beta values for these variables indicate that they both have a positive relationship with Method of Subject Assignment. However, Criminal Justice Site and Academic Author Affiliation are negatively correlated with each other. If Criminal Justice Site is substituted for Level of Taxation (X1), Academic Author Affiliation is substituted for Honesty in Tax Returns (X2), and Method of Subject Assignment is substituted for Government Revenue (Y), the pattern of relationships, and the suppression effects, are exactly the same. For both predictor variables, their indirect effect upon the criterion Method of Subject Assignment would be in the opposite direction of their direct effect reflected by the beta weight. This produces the suppression of the correlation coefficient.

There are other examples of variables in this analysis exhibiting the same pattern. One way to systematically identify suppressive variable relationships is the following: Any two predictor variables that both exhibit the same sign for their final beta weight, either positive or negative, will be in a suppressive relationship when the predictor variables are negatively correlated. Conversely, any two predictor variables with opposite signs for their beta weights will exist in a suppressive relationship when there exists a positive correlation between the predictors.[1]

Table 3 presents three correlation matrices. The upper matrix includes only variables that exhibit positive beta values with the criterion Method of Subject

---

1 This decision rule does not identify all suppressive variable relationships. All suppressive relationships can be identified by determining when any one of the correlations between the two predictor variables and the criterion, r_{Y1}, r_{Y2}, or r_{12}, is less than the product of the other two correlations. This decision rule described in the text has been used in this situation for purposes of clarity in understanding the analysis of Method of Subject Assignment (SH27).

Assignment. In that matrix, all statistically significant ($p<.05$) negative correlations between the predictor variables in the matrix are presented. A negative correlation indicates that a suppressive relationship exists between the predictor variables. (Positive correlations are omitted for clarity.) The second matrix includes only variables that exhibit negative beta values with Method of Subject Assignment. For this matrix, again only statistically significant negative correlation coefficients in the matrix are presented. In the third matrix at the bottom, predictor variables having a negative beta weight are listed along the left side of the matrix. Predictor variables having a positive beta weight are listed across the top of the matrix. For this correlation matrix, a positive correlation between the two predictors is indicative of suppression. The missingness variable for Criminal Justice Site and Coder Confidence in Method of Subject Assignment have not been included for clarity.

In Table 3 it can be seen that for the criterion Method of Subject Assignment many pairs of predictor variables exhibit suppression. One interesting example involves the predictor variable Interaction between Evaluator Involvement and Demonstration Treatment. This variable, initially showing a modest positive correlation coefficient with the criterion, shows a final beta weight of approximately equal magnitude but of opposite sign ($\beta = -.280$, $p = .1171$). Suppression appears between this variable and the predictor variables Evaluator Involvement with the Treatment and Demonstration Program . These two variables both exhibit positive beta weights with the criterion. And, as would be expected because of their role in the creation of Interaction between Evaluator Involvement and Demonstration Treatment, both are strongly and positively correlated with Interaction between Evaluator Involvement and Demonstration Treatment.

Interaction between Evaluator Involvement and Demonstration Treatment was created for this research because it was anticipated that special effects might be found in evaluations where an academic evaluator would be testing an experimental

demonstration program of his or her own design. In particular, it was expected that this variable might capture evaluations with unusually strong methodological characteristics. Judging from these results, and contrary to expectations, it appears this variable was sometimes associated with evaluations where a randomized method of subject assignment was not frequently employed. This result might indicate a situation were development of the experimental treatment is the primary objective and/or the maturity of the treatment program is not yet ready to support a rigorous evaluation.

The question remains how should the overall pattern of suppression among the predictor variables should be interpreted. As previously noted, suppression is often characteristic of homeostatic systems. Homeostasis arises because system variables provide negative feedback when the system deviates from a system norm or equilibrium point. One possible explanation for widespread pattern of suppression with Method of Subject Assignment is that there is an "equilibrium point" for the rigor of the research design in many evaluation settings. The context of an evaluation is usually made up of several key players, or stakeholders, who are interested in the conduct or outcome of the evaluation. In addition, there are important variables such as supporting resources and the availability of program clients, that also impact implementation of the research design. Evaluation stakeholders often do not have common interests in the evaluation, and place conflicting demands on the evaluation. For example, policy makers or senior administrators might want to have the most unbiased assessment of program effects, while program personnel might want to minimize the impact of the evaluation on the program's clients (e.g., by preventing the "unfair" denial or postponement of services to some clients for purposes of randomization). Further, stakeholders may have expectations regarding the evaluation that are inconsistent with the evaluation resources.

