DOCUMENT RESUME

ED 229 429 TM 830 330

AUTHOR	Thompson, Bruce; King, Jean A.
TITLE	Evaluator Types: Do Differences Matter?
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	81
GRANT	NIE-G-80-0082
NOTE	21p.; Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Administrators; Competence; *Educational Researchers; *Evaluation Methods; *Evaluators; *Experimenter Characteristics; Higher Education; Influences; Political Issues; *Researchers; Simulation
IDENTIFIERS	*Evaluator Characteristics; *Evaluator Types (Meltsner)

ABSTRACT

The study investigated whether school administrators (n=128) implicitly perceive the "types" of evaluators proposed by Meltsner (1976) when they consider evaluative information. One simulated evaluation report was written to represent each evaluator, differentiated from one another by their norms, expectations, motivations, and training. Two pilot studies were conducted to assess the validity of the simulation materials. Graduate students were asked to read all four reports to identify which "type" each report represented, and isolate which report features led to their decisions. Administrators were then asked to indicate: (1) the similarity of the report to their school system reports; (2) evaluator fairness; (3) technical competence of the evaluator; and (4) evaluator awareness that political considerations sometimes affect program decisions. Finally, administrators were asked to make a policy recommendation regarding the program's future. Results indicate that administrators do perceive evaluator types when they consider evaluative information presented in reports. However, the results suggest that administrators are more attentive to the technical merit of reports than they are to the evaluator's political sensitivity, as indicated by the homogeneity of the subjects' perceptions of the political sensitivity of the various types of reports. This attention to the technical has been suggested in earlier research. (Author/PN)

EVALUATOR TYPES:

DO DIFFERENCES MATTER?

Bruce Thompson

University of New Orleans 70148

Jean A. King

Tulane University 70118

Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, 1981. The research reported here was funded in part by a grant from the National Institute of Education (NIE-G-80-0082); however, no endorsement by the Institute is implied.

ABSTRACT

The study investigated whether administrators implicitly perceive the "types" of evaluators proposed by Meltsner (1976) when they consider evaluative information. One simulated evaluation report was written to represent each evaluator type. Two pilot studies were conducted to assess the validity of the simulation materials. The results suggest that administrators may implicitly employ the typology. The ecological validity of the results was also empirically investigated.

The literature on program evaluation contains numerous desultory comments regarding the extent to which administrators use evaluative information (Thompson & King, 1981, pp. 3-4). For example, Cronbach et al. (1980, p. 3) lament that "commissioners of evaluations complain that the messages from evaluation are not very useful, while evaluators complain that the messages are not used." Stake (1973, p. 314) has gone so far as to say that "we do not know whether or not evaluation is going to contribute more to the problems of education or to the solutions." Alkin and Daillak (1979, p. 41) sum up the situation by noting that "there have been great hopes for evaluation, not only among evaluators themselves, but also among other educators, elected officials, and the public. Yet these hopes have dimmed."

Unfortunately, the non-use of evaluative information, when that use would be appropriate, respresents an enormous waste of effort (Datta, 1979, p. 22), an enormous waste of money (Patton, 1978, p. 12), and results in less than optimal service being provided to program clients. This situation is tragic because, as Wise (1980, p. 16) notes, "no one else is given the resources and time to question, observe, assess, weigh, probe, and reflect that the evaluator is given." Consequently, several researchers have attempted to identify the factors which do or do not affect use levels.

Several factors which apparently do not effectively promote utilization have been identified. For example, it is now clear that the numerous evaluation "models" (cf. Stufflebeam, Foley, Gephart, Guba, Hammond, Merriman & Provus, 1971) are not panaceas for promoting use. The models have been useful to an emerging discipline trying to conceptualize its role, but the models are not very useful guides (nor perhaps were they meant to be) for evaluation practice. Thus Airasian (1974, p. 148) notes that "the many evaluation models advanced in recent years suggest what decisions need to be made but not how they should be made." Based on his case studies of evaluative practice, Alkin (1979, p. 7, emphasis in original) found that "none of the five cases involved the application of a formal evaluation model." As Brown (1980, p. 4) summarized the matter:

> For a time, it was hardly respectable to be an evaluator without having your own model. You at least had to be a disciple of a proponent of a new model that was on the "cutting edge" in order to maintain some semblance of self-esteem. It is interesting to observe that there were very few wounds inflicted by that "cutting edge."

