UAA Justice Center
UNIVERSITY of ALASKA ANCHORAGE

Scholarworks@UA — UAA Justice Center

March 1981

Converging Science and Practice in Analyzing Evaluation Data

Knowlton W. Johnson

Suggested citation

Johnson, Knowlton W. (1982). "Converging Science and Practice in Analyzing Evaluation Data." Paper presented at the annual meeting of the Academy of Criminal Justice Sciences, Louisville, Mar 1981; revised 1982.

Summary

A strategy is presented for converging science and practice which focuses on the needs of scientists and policymakers in analyzing evaluation data. Emphasis is placed on employing powerful statistical techniques that maximize the evaluators' confidence in their results. Attention is also drawn to the need for producing results which can be easily communicated to and interpreted by policymakers. In regard to these requirements, the discussion concerns application of four statistical techniques: factor analysis, Guttman scalogram analysis, multiple classification analysis and cross-break analysis. Each statistical analysis technique is described as to its value in evaluation research for dealing with problems known to inhibit the convergence of science and practice. The application of these techniques is demonstrated by illustrations taken from previous evaluation studies. The paper concludes with implications for stimulating the extent and quality of evaluation use.

CONVERGING SCIENCE AND PRACTICE IN ANALYZING EVALUATION DATA

by

Knowlton W. Johnson

Justice Center
University of Alaska, Anchorage

JC 8219

Paper presented at the annual meeting of the Academy of Criminal Justice Sciences, Louisville, Kentucky, March 1981.

Revised 1982.

*This article is a revision of a paper presented at the Academy of Criminal Justice Sciences meeting, Louisville, Kentucky, March 1981. The author wishes to thank Drs. Sandra Baxter, Charles Beirnard, Richard Bennett, Richard McCleary, Jill McKelvy, and Anne Schneider for their helpful comments on earlier drafts.

ABSTRACT

A strategy is presented for converging science and practice which focuses on the needs of scientists and policymakers in analyzing evaluation data. Emphasis is placed on employing powerful statistical techniques that maximize the evaluators' confidence in their results. Attention is also drawn to the need for producing results which can be easily communicated to and interpreted by policymakers. In regard to these requirements, the discussion concerns application of four statistical techniques: factor analysis, Guttman scalogram analysis, multiple classification analysis and cross-break analysis. Each statistical analysis technique is described as to its value in evaluation research for dealing with problems known to inhibit the convergence of science and practice. The application of these techniques is demonstrated by illustrations taken from previous evaluation studies. The paper concludes with implications for stimulating the extent and quality of evaluation use.

Historically, program evaluation in the human services has been plagued with studies in which researchers have employed weak evaluation designs, faulty measures and inappropriate analysis techniques. Bernstein and Freeman (1975), found that less than 20 percent of the evaluation projects funded by the federal government consistently followed generally accepted procedures in regard to design, data collection and data analysis. Lounsbury et al (1980) and Novaco and Monahan (1980) have shown that research rigor in community psychology is less than desirable, and Bailey (1966) and Logan (1972) have highlighted the lack of rigor in evaluating criminal justice programs. Put most starkly, evaluation in the human services has lacked methodological sophistication.

Recently this absence of research rigor appears to be changing. Beginning in the mid - 1970's there were numerous reports on issues concerning research rigor and on application of sophisticated evaluation designs, data collection procedures and analysis techniques.[1] During this period two evaluation research associations were formed and a number of graduate programs now have an evaluation research concentration. There is a very clear push for defining evaluation research as a distinct discipline.

In reality, this increase in methodological rigor creates a quandary for evaluators and policymakers. On the one hand research sophistication is known to enhance the scientific quality of evaluations (Cook, et al, 1977). Furthermore, scientific quality has been found to be important to policymakers (Weiss and Bucuvalas, 1980). On the other hand, the application of rigorous quantitative methods is known to disrupt the organizational environment within which the evaluations are conducted (Schulberg and Jerrell, 1979) and to produce results that are difficult to communicate to policymakers (Rothman, 1980).

This research rigor predicament has been couched in the past in debates about the appropriateness of experimental methods and quantitative methodologies (Conner, 1981). These debates have clearly reflected either a scientific perspective or a practical point of view. Riechern, et al (1974) and Cook and Campbell (1979), for example, advocate the need for increased rigor through randomized experiments and quantitative methods. Conversely, Gube (1978) and Patten (1980) strongly support a naturalistic inquiry or qualitative evaluation which interfaces with the user's world.

Moving into the 1980's, the methodological interest in evaluation research has shifted to a mix of quantitative and qualitative methods (Cook and Reichardt, 1979; Conner, 1981:8; Bell and Anderson, 1982). This trend focuses attention on the application of two complementary sets of methodologies, one set addressing concerns for scientific rigor and the other set focusing on concerns of the user of evaluation results. In regard to this emerging interest, this paper offers an alternative to mixing quantitative and qualitative methods. Instead of using two different sets of evaluation methodologies, an alternative strategy is to concentrate on rigorous quantitative methods and give special attention to converging science and practice at the analysis stage. The focus is on using a multivariate approach to analyzing evaluation data which takes into consideration the scientific and policymaker's perspective.