In this context, the evaluation enterprise is similar to a homeostatic system. Various stakeholders, resources, and program and client factors provide negative feedback to the evaluator when the evaluation design deviates too significantly from their expectations or limitations. The net effect of these influences is that many evaluations may reach an equilibrium point in the rigor of the research design. This equilibrium point would be the method of subject assignment that is consistent with the social, political, and economic context of the evaluation. From this perspective, the frequent suppression evident with the predictor variables in this analysis represented the negotiated, evolutionary process that occurred in the formation of the research design.

Subject Assignment Procedures in the *Here's Looking at You, 2000* Evaluation

An evaluation of *Here's Looking at You, 2000*, currently in progress in California, provides a concrete example of the above findings. The *Here's Looking At You, 2000* curriculum is a comprehensive, K-12 curriculum designed to prevent or reduce the student's use of alcohol, tobacco, and marijuana (Roberts, Fitzmahan & Associates, 1986). *Here's Looking at You, 2000* focuses on these three drugs because they are believed to act as "gateway drugs," meaning that adolescents typically begin their drug careers with the use of these drugs prior to beginning use of other illicit drugs. *Here's Looking at You, 2000* is designed specifically to counteract the known risk factors leading to adolescent drug abuse: family history of abuse of alcohol and other drugs, permissive family attitudes towards alcohol, tobacco and other drugs, academic failure, lack of social bonding, peer influences, and the student's favorable attitudes towards drug use. The lesson plans developed by *Here's Looking at You, 2000* focus on developing the student's internal resources which will be necessary to avoid drug use. Specifically, *Here's Looking at You, 2000* 1) promotes bonding between the student and society, 2) provides students information on the effects of drugs,

and 3) provides the students the necessary skills by which they can resist peer and societal pressures to use alcohol, tobacco and marijuana.

Here's Looking at You, 2000 is the most popular drug prevention curriculum in school districts in California (Duerr Evaluation Resources, 1989). However, other than anecdotal evidence, there is currently little empirical evidence measuring the effectiveness of the *Here's Looking at You, 2000* curriculum. Given that a substantial amount of money is being spent by California Schools on implementation of the *Here's Looking at You, 2000* curriculum, there is good reason to evaluate whether measurable, positive effects of the program can be identified.

The district in which the *Here's Looking at You, 2000* evaluation is being conducted is a medium sized district with a large geographic are encompassing bot urban and rural areas. In addition, it has a wide variation in SES among its students. The district has been active in drug prevention efforts for many years, beginning to develop comprehensive services in the early 1980's. In general, the district is open to the development and evaluation of drug prevention efforts perceived to be beneficial to its students.

Several stakeholders in the district played a significant role in the development of the evaluation. However, not all of them were equally enthusiastic or committed to the conduct of the evaluation. Most stakeholders' attitudes towards the prospective evaluation were a function of their perceived gains or losses as a result of the conduct of the evaluation. Their positions are reflected in Table 4.

From Table 4 it can be seen that there was a relative balance in the position of the stakeholders between those who wanted to adopt a rigorous method of subject assignment (e.g., randomization of individual students) and those who resisted this approach. Those stakeholders who resisted randomization did so primarily because of the disruption and the logistical difficulties that would be encountered if individual students were assigned either to receive the *Here's Looking at You, 2000* curriculum or would receive no services. Because of the amount of classroom time the *Here's Looking at You, 2000* curriculum requires, it is no

small inconvenience to be randomly pulling students from classrooms to receive the curriculum.

The reasons for other stakeholders not accepting less rigorous methods of subject assignment primarily focused on the lack of usefulness of the results if a poor quality subject assignment procedure was adopted. The evaluator was not interested in committing a significant amount of time or resources to the project if uninterpretable or ambiguous results were unavoidable. Administrator #1 had a similar, research "minded" orientation and also was not very interested in the evaluation if minimum standards could not be maintained.