It has also become clear that the methodological aspects of evaluations are not the primary determinants of use. In fact, there is some empirical evidence (Brown & Newman, in press) that inclusion of inferential statistics in evaluation reports even tends to lessen use. Although this result may

have partially been a function of the manner in which the inferential statistics were presented (Thompson, 1981a, p. 6), nevertheless, a growing body of evidence suggests that "the major barriers to successful evaluation are not technical and methodological, though these are certainly important and worthy of further effort, but are rather the structural constraints and requirements and the interpersonal relationships which characterize the evaluation endeavor" (Gurel, 1975, pp. 27-28).

One of the most critical determinants of evaluation use discussed in the literature is what Patton (1978) has termed "the personal factor." As Cronbach et al. (1980, p. 6) summarize, "nothing makes a larger difference in the use of evaluations than the personal factor--the interest of officials in learning from the evaluation and the desire of the evaluator to get attention for what he knows." Evaluators who are sensitive to this factor will respond in two ways. First, following the suggestion of Patton (1978), evaluators will identify the evaluation's relevant information users, and they will continually work on affecting utilization throughout the course of the evaluation. Second, evaluators will also attempt to establish close working relationships with evaluation clients.

One theoretical framework regarding these interpersonal relationships has been offered by Meltsner (1976), who argues that evaluators are differentiated from each other by their norms, expectations, motivations, and training. During the early 1970's Meltsner and a colleague interviewed 116 federal policy analysts. The data from these interviews suggested that evaluators might be differentiated from each other primarily by how much analytical skill they possess and by how much political skill they have developed. This conceptualization led to a fourfold typology of evaluators.

"Entrepreneurs" are characterized as having both good political and good technical skills. "Technicians" are characterized as having good technical skills and less satisfactory political skills. "Politicians" are characterized as having good political but weak technical skills. "Pretenders" are characterized as being weak on both skill dimensions.

Some research suggests that evaluators implicitly perceive themselves in terms of Meltsner's conceptual model (Thompson, 1980a, 1980b). However, it has not yet been demonstrated that the model helps to explain how school administrators perceive evaluators. The research reported here was conducted to investigate this possibility.

Specifically, the study addressed three questions. Do school administrators implicitly use an evaluator typology (Meltsner, 1976) when they consider evaluation reports? Do these perceptions influence the policy recommendations which the administrators offer? Finally, how do perceptions of report technical quality and political sensitivity interact with each other and other variables?

Method

Subjects

The subjects (n = 128) in the study were principals in public school systems in the Southern United States. Only principals were included to avoid any confounding of results that might have been introduced by variations in administrative perspective. Furthermore, principals form an important evaluation audience which has not been considered frequently enough in previous evaluation research. There have been use studies involving principals as subjects (Granville, 1977; Hamilton, 1980), but the number of such studies does not seem proportional to the influence which principals may exert over school decision-making (Lipham, 1980, p. 3).

Materials

The study used simulated evaluation reports for three reasons: first, evaluators' contact with principals' frequently takes the form of such written reports; second, the variables of interest could be carefully controlled; and

third, earlier simulation studies suggest the value of this approach (see, for example, Newman, Brown & Braskamp, 1980). While it is true that principals' perceptions of an evaluator typology will be evidenced primarily in their reactions to practicing evaluators, their responses to reports reflecting differences may also provide evidence for the validity of the typology. With this in mind, an evaluation report was written as an exemplar of each of the four evaluator perspectives proposed by Meltsner (1976). The reports were similar in content--they all dealt with the same hypothetical program--but varied in style, e.g., in the amount of technical information included and in the number of references to specific individuals.

Two pilot studies were conducted in order to refine and validate the study's simulation materials. In the first pilot study 23 graduate students were taught the "types" and were then asked to read all four reports, attempt to identify which "type" each report represented, and isolate which report features led to their decisions. Based upon these results the four reports were revised and a second pilot study was conducted. Ten different graduate students from a different university were taught the "types" and were then asked to identify which "type" each report represented. The number of correct identifications for the "entrepreneur," "technician," "politician," and the "pretender" were, respectively, seven,

nine, nine, and eight. All four values represent statistically significant deviations from expected chance performance ($p < .01$). The numbers of correct identifications were evaluated using a binomial test distribution with the a priori probability of a correct identification being considered .25 (Freund, 1971).