In presenting this approach, emphasis is placed on data analysis problems known to inhibit the convergence of scientific and policymaking perspectives. Additionally, particular multivariate techniques are presented which have been found to facilitate the convergence process. The application of these techniques are illustrated in case studies. Most importantly, the question is addressed as to how linking science and practice may potentially impact use and misuse of evaluation results.

Data Analysis Problems Inhibiting Convergence

The schism between evaluators and policymakers appears to be critical at some evaluation stages more than others. While it is well known that the rigorous application of theory and experimental designs create problems for science and practice, our focus is on difficulties stemming from the increased use of sophisticated statistical techniques (Conner, 1979; Silberman, 1980). From the point of view of science, the appropriate application of rigorous methods reduces uncertainty, and thereby enables the evaluator to make fewer ambiguous and inconclusive statements about the results. That is, methodological sophistication can be viewed as increasing the evaluator's confidence and preciseness in reporting the results. From the policymaker's perspective, they cannot use results that are not understood. In short, the evaluator must produce results that are easily communicated to policymakers.

Boruch and Rindshopf (1977), Cook, et al (1980) Schneider, et al (1978) and others have addressed a number of data analysis problems that relate to the issue of the convergence of evaluator and policymaker's needs systems. Of these, we focus our attention on two commonly known concerns, invalid research measures and artificial relationships. In regard to invalid measures, evaluators know that in many instances record data are inadequate for the use of powerful statistical techniques. Therefore, primary data have to be collected about program outcomes and treatment processes in order to use various statistical procedures. To make matters worse, most measurement construction for evaluation research has to begin with only constructs, for there are few standardized measures available; of the available ones, there is a serious problem of transferability.[2] Given these realities of evaluation research, it becomes necessary in many evaluation studies to engage in scale construction.

Both reliability and validity are necessary requisites of scale construction; however, it is felt that the latter, not the former, is the more critical requisite to address in converging scientific and policymaking requirements of evaluation research. From a scientific perspective, the problem of validity can be overcome by establishing content, construct and criteria related validity, i.e., concurrent or predictive checks (Carmines and Zeller 1979; Nunnally, 1978). From the policymaker's point of view, however, such procedures may take too long, require too much of program staff and not be relevant other than as a necessary yardstick for evaluating their program. The question which has to be answered is how to maintain the scientific integrity of evaluation measures when operationalizing policy relevant concepts and simultaneously produce understandable results.

Artificial relationships among variables also inhibit the convergence of the scientific and policymaker's perspectives. Cook, et al (1980:484) addresses this problem in terms of inaccuracy stemming from generating, disseminating and interpreting research findings. In generating results, evaluators have to concern themselves particularly with distortion stemming from nonlinear relations, spuriousness and nonspecification.

Nonlinearity most frequently occurs in an evaluation context when the analysis strategy incorporates the use of techniques with a linearity assumption, e.g., analysis of covariance or various correlational analyses. There are techniques which can accurately describe nonlinear relationships without violating this requirement, but the results are difficult to communicate to policymakers.

Spuriousness may occur in evaluation research when the relationship between an outcome and a particular program process stems, not from a connection between them, but from the fact that each of these is related to some characteristic of the unit being studied or to an element of the environment in which the study takes place or to another program process variable. Unfortunately, evaluators, because of time and funding constraints, often produce results based on only bivariate relationships or limited multivariate relationships. In these instances, distortion is inevitable.

The inference that an assumed relationship is spurious is made only if a third or combination of variables reduces or eliminates the original relationship. However, if the introduction of a third variable leads to intensification of the relationship within one subgroup and reduces it in another, then without this third variable in the analysis, there is a problem of lack of specification. Very seldom do evaluators produce results consisting of conditional relationships that elaborate on the original relationship.

These data analysis concerns regarding artificial relationships are relatively straightforward when scientific requirements are the focus; however, when the policymaker's capacities and needs are taken into consideration, the appropriate analysis strategy is not so obvious. It is well known that policymakers desire to review results which are in summary form and which are easily understandable. Therefore, the evaluator's choice of statistical techniques directly relates to overcoming inaccuracy in disseminating and interpreting evaluation results. This concern raises another important question addressed in this paper, which is how to conduct an examination of relationships, and to present the findings to policymakers in a form that is understandable and is easily translated into practice.

In response to these problems of invalid measures and artificial relationships, a multivariate approach is being offered which takes into account the needs of the evaluator and policymaker. Emphasis is on particular multivariate techniques that can produce results which meet both scientific and practical requirements. This mandate places primary responsibility on evaluators to generate results that are scientifically reliable and valid, approximate reality, and can be easily communicated and interpreted.

Multivariate Analysis In A Policymaking Context

Operationalizing Policy Relevant Concepts

Scale construction procedures are well established, but in many instances in evaluation research both scientific and practical constraints prohibit the use of some validation procedures. For example, the evaluator is often confronted with the lack of an outside criterion in a universe of acceptable controls relating to the qualities to be measured. As Carmines and Zellers (1979) have suggested, the lack of universe of content and outside criteria characterizes most social science research and, therefore, these voids prevent the use of desired validation approach such as concurrent and predictive validity checks. In the case of predictive validity, the longitudinal data requirement also creates practical constraints concerning time and costs.

