DOCUMENT RESUME

ED 109 167 TM 004 614

AUTHOR Wick, John W.
TITLE On Evaluating a Project: Some Practical Suggestions. NCME Measurement in Education, Vol. 6, No. 1.
INSTITUTION Michigan State Univ., East Lansing. Office of Evaluation Services.
PUB DATE 75
NOTE 9p.
AVAILABLE FROM Office of Evaluation Services, Michigan State University, East Lansing, Michigan 48823 (Yearly Subscription Rate $2.00, single copy $0.50, single copies when quantity is 25 or greater $0.35)

EDRS PRICE MF-$0.76 PLUS POSTAGE. HC Not Available from EDRS.
DESCRIPTORS *Cost Effectiveness; Data Collection; *Educational Objectives; *Evaluation Criteria; Evaluation Needs; Feedback; *Formative Evaluation; Program Administration; Program Costs; *Program Evaluation; Program Improvement

ABSTRACT

Prime indicators for realistic short term/long term project goals are budgets and timetables. Concrets, identifiable objects are useful in separating eloquent rhetoric from actual promises. Similarly, an external evaluator should be able to separate proposals with intentional misrepresentation of funding and goals from those which need further organization. Once a project begins, the evaluator should know whether the data being collected and analyzed will be used for internal public consumption, external public relations, or both. This may depend on whether the evaluators' primary allegiance is to the funding agency or to the project. In any evaluation traditional staff roles and lines of authority should be recognized and better communication facilitated. Technical expertise and the political realities of a system should be reconciled. (BJG)

Vol. 4, No. 1

On Evaluating A Project: Some Practical Suggestions[1]

John W. Wick

ABOUT THIS REPORT

This article gives some very practical advice to the external evaluator on potential pitfalls in program evaluation. Many factors relating to the success or failure of a project can be found in the proposal itself. Schedule, budget, lines of authority and staff commitments all have important implications. Examine what the proposal promised to do. Determine in advance what type of report is wanted—one of critical evaluation for internal consumption or an external one for public relations, or both.

Dr. John W. Wick, Associate Professor of Education at Northwestern University, from his own experience as an evaluator lists five basic characteristics important to a good evaluator: 1) gather sufficient baseline data, 2) collect data continuously, 3) provide low cost feedback in understandable format, 4) use trend analysis, and 5) sample where possible.

Dr. Wick has been active professionally in areas generally concerned with educational measurement, evaluation, testing and statistics. He is the author and co-author of numerous articles and books in his field.

CJF

This article is directed toward an external evaluator. This could be someone operating in the common situation wherein competitive bids are required for the external evaluation of a project. But "external" could also apply to a line or staff person within an agency, serving as a monitor for projects operating in another area of the bureaucracy, or to a staff person in a funding agency monitoring outside projects. An "external" evaluator is one who wishes to evaluate a project for some reason, and who is not directly involved with the operation of the project. This does not mean that the suggestions and comments included herein could not be used by people internal to the project. They are simply not the target audience I am attempting to address.

Short Term/Long Term: The Impact of the Project

The Assistant Superintendent of an elementary school district noticed two interesting things: First, some of the students were not making progress at learning to read as fast as he hoped they would; and second, some of the teachers weren't particularly good at diagnosing the specific problems these children were having. Being a good grantsman, he wrote a proposal to an agency and subsequently received funding for a project which was designed to solve both problems.

Soon a project administrator was hired; testing people were brought in; and a reading specialist was assigned to each ten teachers. The reading and testing specialists were to help the teachers diagnose student problems. The Assistant Superintendent was happy; the school board and townspeople were impressed; the salaries of a number of people were augmented; and all seemed right in the world.

Then the money ran out.

The administrators moved on to other "soft money"; the consultants too searched for other projects; the teachers started misdiagnosing or ignoring

[1] Much of this article is based on Chapter 12 of *EDUCATIONAL MEASUREMENT: Where are we going and How will we know when we get there?* (Columbus, Ohio. Charles Merrill Publishers, 1973).

problems again; and the old reading problem made a dramatic reappearance.