Ultimately, the net effect of these forces were that a compromise was reached. The *Here's Looking at You, 2000* program was presented in 7th grade social studies classes at each of the four middle schools in the district. This was the first year that the 7th grade social science teachers would present the curriculum, and all had to be trained prior to their presenting the program. Further, because of constraints in the number of substitute teachers who could cover classes of the social science teachers as they were being trained, and because of limitations in the amount of trainer time that was available, teachers training in the *Here's Looking at You, 2000* curriculum had to be spread out across the entire year. It was decided to take advantage of these facts by randomly assigning the 7th grade social science teachers to receive their training either in September, February, or May of the 1989-90 school year. Each of the three 7th grade social science teachers at each middle school was randomly selected to participate in one of the three training sessions. Since the teachers were to present the *Here's Looking at You, 2000* curriculum immediately after the training, this meant that one-third of the 7th grader social science students received the program in October, one-third in March, and one-third would not receive the curriculum at all in this school year. In short, while students were not

randomly assigned, the teachers were randomly assigned within each of the four middle schools.

The final form of the evaluation design was the end product of several competing and conflicting forces within the evaluation setting. It could be said that the final form of the subject assignment procedure was a compromise among the differing expectations and demands of the stakeholders. But more importantly, the final form was a kind of "equilibrium point" around which the final evaluation plan probably could not have significantly differed. The various stakeholders, because of their "drop dead" criteria, continually provided negative feedback to the negotiation process when the proposed plan differed too widely from their expectations. From this perspective it appears that the most likely outcomes for this evaluation were that this particular method of subject assignment would be adopted or that the evaluation would not be conducted at all. This kind of mutual constraint among the stakeholders was similar to what was described in the meta-analysis research.

Two other broad conclusions emerge from these findings. These conclusions are both consistent with these results, yet are somewhat contradictory with each other. First, given that these results provide the evaluator with more insight into the forces surrounding the implementation of subject assignment procedures, the evaluator may be able to plan more effectively. Recognizing the probable forces that surround the implementation of an evaluation, the evaluator could more effectively negotiate among the various stakeholders, eventually producing a more rigorous evaluation methodology.

The second broad conclusion, as mentioned, is somewhat contradictory to the above comments. The evaluator should be realistic as to what he or she will be able to achieve in implementation of research design. Trying to force a rigorous method of subject assignment in an evaluation whose context will not support that effort may be doomed to failure. A typical evaluation problem is that what was planned as a rigorous investigation of the effectiveness of a treatment degrades into an evaluation that is neither as rigorous or as

informative as was hoped. If the analysis of this paper is correct, an evaluator trying to impose a rigorous design in a system whose "equilibrium point" for method of subject assignment is substantially lower may be at high risk for frustration.

These comments are not meant to detract from the work of authors such as Boruch and Wothke (1985) who argue that more rigorous evaluations could be implemented than is now routinely the case. In fact, their position is quite consistent with these views. They argue that the evaluator can enhance their evaluation by modifying exactly the kinds of factors in the evaluation setting that are represented in this analysis. Boruch and Wothke state that: "Interest and support at higher levels in loosely coupled social systems do not guarantee randomization at operating levels. Incentives and benefits that offset the sharing of control have to be planned to engage participation and to sustain agreements at lower levels." (p. 108) While the language used to describe the results of this research is sometimes different than that used by Boruch and Wothke (1985), the results are consistent with the conclusions drawn from their experience. The evaluation enterprise, particularly when it concerns the method of subject assignment, is characterized by numerous social and political factors that are mutually antagonistic and that act to counteract each other's influence. Left unrestrained, these factors ultimately result in a "compromise" in the method of subject assignment. This compromise typically gives none of the stakeholders all that they want, and results in proliferation of evaluation studies with less than optimal research designs. Boruch and Wothke continue: "It is not easy to design the randomization process so that it is robust against indifference, incompetence, and corruption. Still it is a task that seems worth the effort." (p. 98)

REFERENCES

Boruch, R. F., & Wothke, W. (1985). Seven kinds of randomization plans for designing field experiments. New Directions for Program Evaluation, 28, 95-114.

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis in field settings. Boston: Houghton Mifflin.

Duerr Evaluation Resources (1989). An assessment of alcohol and other drug prevention program components in California school districts during the 1987/88 school year. Unpublished manuscript.

Kirk, R. E. (1982). Experimental design: Procedures for the behavioral sciences (2nd ed.). Belmont, CA: Brooks/Cole Publishing Company.