An alternative validation procedure is to develop research measures that satisfy construct validity requirements. According to Crumbach and Meehl (1955: 282) "[c]onstruct validity must be investigated whenever no criterion or universe of content is accepted as entirely adequate to the quality to be measured." In evaluation research this type of validity would concern the extent to which a particular indicator relates to other indicators consistent with some concept concerning a program outcome or treatment process. Construct validity can be established by using factor analysis or Guttman scalogram analysis. Illustrations of the use of these techniques for this purpose are presented, using evaluation studies in criminal justice.

Factor Analysis Application in Evaluating Programs

The logical foundation of factor analysis rests with a set of various statistical procedures that allow the analyst to search for a linear combination of a set of measures with some underlying factor. (Kim and Mueller, 1978a). A factor analytic approach can be useful as a heuristic device or as an exploratory method. The former use may be thought of as an informal method of ascertaining underlying factor structure of some preconceived conceptual dimension. The latter use of factor analysis can be equated with exploring the data for possible data reduction. We have found that factor analysis can serve as a heuristic device in constructing outcome measures in both formative and summative evaluation studies. It is also particularly useful in formative evaluation as a exploratory technique for reducing a large number of program process variables to a smaller number.

As a heuristic device, we have used factor analysis in a number of evaluative studies to uncover multi-dimensions of a policy relevant construct that was conceptualized originally as unidimensional. For instance, in a formative evaluation of inmate stress in a medium size adult correctional facility, the evaluation team conceptualized inmate stress as a single dimensional construct (Johnson et al, 1979). A factor analysis of self-reported data from 107 inmates showed, however, that stress was not one, but a two dimensional phenomenon, i.e., stress stemming from boredom and stress stemming from fear of being harmed. A subsequent multivariate analysis revealed that particular program staff practices were highly correlated with boredom related stress, whereas not any of the program or organizational process variables were correlated with fear related stress. In this case the results that were presented to policymakers would have been incorrect if we had assumed stress to be an unidimensional phenomena and had simply based the analysis on a single stress index.

Another example of using factor analysis as a heuristic device was in connection with a statewide evaluation of community based treatment programming for juveniles. In this two year formative evaluation project, the evaluation team developed an empirically based evaluation system (Johnson, Rusinko and Girard, 1979:80). Factor analysis results which were produced at different points in time enabled us to regroup items into clusters that were internally consistent across time for behavioral such as youth responsible behavior, and psychological outcomes such as youth's self reliance and staff burnout.

Factor analysis has also been valuable in reducing the number of program process variables to a smaller number. When using this technique for data reduction purposes, the analyst usually does not have a clear conceptual map of the factor structure. As such, the procedure is more of an inductive method than a deductive one. For example, in the statewide child care evaluation project described above, we measured the treatment environment by writing items which, on the basis of content, defined the major treatment modalities that were discussed in the literature (e.g., reality theory, behavioral modification, etc.). Using youth-reported and staff-reported behavioral data on individual youths, a factor analysis was used to define the actual treatment environment by reducing the original number of items to a much smaller number. In some instances, only two items were left to describe a specific element of treatment, e.g., youth's report on staff's use of authority, and in other cases, as many as five items were summed to form a composite indicator, e.g., staff report on level of supported behavior provided to youths.

Using factor analysis to establish construct validity, and subsequently to increase the evaluator's confidence in the results is important to the world of science; however, from a policymaker's perspective, there is a more pragmatic need for the multivariate technique to serve. This need centers on the results being easily communicated to policymakers; that is, they should describe the particular phenomenon being measured. Producing the most descriptive results involves decisions about constructing the composite or scale scores from those indicators that load on the same factor.

According to Kim and Mueller (1978b), there are three ways to construct scale scores using factor analysis results. These approaches are referred to as factor score scaling, component factor scaling and factor based scaling. First, the scale can be based on the values of individual indicators which have been weighted using factor score coefficients. A coefficient is obtained by regressing the factor on its indicator, and is the equivalent of a regression equation's beta weight (Marradi, 1981). It may be argued that this procedure of scale construction is the best scientifically speaking (Marradi, 1981: 70-71); however, Alwin, (1973) and Kim and Mueller (1978b: 70-71) contend that component factor scaling and factor based scaling are appropriate procedures, provided the factor loadings are not grossly different. The component factor technique sums across raw values after each has been weighted by the principal component factor loading. This weighting scheme is like building factor scores, but only zero-order correlation coefficients are used in the regression equation. Factor base scaling simply assumes equal weighting of the indicators with the highest factor loadings being summed to form a scale score.

The Hartwig and Dearing (1979) discussion of location, spread and shape of distributions helps to explain why one scale construction method produces results that are more or less understandable than another one from the policymaker's perspective. Regardless of the scaling procedures used, the shape of the distribution remains the same, but not its location, which is often measured by the mean, or the spread which is often described by the standard deviation. These differences will not distort the results of subsequent analyses; however, there is distortion when relating the scale score to the distributions of individual indicators included in the scale score. That is, the most understandable distribution of scale scores are produced by the factor based scaling procedure which simply sums the raw values of the individual indicators, i.e., equal weighting. The location, spread and shape of this distribution can easily be connected to the individual indicator distributions. Conversely, the location and spread of scale scores are changed from the summated raw scores when component factor and factor scaling methods are employed. In these cases the weighted values are summed; and because these values are smaller, the mean is smaller and the spread is less. As such, it tends to be more difficult to communicate to policymakers the distribution of these weighted scale scores than the unweighted scale score distributions.