Procedure

Each subject was randomly assigned one "type" of report. Subjects were mailed the reports, asked to answer five brief questions, and then return the responses in stamped, pre-addressed envelopes. Specifically, the subjects were asked to indicate how similar the report was "to evaluation reports written about programs in my school system" (hereafter labelled "similarity"), whether "the report gives me the impression that the evaluator was attempting to be fair," whether "the report gives me the impression that the evaluator has the technical competence to address evaluation issues," and whether "the report gives me the impression that the evaluator recognizes that political considerations sometimes affect decisions about programs." In order to maximize reliability, responses were gathered using a nine-interval Likert scale (Thompson, 1981b). Finally, the subjects were asked to make a policy recommendation regarding the program's future. Participation was anonymous and 128 usable responses represented a response rate of roughly 40%.

Results

The first research question posed in the study asked whether school administrators implicitly employ Meltsner's (1976) evaluator typology when they consider evaluative information. The question was addressed by conducting a discriminant function analysis. The four report types constituted a nominally-scaled criterion variable. Subjects' ratings of report "similarity," fairness, technical quality and political sensitivity constituted the predictor variables. One statistically significant ($\chi^2 = 26.9$, $df = 12$, $p < .01$) discriminant function was identified.

However, the preponderance of the function's predictive efficiency was contained in the technical-quality variable. This was indicated in several ways. First, only this variable had a statistically significant univariate F-ratio ($F = 4.4$, $df = 3,119$, $p < .01$). Second, the variable's structure coefficient (Thompson & Frankiewicz, 1979) was -.7; the structure coefficients for all the other predictor variables were less than an absolute value of .1.

A numerically higher rating on the technical-quality variable represents perception of greater technical merit. The mean ratings for the "technician" and "entrepreneur" reports were, respectively, 7.0 ($SD = 2.2$) and 6.7 ($SD = 2.2$). The mean ratings for the "politician" and "pretender" reports were,

respectively, 5.5 (SD = 2.8) and 5.0 (SD = 2.6).

The second question posed in the study asked whether perceptions of evaluator type influenced administrators' policy recommendations. There were no noteworthy variations in policy recommendations associated with which reports the subjects received. However, this result may have been an artifact of the fact that most of the respondents (86%) recommended that the portrayed program be modified to determine whether it could be made more effective.

The third question posed in the study asked how perceptions of report technical quality and political sensitivity interact with each other and with other variables. This question was addressed by computing bivariate correlation coefficients. They are presented in Table 1. All the values are statistically significant ($p < .01$).

INSERT TABLE 1 ABOUT HERE.

Discussion

One distinguishing feature of the research was an effort to determine the ecological validity of the results. Subjects were asked to indicate whether the study's reports were similar to reports produced in their school system. Approximately 61% of the subjects agreed that the reports were typical. Furthermore, all analyses were reconducted eliminating subjects

who felt the reports were atypical. The findings were unchanged. This suggests that the study's results have some generalizability.

The results of the analysis addressing the study's first research question suggest that administrators do perceive evaluator types when they consider evaluative information presented in reports. However, the results suggest that administrators are more attentive to the technical merit of reports than they are to the evaluator's political sensitivity, as indicated by the homogeneity of the subjects' perceptions of the political sensitivity of the various types of reports. This attention to the technical has been sugested in earlier research. In summarizing a series of simulation studies, Newman, Braskamp, and Brown (1980) note that evaluation reports containing jargon and data--both of which make a report more technical--were generally rated more useful by readers, regardless of their field. The failure to find differences on the political-sensitivity variable may have been a function of the study's design. Administrators may formulate impressions of evaluator technical ability on the basis of evaluation reports; they may prefer to judge evaluator political sensitivity on the basis of personal interactions. This possibility remains to be explored in future research.

The results addressing the study's second research question are ambiguous regarding the study's theoretical underpinnings, but may offer a valuable comment on simulation research. The failure of more subjects to offer a policy recommendation other than program refinement may reflect a hesitancy to make "negative" recommendations on the basis of brief simulation reports. Simulation researchers confront a dilemma. They can write longer, more detailed simulation reports which may better generalize to natural ecologies, but which may take more time to administer and may lower response rates. Or researchers can write brief reports which inherently will not represent very important dynamics affecting administrator decisioning (see Patton, 1978, especially p. 266). This does not mean that simulation research should be abandoned, but its limits must certainly be recognized.

The results addressing the study's third research question are important theoretically. The primary finding is that perceptions of technical quality and political sensitivity are reasonably independent (r squared = .04). This finding is noteworthy because it suggests that the two dimensions of Meltsner's typology are unrelated, as postulated.