The project was a failure. Or was it?

That depends on whether you're talking about short-term benefits or long term changes. A good evaluator must keep these two separate.

In the short term, a lot of students obviously benefited from the program. The teachers and district personnel probably learned from the procedures. Some interest, excitement, espirit de corps, and activity were generated in the district. And the program administrators, staff, and consultants did all right, too.

In the long run, the procedures died with the project funding, but maybe some positive things did occur. Maybe the funding agency learned that such an approach was not feasible without support external to the district.

A first and very important decision by an evaluator is this one: Are *these* funds, time, and effort being expended to help or to change *these* students or *these* teachers or *these* administrators? Or are long term changes envisioned—changes which will live long after the project funding ends? If the stated goal of the hypothetical project described above was a long-term change, it must be termed a failure. If the project outline only covered the particular student and teacher population which existed at the time of the funding, it did not fail—assuming each administrator, specialist in reading or testing, and consultant did, in fact, "do his thing" well. Most funded projects must be viewed as having some goals which are primarily of the long-term variety. These long-term goals may not be particularly explicit in the proposal, but they are frequently implicit in the relatively large amount of money spent on a small number of people. That is, when a district suddenly begins to spend 50 percent more per pupil on a certain group of students for a three year period—then turns the water off—it must be assumed they had something long-term in mind. If not, the district probably would have spread the money equally among all of the pupils.

How can you predict whether or not the project will have a long-term impact? Some warning lights for projects which probably will *not* have such impact can be offered. If a school district has a curriculum building project which does not closely involve its regular curriculum people, or a college has a project which is staffed primarily by "outsiders" on "soft money", or if the "regulars" in *any* agency are not closely involved with the day-to-day operations of a project, the possibility of long-term changes is clearly limited.

To find out if your project is headed toward the "short-term oblivion" route, do this little test. Periodically, say one day every month, make a checklist of all the things that happened in the project during the day. Decide which of these activities will continue to occur when the funding runs out. If most of the activities are dependent on outside money, the prediction is pretty clear.

Evaluating Objectives: Two Interpretations

Criterion-referenced tests, mastery learning, performance contracting, learning packages, behavioral objectives— these currently popular terms are all closely related to the notion of *stating objectives specifically*. The evaluator has the role of "evaluating the objectives." People interpret this role in two very different ways, and it is important that the project people and the evaluator are both "singing out of the same hymnbook" regarding the interpretation.

To most evaluators, "evaluation of objectives" means "evaluating to see if the objectives have been attained." This implies measures which are most suited to determining if an objective has been reached by the people toward whom the project was geared. The whole range of techniques, mastery tests, questionnaires, and interviews can be brought to bear on the objective.

However, some people interpret "evaluating the objective" to mean comparing the objective to other possible objectives. That is, they see this as requiring a *value judgment* of the objective, compared to others. For example, take this objective: "The student shall recall the equivalencies between the common metric and English units of time, length, volume and weight." To evaluate the attainment of the objective, one would devise some sort of an achievement test asking the student to recall all, or a random sample of, these equivalencies.

On the other hand, to place a value judgment on the objective would require asking questions like: Why should a student *recall* these equivalencies? Perhaps the student should simply *recognize* them, *use* them in context. Maybe *estimating* lengths, masses, and volumes is the proper manner in which students should "know about" the relation between English and metric units. Even broader, why is this information important at all? Valuable school time will be taken if the student is to reach this objective. Couldn't that time be better spent elsewhere, for example, in reading a newspaper or socializing with his peers?