We have found in our evaluation work that unless factor loadings or factor score coefficients are grossly different from each other, they are highly correlated. Therefore, because of the ease in connecting the scale score to distributions of individual items included in the scale, we have presented to policymakers the distributions of equal weighted scale scores whenever possible. When the factor loadings are quite different, we have used weights based on the component factor and factor score methods. The principal component factor loadings have been used as weights when a phenomenon was being described as a single dimension. Factor scores have been used whenever there was more than one scale involved and therefore standardization was necessary for comparative purposes. When scales have been standardized the distributions have been presented to decision makers in the form of histograms or line graphs. Regardless of the scaling method used, however, emphasis has been placed on presenting results in an understandable form.

Guttman Analysis and Program Evaluation

In addition to using a factor analytic technique for scale construction, Guttman scalogram analysis has received much attention as a scaling device. Whereas factor analysis orders indicators regarding some underlying factor structure, Guttman analysis orders both indicators and subjects with respect to some common cumulative dimension (McIver and Carmines, 1981). Construct validity can be established by requiring a set of indicators to meet two conditions. First, the scale must be cumulative; that is, items can be ordered by the probability of subjects responding to specific items (degree of difficulty). A person who replies positively to a difficult item will also respond positively to less difficult items and vice versa. The statistic, coefficient of reproductibility, can help the analyst assess the pattern of responses by identifying which items they least likely answered positively. It ranges from 0 to 1 and should be .90 or above (Guttman, 1947). A second condition of Guttman scaling is unidimensionality; that is, items should be measuring the same thing. The coefficient of scalability measures unidimensionality and should be above .60 on a continuum from 0 to 1 (Menzel, 1953).

In a scientific sense, Guttman analysis can produce results that are equally valid as results generated by factor analysis; however, its value as an heuristic or as an exploratory device is not as great as factor analysis. It is well established that Guttman scaling is appropriate for uncovering a set of indicators that operationalizes some policy relevent concept; but because this device is designed as a unidimensional analytical tool, it can not assist the analyst in defining an unanticipated second or third dimension. Further, Guttman analysis is not practical for data reduction because it is not designed to handle a large number of variables.

From a policymaking perspective, Guttman analysis can be viewed as superior to factor analysis in that the former scaling strategy provides more descriptive information than the latter strategy. That is, in addition to producing a description of the distribution of scale scores, Guttman analysis produces the pattern of responses for a given set of indicators. This pattern description stems from cumulative scoring which determines not only how many items a respondent might answer affirmatively, but also which items he/she answered affirmatively. An additional advantage of a Guttman scaling approach is that a composite score is computed by simply adding up yes responses for each dichotomous item being included in the scale score. The results which describe the frequency distribution of the scale are easily communicated to policymakers in that the equal weighting reflects the exact numer of items included in the scale.

The application of Guttman analysis in an evaluation context is illustrated in a study conducted in a County Department of Corrections (Johnson, 1980). In this formative evaluation, policymakers were primarily interested in ways to improve their community outreach programming, and secondarily, they wanted to know the extent of interagency support that could be mobilized for a push for a budget increase. Toward these purposes, we operationalized interagency support in two ways. First, 85 policymakers from 59 human service agencies answered questions that indicated their degree of commitment to a number of hypothetical situations involving money and time donations, public hearing appearances, political contacts and letter writing. Additionally, the department was interested in establishing to whom and to what extent outside agency personnel had shared positive opinions about their contact with the department of corrections.

The scientific requirements of scaling were satisfied for both measures of interagency support by constructing scales with coefficients of reproducibility and scalability above the acceptable .90 and .60 respectively. Policy making requirements were also satisfied by the analysis producing a description of the patterns of actions, ranging from the most likely type to the least likely type of action that respondents would endorse. Further, the Guttman analysis uncovered those respondents who had most likely shared positive opinions to the least likely recipient of such opinions. We found that these results could be easily communicated in group meetings with decisionmakers and could be presented in a report in an understandable form.

In summary, we have shown that factor analysis and Guttman scalogram analysis can increase the confidence of evaluators and also yield results that can be communicated to policymakers. Factor analysis tends to be more powerful than Guttman analysis in uncovering hidden dimensions of originally preconceived policy relevant constructs. This analytical tool can also be used more effectively as a data reduction device than can Guttman analysis. Conversely, Guttman analysis produces more descriptive information about frequencies and patterns of the phenomenon under study than is produced by factor analysis. Further, the Guttman scalogram results are easier to communicate to policymakers than are results from factor analysis.