In summary, the results represent an initial attempt to determine whether school administrators perceive evaluators in terms of the typology proposed by Meltsner (1976). The results

are not fully conclusive, but generally indicate that the model may be valid. Thus, the model may provide valuable assistance to efforts to understand what Gurel (1975, pp. 27-28) has termed the structural and interpersonal constraints on the evaluation endeavor.

References

Airasian, P.W. Designing summative evaluation studies at the local level. in W.J. Popham (Ed.), Evaluation in education: Current applications. Berkeley: McCutchan, 1974.

Alkin, M.C. Using naturalistic research for the study of evaluation utilization. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1979. (ERIC Document Reproduction Service No. ED 170 378)

Alkin, M.C., & Daillak, R.H. A study of evaluation utilization. Educational Evaluation and Policy Analysis, 1979, 1, 41-49.

Brown, R.D. The evaluator as artist: Primitive, impressionistic, or paint-by-number? CEDR Quarterly, 1980, 13, 4-6.

Brown, R.D., & Newman, D.L. An investigation of the effect of different data presentation formats and order of arguments in a simulated adversary evaluation. Educational Evaluation and Policy Analysis, in press.

Cronbach, L.J., Ambron, S.R., Dornbusch, S.M., Hess, R.D., Hornik, R.C., Phillips, D.C., Walker, D.F., & Weiner, S.S. Toward reform of program evaluation. San Francisco: Jossey-Bass, 1980.

Datta, L. O thou that bringest the tidings to lions: Reporting the findings of educational evaluations. Paper presented at the annual Johns Hopkins University National Symposium on Educational Research, Baltimore, 1979.

Freund, J.E. Mathematical statistics (2nd ed.). Englewood Cliffs: Prentice-Hall, 1971.

Granville, A.C. An experimental study of factors that influence school decisions: Implications for evaluation. Paper presented at the annual meeting of the American Educational Research Association, New York, 1977. (ERIC Document Reproduction Service No. ED 139 824)

Gurel, L. The human side of evaluating human service programs: Problems and prospects. In M. Guttentag & E.L. Struening (Eds.), Handbook of evaluation research (Vol. 2). Beverly Hills: Sage, 1975.

Hamilton, G.E. Title I evaluation from the perspective of the local school. Paper presented at the annual meeting of the American Educational Research Association, Boston, 1980.

Lipham, J.M. Change agentry and school improvement: The principal's role. In Interorganizational arrangements for collaborative efforts. Portland, OR: Northwest Regional Educational Laboratory, 1980.

Meltsner, A.J. Policy analysts in the bureaucracy. Berkeley: University of California Press, 1976.

Newman, D.L., Brown, R.D., & Braskamp, L.A. Communication theory and the utilization of evaluation. In L.A. Braskamp & R.D. Brown (Eds.), Utilization of evaluative information. San Francisco: Jossey-Bass, 1980.

Patton, M.Q. Utilization focused evaluation. Beverly Hills: Sage, 1978.

Stake, R. Evaluation design, instrumentation, data collection, and analysis of data. In B.R. Worthen & J.R. Sanders (Eds.), Educational evaluation: Theory and practice. Worthington, OH: Charles A. Jones, 1973.

Stufflebeam, D.I., Foley, W.J., Gephart, W.J., Guba, E.G., Hammond, R.L., Merriman, H.D., & Provus, M.M. Educational evaluation and decision making. Itasca, IL: Peacock, 1971.

Thompson, B. Comparison of two strategies for collecting Q-sort data. Psychological Reports, 1980, 47, 547-551. (a)

Thompson, B. Validity of an evaluator typology. Educational Evaluation and Policy Analysis, 1980, 2, 59-65. (b)

Thompson, B. Administrators' perceptions of various evaluation report styles. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, 1981. (a)

Thompson, B. Factor stability as a function of item variance. Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, 1981. (b)

Thompson, B., & Frankiewicz, R.G. CANON: A computer program which produces canonical structure and index coefficients. Educational and Psychological Measurement, 1979, 39, 219-222.

Thompson, B., & King, J.A. Evaluation utilization: A literature review and research agenda. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, 1981.

Wise, R.I. The evaluator as educator. In L.A. Braskamp & R.D. Brown (Eds.), Utilization of evaluative information. San Francisco: Jossey-Bass, 1980.

Table 1
Correlation Coefficients

	F	T	P
Fairness (F)	---		
Technical Quality (T)	.68	---	
Political Sensitivity (P)	.25	.21	---