Evaluating to see if the objective has been attained is clearly the evaluator's job, but placing value judgments usually is not. These judgments should be made by the people who are served by the agency housing the project. If the "housing agency" is a school, then the people of the district should make the value judgments. In cases where the situation at hand *forces* the evaluator to make these value judgments of objectives, I believe the evaluator should clearly delineate the two kinds in the final report. The client has the right to know which evaluations are technical judgments, and which are basically the evaluator's opinion.

Concerning Car Salesmen: What Specifically Did the Project Writers Promise to Do?

"This lil sweet 'arts 'bout the bes' '67 in town. Lo miles . . . bin treat'd like a baby . . . bes' bargain in town at four-fifty."

The car salesman wraps one single statement of fact in the same package with some half-truths, implications, and fast talk. The only fact was the price. If the car was not actually the best '67 in town or if it is not the best bargain in town, you won't really have any recourse later. The only thing he actually said he would *do* or guarantee was to sell the car for "four-fifty."

Proposal writers are often like car salesmen. The evaluator has to separate the actual promises—the things the project will do—from the other implications. Usually, the project staff will only be held accountable for things they specifically promised to do—just as the salesman is only accountable for one fact in the statement above.

Now, most proposal writers are not dishonest. The writer has to build a case to show the background conditions which lead to the need for additional funds. A competent proposal writer makes the best possible case for the proposed project. Sometimes it's hard to separate the "we will *do* this" statements from the "flag and motherhood" parts. Proposal writers are clever at mixing them up. It is possible to cut through the rhetoric, and go right to the heart of a proposal by looking immediately at a few key places. The two primary starting points are the budget and a project timetable, which is usually required in proposals.

Use these two to set down a list of "will do" items. If the budget lists money for a "field coordinator," then you can assume this is a "will do" item and not just rhetoric. The timetable will probably tell you what this field coordinator is supposed to be *doing*. I almost always begin reading a new proposal at the budget section, and then move to the activity section. Where is the project going to spend the money? The answer gives us a good clue as to the real objectives.

The budget has another important use. With it you can develop a hierarchy of objectives, from most to least important. Think back to the hypothetical situation outlined previously where the Assistant Superintendent tried to solve a reading problem. Suppose the budget looked like this:

Administration	$ 20,000
Reading Specialists	$ 84,000
Test Specialist	$ 10,000
In-Service Training	$ 1,000
Secretarial	$ 6,000
Consultants	$ 5,000

Now the evaluator prepares a score sheet:

	GOOD	INDIFFERENT	BAD
Administration	X		
Reading Specialists			X
Test Specialist	X		
In-Service Training	X		
Secretarial	X		
Consultants	X		

That's 5 to 1—a good project. Right? Wrong! That's 84 to 42, or 2 to 1—a bad project. That is, $84,000 was spent on a section which didn't work out, while $42,000 was spent on five sections which rated "good."

To develop the hierarchy of important objectives, try to allocate the money in the budget to the different objectives. Some items, such as staff, travel and materials are easy to allocate, but others may have to be skipped. After allocating as many times as possible, make some sort of chart—a "pieces of pie" circle graph, for example—to get an idea of the relative importance of the objectives. The evaluation efforts should be similarly distributed.

Timetable: If a Child is to be Born in October, Conception Must Occur Somewhat Earlier

A proposal should contain a timetable of events. If it does not, then the evaluator should help the project develop one. If the one already developed is unrealistic, then the evaluator should help get it revised. Look at this objective:

"By June 1, 1975, the project staff will have received the approval of the School Board to pilot test a series of special reading programs in three schools in the district."