Remedies to Artificial Relationships

In addition to the problem of invalid measures, evaluation research has to contend with artificial relationships stemming from nonlinearity, spuriousness and the lack of specification. In regard to solutions, several multivariate statistical techniques might be considered. In particular, there are two techniques that we have found useful. These are multiple classification analysis (MCA) which is an extension of an analysis of variance (Sonquist, 1970; Kim and Kohout, 1975:409-410), and crossbreak analysis which combines the analysis of variance and contingency table analysis (Nie et al, 1975).[4] These techniques can handle problems of nonlinearity, spuriousness and nonspecification and can produce results that are easily communicated to policymakers.

MCA and Crossbreak Analysis Application

Multiple classification analysis is based on an extension of an analysis of variance, but without the linearity assumption. MCA can depict the mean value of some outcome variable across a selected number of treatment or process variables. Additionally, this statistical technique can both control for spuriousness and accurately describe nonlinear relationships.

Results from MCA can also be easily communicated to policymakers. The output is in terms of unadjusted and adjusted deviations from the grand mean. These deviations can be easily transformed to subcategory means for presentation of linear or nonlinear patterns. In addition, the policymaker can see the effects on the subcategory means when spurious relationships are taken into account by comparing differences between the unadjusted means and the adjusted means. The overall explanatory power of each independent variable is also calculated as "eta" coefficients, which are without covariate adjustments and "beta" coefficients which are with covariate adjustments. These coefficients have seemed to have little meaning to policymakers, therefore we have in the past presented these results in a table footnote.

MCA does not provide information about interaction effects. What we have done is to conduct preliminary analysis using Automatic Interaction Detection (AID) or Analysis of Variance routines to explore for interaction effects (Sonquist and Morgan, 1964; Kim and Kohout, 1975:399-433). If such effects are found, interaction terms are constructed and entered into the Multiple classification analysis (Sonquist, 1970). These effects are more difficult to communicate to policymakers than the main effects which are the type of results routinely produced by MCA.

An illustration of MCA's handling of nonlinear relationships was in connection with the statewide child care program evaluation described earlier. In one facet of the analysis, we found that group home staff reported they were using positive reinforcements (e.g., allowances, later curfews) to varying degrees in all of the facilities studied. It was assumed that the more positive reinforcements received by youth, the more responsible their behavior would be. When MCA was used to analyze the relationship between this treatment variable and responsible behavior, a policy relevant curvilinear relationship was uncovered. It was found that when staff reported using positive reinforcements sparingly or a great deal, youths reported low involvement in responsible behavior related activities. Conversely, youth who received a moderate number of positive reinforcements reported high involvement in responsible activities. These results indicate that there is a direct relationship between positive reinforcement and responsible behavior, but that there is an optimal level at which positive reinforcement appears to be effective before its effectiveness diminishes.

The importance of controlling for spurious relationships was highlighted in another formative evaluation of witness/victim assistance provided by a county states attorney's office (Johnson et al, 1979). In this evaluation, telephone interviews were conducted with 100 witnesses who had been the key witness in a felony case, and who had some contact with personnel in the states attorney's office. The importance of controlling for spuriousness was realized in a multiple classification analysis that focused on relationships between states attorney's office communication with witnesses and witness satisfaction. When examining the bivariate relationships between witness satisfaction and whether witnesses had been informed of case progress, available services, the trial date and prosecuting attorney and the case outcome, all relationships were significant. However, when these four process variables were entered into an equation simultaneously, only "informed of case progress" and "informed of available services" emerged as significant correlates. Relationships between witness satisfaction and being informed of the trial date and attorney and informed of the case outcome disappeared. In this situation, controlling for spuriousness produced more accurate results to be communicated to policymakers.

Specification may also be policy relevant in evaluation research as it points to conditions under which a treatment has varying degrees of impact. In this instance, an analysis technique should be selected to elaborate on rather than control for these effects. One especially effective technique which we have used to elaborate on hidden effects is a "crossbreak" analysis. This technique is a hybrid of an analysis of variance and contingency table analysis (Nie et al, 1975:266-268). The results appear in a contingency table format, but instead of reporting cell frequencies and percentages, the means of the outcome variables are specified for the subgroup of a third or fourth variable.

Crossbreak analysis was used in a recent impact evaluation of an innovative program in juvenile corrections (Johnson, 1980b). This analysis strategy was considered because of a need to elaborate on particular program effects. The program being evaluated entailed assigning probationers to a probation officer by the school they attended rather than by where they lived in the county. Further, probation officers in the experimental program were provided an office in the high school that youth assigned to his or her caseload attended.

In the evaluation, two nonequivalent groups of probationers, 78 probationers assigned by school and 98 assigned by residency, were compared according to three sets of outcome criteria. These criteria were probation officer's supervision intensity, probationer adjustment in school and probationer community adjustment. In assessing the experimental program's effect on school adjustment, we used crossbreak analysis to elaborate on a favorable program effect on school adjustment as measured by absenteeism; the experimental group had fewer absentees over a school year then did members of the control group. This analysis enabled us to show that this positive impact on absenteeism was greater among probationers with two more previous offenses than those with one or no priors. Given these results, we were able to provide policymakers with more specific information about the program effects.