Lipsey, M. W., Crosse, S., Dunkle, J., Pollard, J.,, & Stobart, G. (1985). Evaluation: The state of the art and the sorry state of the science. New Direction for Program Evaluation, 27., 7-28.

Lipsey, M W. (in press). Juvenile Delinquency Treatment: A Meta-analytic inqr iry into the variability of effects. Meta-analysis for Explanation: A Casebook. Russel Sage Foundation, Spring, 1991.

Pollard, J. A. (1990). The methodological conduct of juvenile delinquency treatment evaluations and the reported effectiveness of the treatment. Unpublished doctoral dissertation, Claremont Graduate School.

Roberts, Fitzmahan & Associates (1986). *Here's Looking At You, 2000: A Teacher's Guide for Drug Education*. Seattle, Washington.

Rossi, P. H., & Freeman, H. E. (1982). Evaluation: A systematic Approach. Beverly Hills, CA: Sage.

Sherman, L. W., & Berk, R. A. (1984). The specific deterrent effects of arrest for domestic assault. American Sociological Review, 49, 261-272.

Sherman, L. W., & Berk, R. A. (1985). The randomization of arrest. New Directions for Program Evaluation, 28, 15-26.

Winer, B. J. (1981). Statistical procedures in research design. New York: McGraw-Hill.

Table 1

Descriptive statistics of the variables used in this study.

RATIO VARIABLES		Mean	S.D.	Valid N	Miss.
Year of Publication		1976.76	6.105	258	4.1%
CATEGORICAL VARIABLES:	**Category**			**Pct.**	**N**
Affiliation of Senior Author	Other Affiliation			31.2	84
	Academic Affiliation			62.1	167
	Missing			6.7	18
Evaluation Funding	Not Funded			30.5	82
	Funded			67.3	181
	Missing			2.2	6
Evaluator Involvement with Treatment	Independent of Treatment			33.8	91
	Influential, No Dir. Role			14.9	40
	Plan-Control-Supervise Treatment			33.5	90
	Delivered Treatment			7.8	21
	Missing			10.0	27
Program/Treatment Age at Time of Research	Relatively New			63.2	170
	Established (2+ yrs.)			32.3	87
	Defunct Program			1.5	4
	Missing			3.0	8
Is Treatment Part of Demonstration Program	Regular Program			62.5	168
	Demonstration Program			34.9	94
	Missing			2.6	7
Mandatory Treatment for Juveniles	Participation Voluntary			3.0	8
	Referred By Family and Friends			.7	2
	Referred By Non-CJ Agency			13.4	36
	Voluntary Referral, CJ Agency			30.5	82
	Mandatory Referral,CJ Agency			45.0	121
	Missing			7.4	20

Table 1 (Cont.)

Descriptive statistics of the variables used in this study.

CATEGORICAL VARIABLES:	Category	Pct.	N
Treatment Conducted at Criminal Justice Site	Non-CJ Treatment Site	50.9	137
	Mixed CJ and Non-CJ Tx.	5.9	16
	CJ Treatment Site	28.3	76
	Missing	14.9	40
Treatment Conducted in Institutional Setting	Not Institutional Setting	69.1	186
	Institutional Setting	25.3	68
	Missing	5.6	15
Interaction of Academic Affiliation of Author and Whether Treatment is Demonstration Program	Regular Program	62.5	168
	Demo. Prog., Eval. Indepen.	3.3	9
	Demo. Prog., Eval. Influen.	2.6	7
	Demo. Prog., Eval. Planned Treat.	19.7	53
	Demo. Prog., Eval. Delivered Treatment	7.1	19
	Missing	4.8	13
Method of Subject Assignment	Weak Quasi-Experimental	24.9	67
	Strong Quasi-Experimental	26.8	72
	Randomized	46.1	124
	Missing	2.2	6
Coder Confidence in Method of Subject Assignment	Very Low	.4	1
	Low	2.6	7
	Moderate	7.4	20
	High	23.8	64
	Very High	62.5	168
	Missing	3.3	9

Table 2

Multiple regression analysis predicting the Method of Subject Assignment from the Study Context variables.