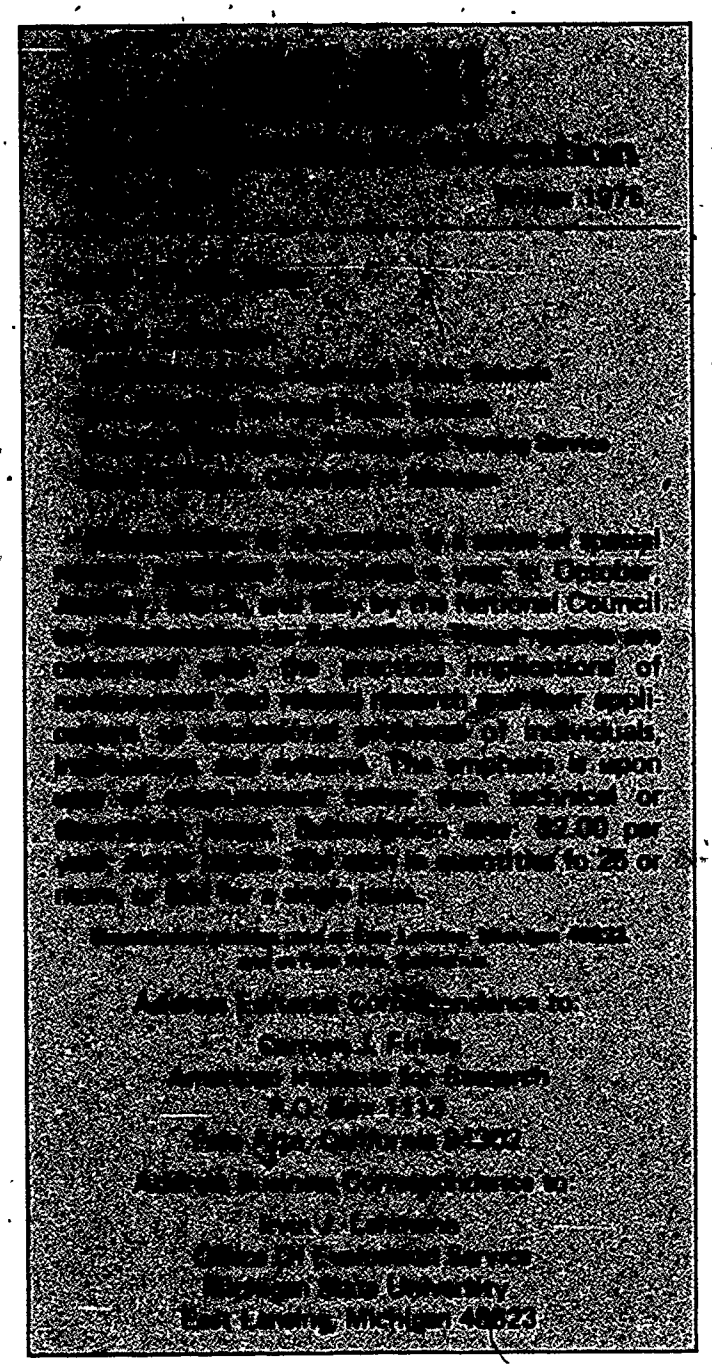
Failure to plan ahead too often leads to frantic and inefficient last minute efforts. What kind of planning goes into the objective above? Start from the completion and work backwards—that usually is the easiest way.

The deadline is June 1st. What is the last School Board meeting before June 1st? May 20th. And how long prior to the meeting must members have material which will be acted upon? One month. (You're back to April 20th.) How long to type and collate the report in the bureaucracy? Two weeks. (Now at April 6th.) How long to plan the program, including obtaining permissions, holding hearings, consulting "learned experts," and things like that? Three months. We are now at the first of the year. Must staff be hired? What other approvals are needed and how long will they take? The press of day-to-day activities frequently can cause a project to avoid long range planning. The evaluator should help keep the project on schedule.

The Project Staff: Concerning Prior Commitments and Real Power

First fable: The Assistant Superintendent wrote the proposal covering the alleged reading and diagnosing problem in his district. The proposal was funded. His Superintendent thought it might be nice for Mr. Assistant Superintendent to direct the project, so the School Board granted him a three-year leave of absence from his regular job and named him project director.