In summary, MCA and crossbreak analysis techniques have been presented as data analysis procedures which can insure scientific integrity and can produce results which are easily communicated to policymakers. These techniques are appropriate for overcoming problems that create artificial relationships which have been known to plague social science research. Moreover, the results are in the form of mean differences which is also known to be one of the easier types of statistical results to understand. By satisfying these scientific and policymaking requirements in analyzing evaluation data, we have begun to converge the worlds of science and practice.

Conclusions and Implications

It has been argued that evaluation research should focus on both the needs of evaluators and policymakers. In regard to evaluators, we drew attention to data analysis techniques that could be used to produce valid research measures, particularly outcome measures; that could accurately describe nonlinear relationships; and that could control or elaborate on original relationships. The point was also made that policymakers are not as concerned about how results are produced as they are about the type of results being produced. Consequently, we focused our attention on multivariate methods that could produce results which are easily communicated to and interpreted by policymakers. This analysis strategy was suggested as a way of converging the scientific and policymaking perspectives in analyzing evaluation data.

One of the most obvious implications that stems from linking science and practice at the analysis stage is its potential impact on increasing the extent and quality of evaluation use (Cook, et al 1980:492-93).[5] Scientifically reliable and valid results that are easily communicated can be used by policymakers for two major types of decisions: those relating to problem definition, program strength and program implementation; and those pertaining to program impact. In the case of the former type of decisions, policymakers frequently have to group and prioritize problems for various administrative reasons. Further, decisionmakers need to have information on the internal and external dynamics of programs to assess the strength and integrity of program implementation.[6] As illustrated in this paper, scale construction techniques such as factor analysis and Guttman can yield results for these types of decisions. In addition to the scientific value of these results, these multivariate techniques can describe interrelationships and patterns of measures that can be easily translated into practice.

Policymakers are also confronted with the second set of decisions concerning program effects. In regard to evaluation use for such decisions, decisionmakers may, for example, be influenced directly by evaluation findings to modify a program, terminate it or develop a new one. Additionally, results may be used to justify a current program or to substantiate present budget for a program.

It is in connection with these decisions that the quality of use is often compromised. Cook, et al (1980:495-96) discussed obvious instances of misutilization stemming from distorted relationships of the Coleman evaluation of school desegregation effects. We have also found evidence of misuse of research in a recent study concerning the 268 human service agencies in Alaska. This study revealed that it was not uncommon for policymakers to make programmatic changes on the basis of artificial relationships (Johnson, 1982; 1983). Multivariate techniques like multiple classification analysis and crossbreak analysis can assist in reducing misuse of evaluation findings by providing more reliable valid and understandable results for making decisions about program effects.

In conclusion, integrating science and practice in analyzing evaluation data may not be a panacea for improving the extent and quality of evaluation use, but employing strategies for convergence at this stage is a necessary consideration. Moreover, the data analysis techniques offered as central to the strategy presented, while not exhaustive, do illustrate ways for producing accurate results that can be easily communicated to policymakers.

NOTES

[1] See Evaluation Studies: Review Annual (Vol. 1-6) for discussions of research rigor and evaluation research application. In addition, there are a number of evaluation journals that address research rigor and other program evaluation issues. These journals include: Evaluation and the Health Professions; Evaluation News; Evaluation Review; Educational Evaluation and Policy Analysis; New Directions for Program Evaluation; CEDR Quarterly; and Evaluation and Program Planning. In the justice area, Schneider et al (1978) and Klein and Teilman (1980) are illustrative of the application of evaluation methods in this problem area.

[2] Moos (1974; 1975) has attempted to construct varied measures of the treatment environment in correctional setting. However, in different evaluation projects concerning residential group homes for juveniles, a halfway house for adults and a residential drug treatment program for hard core users, we found that Moos's subscales could not survive a factor analytic approach to scale construction.

[3] Kim and Mueller (1978a:9; 1978b:46-54) also discuss confirmatory factor analysis which is a formal method of ascertaining underlying factor structure.

[4] Medler (1978) has an excellent discussion of the application of multiple classification analysis in evaluation research. A crossbreak analysis technique is an alternative to continuency tables analysis involving more than two variables. This technique has been traditionally used to specify conditions of relationships seen in a two variable table (Selitiz et al, 1951; Rosenberg, 1968; Johnson, 1981: 315-342).

[5] There is an increasing body of empirical work on the extent and nature of research use. For examples, see Caplan (1976); Weiss (1977); Johnson (1980).

[6] The major questions relating to treatment strength and integrity are: was the program sufficient in strength and was the program design implemented as planned? Research concerning these questions can be found in Sechrest, et al (1979) and Yeaton and Redner (1981).

REFERENCES

ALWIN, D. F.

1973 "The Use of Factor Analysis in the Construction of Linear Composites is Social Research." Sociological Methodology and Research 35: 191-214.

BAILEY, W. C.

1966 "Correctional Outcome: An Evaluation of 100 Reports." Journal of Criminal Law, Criminology and Police Science. 57: 153-60.

BELL, M.

1982 "Toward A Unified Methodology for Evaluation: The Qualitative - Quantitative Continuum." Paper presented at the 1982 Joint Meeting of Evaluation Network and Evaluation Research Society, Baltimore, MD.