Criterion Variable: (N=263)		Method of Subject Assignment (Positive Beta indicates correlation with randomized designs.)		
Predictor Variables:				
Entry At:	Variable Name	r	Final ß	t-value
Step 1:	Year of Publication	-.003	-.008	
	Academic Author Affiliation	.170	.134	2.254*
Step 2:	Evaluation Funding	.002	.067	
	Eval. Involv. in Treatment	.329	.348	4.196***
	Demonstration Program	.271	.318	2.032*
	Eval. Involv. x Demo. Prog.	.283	-.260	-1.572^
	Program/Treatment Age	-.094	.067	
	Mandatory Treatment Part.	-.176	-.191	-3.237**
	Missingness Variable, CJ Site	-.099	-.044	
	Criminal Justice (CJ) Site	.014	.175	2.730**
	Institutional Setting	-.034	-.027	
Step 3:	Coder Conf. in Subject Assign.	.258	.208	3.589***

Summary of Results:

Step:	R Sq. Chg.	F Ratio	Sig.	R Sq. Tot.	F Ratio	Sig.
Step 1:	.029	3.904	.021	.030	3.904	.021
Step 2:	.164	5.689	.001	.194	5.479	.001
Step 3:	.040	12.883	.004	.233	6.333	.001

^p≤ .15 *p≤ .05 **p≤ .01 ***p≤ .001

Figure 1.

Hypothetical variables demonstrating a suppressive variable relationship.

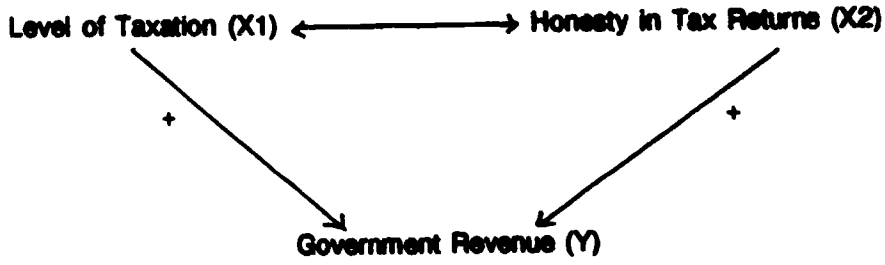

Table 3

Predictor variable correlation matrices identifying suppressive variable relationships.

Correlation matrix for predictor variables exhibiting only positive beta values.

	Evalua. Funding	Affil.* Author	Eval..*** Involve.	Program Age	Demon.* Prog.	CJ** Site
Funding	-----	-.206			-.152	-.122
Affil. Auth.		-----				-.164
Eval. Invl.			-----	-.311		-.145
Prog. Age				-----	-.497	
Demon. Pr.					-----	
CJ Site						-----

Correlation matrix for predictor variables exhibiting only negative beta values.

	Year of Public.	Ev. Inv x Dem. Pr.^	Mand. Tx Part.**	Inst. Setting
Year of Publication	-----	-.110		-.209
Eval. Invol. X Dem.P.		-----	-.120	
Mandatory Tx Part.			-----	
Institutional Setting.				-----

Correlation matrix for predictor variables with positive & negative beta values.

Predictors with positive beta values:

Predictors w. Negative beta values	Evalua. Funding	Affilia.of* Author	Eval.*** Involve.	Program Age	Demon.* Prog.	CJ** Site
Year Pub.				.125		
E.I. x D.P.^		.162	.683		.922	
Mand. Tx.P.**						.252
Inst. Sett.						.388

Note: The significance level of the predictor variable's final beta weight is indicated by

^p≤ .15 *p≤ .05 **p≤ .01 ***p≤ .001.

STAKEHOLDERS	ENTHUSIASM FOR EVALUATION	DESIRED METHOD OF ASSIGNMENT METHOD	"UNACCEPTABLE" ASSIGNMENT METHOD
Evaluator (Including SWRL and WC organizations)	High	Full Randomization	Weak Quasi-Experimental
Administrator #1	High	Full Randomization	Weak Quasi-Experimental
Administrator #2	Low-Medium	Pre-Post or Quasi-Experimental	Randomization by Student
AOD Administrators	Medium-High	Randomization or Quasi-Experimental	None
School Site Administrators	Low	Pre-Post	Randomization by Student
Teachers	Low-Medium	Pre-Post	Randomization by Student

Table 4. Evaluation stakeholders, and their respective positions regarding the implementation of the method of subject assignment.