Mr. Former Assistant Superintendent starts work as project director. But one day, a few weeks later, the Superintendent is faced with a problem that he knows the Former Assistant Superintendent used to handle beautifully. So he asks for a small favor, "just this one time." These *favors* probably will continue, and soon the "project director" is working far less than full time on the project.



The moral of the fable: If someone on the project staff was with the same organization prior to appointment to the project, look very carefully. Organizations—school districts, public agencies, universities, etc.—have a tendency to appoint a person from within for a new job without appointing a different person to the old job. Since the old job often goes unfilled, the new project person carries many of the old responsibilities with him. If a new Assistant Superintendent is not appointed to handle the responsibilities of the Former Assistant, you can rest assured he is not committed full time to the project.

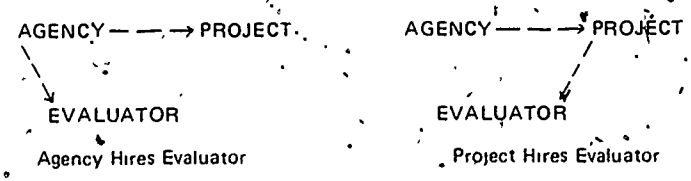
Final Fable: As project director, the Assistant Superintendent recruits two teachers from each building to work with him on the project. The project picks up most of the teachers' salaries. He directs these teachers to begin working on materials for six- and seven-year old children. One week later, he checks again with the teachers and finds two disturbing notes: First, they had not completed nearly as much as he expected; and second, they were also developing materials for pre-schoolers and eight-year olds. In addition, they seemed resistive to his urgings. He found out, after some searching, that while *on paper* the teachers were paid by and responsible to the project, *in practice* anything that happened in a particular building was the responsibility of the building Principal. Even if the teachers had not been holdover teachers from prior years, the lines of authority would have been blurred by the traditions of the district.

The moral of this fable: You can tell who has the "paper power" simply by reading the proposal. Be very sensitive to the other issue, however, of "real" power. Who is always consulted when major decisions are made? Who is able to counter or change directions given by the project personnel?

If the lines of authority are not clear, the efforts of the project personnel can be seriously diminished. If the teachers in the above example do not know which person to respond to (the project director or the building principal), they will probably not do a satisfactory job of either set of directions. Blurred lines of authority lead to all kinds of intrigue and inefficiency. The implication should not be drawn that the project staff is given absolute and unchallenged authority over that which happens under the project's auspices. Obviously, these activities will affect the agency sponsoring the project and the agency needs to have a hand in the decision-making. But the project staff cannot abdicate all decision-making responsiblity in favor of the personnel in the sponsoring agency. A balance must be struck and the project evaluator must find out and make explicit (to all, if possible, but at least to the evaluation team) just what the specific lines of "real" authority are.

And the Evaluator, Too, Must Ask the Question, "Who Am I Working For?"

Sometimes the evaluator is contacted by an agency to evaluate a project which was funded by the agency. The evaluator is working for the agency—helping the agency interact with the project, as in the figure on the left. Sometimes, however, the project makes the contact with the evaluator—perhaps at the urging of the agency. Then the evaluator is working with the project, helping the project interact with the agency, as on the right.



The evaluator must try to answer the question, "Who am I working for?" which usually leads to the question, "To what use are my results going to be

put?" The two questions belong together. So very many times over the past few years I've had a long difficult discussion with project or agency people and finally put the question this way: "Now look, folks, which do you want: A hard-knocking internal evaluation which will tell you what's working—and what's not—or an evaluation which will accentuate the 'positive' in an effort to sell the concept to outsiders? Is my report for *you* or primarily for outside comsumption?"

Usually these people really want both—and that's possible. The evaluator can gather all of the proper information, letting the chips fall where they will, and write two reports. One is for internal consumption and the other for external public relations work. I'm not suggesting that the second report be inaccurate in any way—only that it dwell more on the positive notes, rather than pointing out many flaws in areas which need attention.

Is that unethical?