BERNSTEIN, I. AND H. E. FREEMAN

1975 Academic and Entrepreneurial Research: Consequences of Diversity in Federal Evaluation Studies. New York: Sage.

BORUCH, R. F. AND D. RINDSKOPF

1977 "On Randomized Experiments, Approximations to Experiments, and Data Analysis", in Leonard Rutman (ed.) Evaluation Research Methods: A Basic Guide. Beverly Hills, Calif: Sage.

CAPLAN, N.

1976 "Social Research and National Policy: What Gets Used By Whom, For What Purpose and With What Effects?" Int. Soc. Sci: 28: 187-194.

CARMINES, E. G. AND R. A. ZELLER

1979 "Reliability and Validity Assessment. Beverly Hills, Calif: Sage.

CONNER, R. F.

1981 "Editor's Introduction," in R. F. Conner Methodological Advances in Evaluation Research. Beverly Hills, CA: Sage, pp 7-10.

CONNER, R. F.

1979 "The Evaluator-Manager Relationship, An Examination of the Source of Conflict and a Model for a Successful Union", in Herbert C. Schulberg and Jeanette M. Jerrell (ed.) The Evaluator and Management. Beverly Hills, CA: Sage.

COOK, T. D. AND D. T. CAMPBELL

1979 Quasi-Experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally.

COOK, T. D., F. L. COOK AND M. M. MARK

1977 "Randomized and Quasi-Experimental Designs in Evaluation Research: An Introduction" in Leonard Rutman (ed.) Evaluation Resarch Methods: A Basic Guide. Beverly Hills, DA: Sage, pp 103-35.

COOK, T.D., J. LEVINSON-ROSE AND W. E. POLLARD

1980 "The Misutilization of Evaluation Research: Some Pitfalls of Definition." 1(4) Knowledge: Creation, Diffusion, Utilization, pp. 477-98.

COOK, T. D. AND C. S. REICHARDT

1979 Qualitative and Quantitative Methods in Evaluation Research. Beverly Hills, CA: Sage.

CRONBACK, L. J. AND P. E. MEEHL

1955 "Construct Validity in Psychological Tests." Psychological Bulletin. 52: 281-302.

DAVIDSON, W. S., J. R. KOCH, R. G. LEWIS AND M. D. WRESINSKI

1981 Evaluation Strategies in Criminal Justice. Elmsford, N. J.: Pergamon.

GUBA, E.G.

1978 Toward A Methodology of Naturalistic Inquiry in Educational Evaluation. CSE Monograph series in Evaluation 8. Los Angeles:

Center for the Study of Evaluation, University of California, Los Angeles.

GUTTMAN, L.

1947 "The Cornell Technique for Scale and Intensity Analysis." Education and Psychology Measurement. 7: 247-279.

HARTWIG, F. AND B. E. DEARING

1979 Exploratory Data Analysis. Beverly Hills, Calif: Sage.

JOHNSON, E. S.

1981 Research Methods in Criminology and Criminal Justice. Englewood Cliffs, N. J: Prentice-Hall, Inc.

JOHNSON, K. W.

1983 "Confronting Violence in Alaska: Use of Research in Planning For Change," in K.W. Johnson, Prevention and Control of Violence in Alaska: Conference Proceedings. Anchorage: Justice Center, University of Alaska.

1982 "Utilization of Research in Combating Violence in Alaska: An Ecological Perspective." Paper presented at the 1982 Joint Meeting of Evaluation Network and Evaluation Research Society. Baltimore, MD.

1980a An Examination of Community Outreach Programs Initiated by Prince George's County Department of Corrections. Fairfax, Va: International Training, Research and Evaluation Council.

1980b Evaluation of An Experimental School Juvenile Counselor Program. Fairfax, Va: International Training, Research and Evaluation Council.

1980c "Stimulating Evaluation Use by Integrating Academia and Practice." Knowledge: Creation, Diffusion, Utilization, Vol. 2, No. 2, December 1980, 237-62.

JOHNSON, K. W., W. T. RUSINKO AND C. M. GIRARD

1980 "The Development of an Evaluation System for Community-Based Child Care Programs," in Barbara Price and Phyllis Jo Baunach (ed.) Research Models and Findings in Criminal Justice. Sage Research Progress Series, Sage.

1979 "Group Home Evaluation System Development Project." Fairfax, Va: International Training, Research and Evaluation Council.

JOHNSON, K. W., L. NORTON, M. BINDER, R. HOLT, L. HORNER AND E. KARANDY

1979 An Examination of Witness Assistance Provided by the State's Attorney's Office of Prince George's County. Upper Marlboro, Maryland: Office of Budget and Programming.

JOHNSON, K. W., L. NORTON, J. MENGEL AND J. TOBLAS

1979 "An Examination of the Response of Prince George's County Detention Center to Inmate Stress.

KIM, J. AND C. W. MUELLER

1978a Introduction to Factor Analysis. Beverly Hills, Calif: Sage.

1978b Factor Analysis, Statistical Methods and Practical Issues, Beverly Hills, Calif: Sage.

KIM, J. AND F. J. KOHOUT

1975 "Analysis of Variance and Co-Variance: Sub Programs ANOVA and Oneway", in N. H. NIE, C. H. Hull, J. G. Jenkins, K. Steinbrenner and D. H. Bent (ed.) Statistical Package for the Social Sciences. New York: McGraw Hill.

KLEIN, M. W. AND K. S. TEILMANN

1980 Handbook of Criminal Justice Evaluation. Beverly Hills, Calif: Sage.

LOGAN, C. H.

1972 "Evaluation of Research in Crime and Delinquency: A Reappraisal", in Susette M. Talarice (ed.) Criminal Justice Research, Approaches, Problems and Policy. Cincinnati: Anderson.

LOUNSBURY, J. W., D. S. LEADER, E. P. MEARES AND M. R. COOK

1980 "An Analytic Review of Research in Community Psychology." American Journal of Community Psychology. 8: #4 415-441.

MARGOLIS, H.

1973 "Technical Advice on Policy Issues." Sage Professional Paper in Administrative and Policy Studies. Beverly Hills, CA: Sage.

MARRADI, A.

1981 "Factor Analysis As An Aid in the Formation and Refinement of Empirically Useful Concepts", in David Jackson and Edgar F. Borgatta (ed.) Factor Analysis and Measurement in Sociological Research. Beverly Hills, CA: Sage.

McIVER, J. P. AND E. G. CARMINES

1981 Unidemensional Scaling. Beverly Hills, CA: Sage.

MEDLER, J.

1978 "Prediction Methods", in Anne L. Schneider, Peter R. Schneider, L. A. Wilson, William R. Griffith, Jerry F. Medler, Howard F. Feinman (ed.) Handbook of Resources for Criminal Justice Evaluators. Washington, D. C: National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U. S. Department of Justice.

MELTSNER, A.

1976 Policy Analysts in The Bureaucracy. Berkeley: University of California Press.

MENZEL, H.

1953 "A New Coefficient for Scalogram Analysis." Public Opinion Q 17: 268-280.

MOOS, R. H.

1975 Evaluating Correctional and Community Settings. New York: John Wiley and Sons.

1974 Community Oriented Programs Environment Scale. Palo Alto: Consulting Psychologists Press, Inc.

NIE, N. H., C. H. Hull, J. G. JENKINS, K. STEINBRENNER AND D. H. BRENT

1975 Statistical Package For The Social Sciences. New York: McGraw Hill.

NOVACO, R. W. AND J. MONAHAN

1980 "Research in Community Psychology: An Analysis of Work Published in the First Six Years of the American Journal of Community Psychology." American Journal of Community Psychology. Vol. 8, No. 2, 131-144.

NUNNALLY J. C.

1978 Psychometric Theory. New York: McGraw-Hill.

PATTON, M.Q.

1980 Qualitative Evaluation Methods. Beverly Hills: Sage

RICH, R. F.

1979 The Power of Social Science Information and Public Policy-Making. San Francisco: Jossey-Bass.

RIECKEN, W. R., R. F. BORUCH, D. T. CAMPBELL, N. CAPLAN, T. K. GLENNAN, JR., J. W. PRATT, A. REES AND W. WILLIAMS

1974 Social Experimentation: A Method for Planning and Evaluation Social Intervention. New York: Academic Press.

ROSENBERG, M.

1968 The Logic of Survey Analysis. New York: Basic Books.

ROTHMAN J.

1980 Using Research in Organizations, A Guide to Successful Application. Beverly Hills, CA: Sage.

SCHNEIDER, A. L., P. R. SCHNEIDER, L. A. WILSON II, W. R. GRIFFITH, J. F. MEDLER AND H. F. FEINMAN

1978 Hand book of Resources for Criminal Justice Evaluators. Washington, D. C: National Institute of Law Enforcement and Justice, Law Enforcement Assistance Administration, U. S. Department of Justice.

SCHULBERG, H. C. AND J. M. JERRELL

1979 The Evaluator and Management. Beverly Hills, CA: Sage.

SECHREST, L., S. G. WEST, M. A. PHILLIPS, R. REDNER, W. YEATON

1979 "Introduction" in L. Sechrest, S. G. West, M. A. Phillips, R. Redner, W. Yeaton Evaluation Studies Review Annual. Beverly Hills, CA: Sage, pp 15-35.

SILBERMAN, G.

1980 "Recent Advances in Evaluation Methods" in M. W. Klein and K. S. Teilmann (ed.) Handbook of Criminal Justice Evaluation. Beverly Hills, CA: Sage.

SELLTIZ, C. M. JAHODA, M. DEUTSCH AND S. COOK

1951 Research Methods in Social Relations. New York, NY: Holt, Rinehart and Winston.

WEISS, D. G.

1977 Using Social Research in Public Policy-Making. Lexington, MA: D.C. Heath.

WEISS, C. H. AND M. J. BUCUVALAS

1980 "Truth Tests and Utility: Decision-makers' Frames of Reference for Social Science Research," 45 (2) American Sociological Review, 302, 13.

YEATON, W. H. AND R. REDNER

1981 "Measuring Strength and Integrity of Treatments: Rationale, Techniques, and Examples" in R. F. Conner _Methodological Advances in Evaluation Research._ Beverly Hills, CA: Sage, pp 61-87.