Clearly it's unethical if the "internal" report contains information about serious problems and the information is ignored. One part of the agreement must be that the project *act on* the suggestions given in the internal report. If this is not done, then the evaluator should make the private report public.

But shouldn't this always be done — share all the results with everyone involved — project funding agency and public? In the best of all worlds the answer is clearly in the affirmative, especially when we see the devastating effects a "cover-up" can have. (However, a "cover-up" is not what is involved here. The true information would still go to the project.) But here is the other side of the coin—and in the real world it's worth considering: A project or an agency can easily find a "house evaluator." A "house evaluator" is someone who will figure out a way to get precisely the results the project wants to see. Now, if ten projects are vying for second year funding and nine of them hire a "house evaluator," while the tenth gets a thorough evaluator, then the only one which will *appear* to have problems is the tenth one. Unfortunately, funding agencies frequently do not look any deeper than the evaluation report, and this tenth project would not be refunded. When this is the situation—and the picture drawn is not a hypothetical one, but very real—then the idea of an internal and external report is more defensible.

For me, the most unhappy situations occur when the project people want *only* an "accentuate the positive" report, but I wasn't insightful enough to see this until all the work was done. A project I worked on had about $20,000 to increase the reading and math performance of 100,000 inner-city children, plus develop more positive affect within a four month span. When the obvious result occurred, I reported it. My role as project evaluator was then terminated. This happened only one other time, and the unhappy part is knowing that you've essentially wasted that time, since the people are not going to implement any of your

REPORTS AVAILABLE

Back issues of *Measurement in Education* are available at 35¢ each in quantities of 25 or more for a single issue.

Vol. 1, No. 1	*Helping Teachers Use Tests* by Robert L. Thorndike
No. 2	*Interpreting Achievement Profiles—Uses and Warnings* by Eric F. Gardner
No. 3	*Mastery Learning and Mastery Testing* by Samuel T. Mayo
No. 4	*On Reporting Test Results to Community Groups* by Alden W. Badal & Edwin P. Larsen
Vol. 2, No. 1	*National Assessment Says* by Frank B. Womer
No. 2	*The PLAN System for Individualizing Education* by John C. Flanagan
No. 3	*Measurement Aspects of Performance Contracting* by Richard E. Schutz
No. 4	*The History of Grading Practices* by Louise Witmer Cureton
Vol. 3, No. 1	*Using Your Achievement Test Score Reports* by Edwin Gary Joselyn & Jack C. Merwin
No. 2	*An Item Analysis Service for Teachers* by Willard G. Warrington
No. 3	*On the Reliability of Ratings of Essay Examinations* by William E. Coffman
No. 4	*Criterion-Referenced Testing in the Classroom* by Peter W. Airasian and George F. Madaus
Vol. 4, No. 1	*Goals and Objectives in Planning and Evaluation: A Second Generation* by Victor W. Doherty and Walter E. Hathaway
No. 2	*Career Maturity* by John O. Crites
No. 3	*Assessing Educational Achievement in the Affective Domain* by Ralph W. Tyler
No. 4	*The National Test-Equating Study in Reading (The Anchor Test Study)* by Richard M. Jaeger
Vol. 5, No. 1	*The Tangled Web* by Fred F. Harcleroad
No. 2	*A Moratorium? What Kind?* by William E. Coffman
No. 3	*Evaluators, Educators, and the Publics. A Detente?* by William A. Mehrens
No. 4	*Shall We Get Rid of Grades?* by Robert L. Ebel

suggestions. To avoid this type of situation, the evaluator should avoid people who are so jealous or insecure of the project that they cannot accept anything but positive evaluations!

Continuous Assessment

One very special domain of evaluation situations requires a special approach. These special situations involve schools. My experience in dealing with evaluation problems brought by school administrators—excluding those associated with projects funded by outside sources—is that they fall into three groups:

(1) Evaluation of small, in-house projects to help decide whether or not they should continue. Small changes in curriculum, instructional delivery system, or administrative organization of teachers and students fall under this heading.

(2) Addressing a difficult question raised by someone or some group to whom the administrator must attend. The question might come from a teacher or teacher group, a parent or parent group, a board member, or a newspaper reporter. Two real situations into which I've been drawn in the past few months are these (cleverly reworded, I hope, to hide the actual cases):

"Are the River Edge School's children falling behind the Creek Bank School's children in basic skills?" and "Has the ungraded classroom solved last year's scapegoating syndrome at West School?"

(3) Techniques for providing a systematic feedback system to the public in a manner which will not be misunderstood. "Misunderstood" frequently means no more than cases where the public applies *absolute* standards where *relative* interpretations are more appropriate. Like "80% mastery! Why, in this world of numbers we need 100%—yes, by God!—100%! When I was a child . . ."

If prior baseline data are not available—and such data almost never are—then these questions are difficult to answer. Under such conditions, the cost of getting an evaluation answer is simply unrealistic in the face of the conceivable benefits which could accrue therefrom. Needed is an inexpensive technique for continuously accumulating baseline data for all of the major constituencies of the district.

Such a program would have the following characteristics:

(a) Baseline data would be gathered in all important areas of the school's program. Included should be data on student achievement, affective measures from current and former students, affective information from faculty, measures of community knowledge, interest, and attitude, demographic information from the supporting area, and patterns of student flow through the various instructional programs of the school.

(b) Continous data collection. "Continuous" obviously does not mean daily, but means that data will be gathered at fixed, pre-set intervals. The interval length should be a function of the measure. For example, demographic data change slowly, and a biannual interval would be satisfactory; whereas student achievement changes more quickly, and quarterly samples would be defensible. The faculty is fairly stable, and yearly samples seem appropriate. The size of the sample interval should be set with the district administrators to reflect their perceptions of the change dynamics of each variable.

(c) Low cost feedback is an understandable format. If the system would require outside funding for continued operation, or if the feedback would require consultation with measurement or computer experts every time the administration seeks to use it, then the system will not have wide-ranging applicability. The system should be initially established with a computer feedback system such that current district personnel can feed the most recent measures into the system to provide updated reports on all of the measures involved. Most districts already set aside some monies for evaluation and research work, as well as having at least one administrative person devoted to spending time on these activities. Once the continuous assessment system is in operation, the district should not have to invest substantially different amounts of time and money as it had invested in the past.

(d) Trend analysis. Educational data are, by and large, ordinal at best. A statistic—a mean or a grade equivalent—is usually more useful in the relative than in the absolute sense. In making long-range decisions, or in deciding when to intervene in an existing program, trends are frequently more understandable than a table of figures. Computer generated graphs, with automatic statistical tests of significant changes in the trends are an appropriate feedback system for each measure. Where a trend has changed statistically, tabular data going back as far as possible would be provided, along with a description of the meaning of the observed change. The system should also note the interrelatedness of the measures and the trends to show the administration the points where events tend to happen together.

(e) Sampling is a key word for data collection. The primary data targets involve *groups*. Individual measures are not required from each person in the various groups. Taking the two key questions together—"How accurate must the results be?" and "How much time and money are available?"—the proper sample size can be determined. A very important phase of the initial set-up of the project would be the generation of specific random sampling techniques from the different populations. These directions would then be used by the district in its systematic data collection efforts.

Although I have sought a hospitable school district for the establishment of such a system for more than five years now, it is only recently that a tentative agreement to begin operations in a district has been reached. Aside from being a commentary on my

ineptness as a salesman, this difficulty offers a point on the "crisis orientation" of school administrators which is worth noting by evaluators. This week's crisis has never happened before and may never happen again. Next week's or next month's crisis isn't known yet. How can one be sure the baseline data will help resolve these currently unforseen events? Additionally, a system based on trends cannot really begin functioning properly until the measurements have gone through several cycles. The idea of investing in a system which cannot be expected to pay off for a year or two is not attractive in the face of pressing current problems.

Some Other Evaluator Roles

Regardless of the size, funding, or location of the project, the evaluator has a unique opportunity to fulfill certain other less-well-known roles. Three of these are keeping communication lines open, translating numerical results into understandable terms for the project staff, and creating an "action now" philosophy about evaluation results. Let me expand briefly on each of these.

Keeping the communication lines open. Every project reaches a variety of groups. The groups usually have their own interests and have probably learned to communicate with each other in well-established ways. For example, the program with the Assistant Superintendent described earlier will involve students, teachers, administrators, some reading specialists, probably a few university professors, and maybe even some graduate students. Teachers and administrators usually have an employee-employer sort of relationship. Two-way communication, even though badly needed by the project, will be difficult to establish where prior tradition is strong. Building administrators have established communication channels with central office administrators; the manners in which reading specialists communicate with university professors may be fixed, and so forth. To be successful, the project may require a level of communication among groups which is not likely to occur without some sort of externally imposed greasing of the communication skids.

The evaluator is in a good position to be a communication facilitator. This is especially true if the evaluator is external—that is, not a regular or previous member of the agency housing the project. If the evaluator is external to the agency funding the project, he will be relatively free from intimidation or coercion from any group on the project or in the agency. From this perspective, it will be possible for him to establish communication between different groups by insuring anonymity to all who respond to his questions. Without anonymity, most people hesitate to make comments about persons higher up in the organization, especially if the comments might be construed as being critical. If the evaluator can insure anonymity and establish his own credibility, this hesitancy will evaporate and two way communication will be possible.

Translating numerical results into practical meaning. The very last part of the sentence above needs a few additional comments. The evaluation results aren't of any value at all if the people who make the decisions in the project cannot interpret them. Many project administrators are quite uncomfortable with numerical results of any kind. The evaluator must provide the project staff with more than just the results. Also provided must be information on probable implications, points of inaccuracy, information which may be unreliable or biased, and suggestions for further data gathering. Evaluation reports which are designed to help the project decision-makers should not be written as scholarly articles for peers in the evaluation business. They must be understandable by those who need the results. The test of the evaluation is not the technical beauty of the measurement devices. The test is the usefulness of the results.

An action now philosophy. School administrators and project personnel—with notable exceptions, of course—view evaluators as not being within the mainstream of the system or project. The evaluator's work is viewed as a necessary evil at best, and as a threat at worst. In cases where the evaluator senses these kinds of attitudes, some public relations efforts are necessary. These efforts should be directed toward convincing the staff that thorough evaluation efforts can enhance the probability that the objectives of the project can be reached. The evaluator needs to convince the staff that he or she is not hiding behind the curtains, waiting to expose an embarrassing mistake. Evaluation, it should be argued, involves an "action now" orientation. That is, data gathered throughout the project's life and fed to the staff very quickly can be translated into early intervention in cases where things aren't working as well as had been envisioned. The evaluator needs to convincingly demonstrate that evaluation can transcend summative statements and that good formative evaluation can make positive contributions.

A Final Word

This paper clearly does not constitute a "theory of evaluation." The comments are based on lessons learned in working with a diversity of evaluations projects. Funding on these projects ranged from zero to millions of dollars. Not all of the lessons were pleasant experiences. Unfortunately, more bitter lessons are probably ahead.

Is a definitive "theory of evaluation" possible? Certain well-known statements on evaluation have been written and they describe interesting and thought-provoking general evaluation approaches. But I have failed in efforts to operationalize them. The theories are good ways to think about evaluation concepts, but too general to apply to specific projects. The projects are simply too situation specific.

Whether or not you agree with this philosophy, one current need is acutely apparent to me. This is the

need for a clearing house for evaluators to describe unique approaches which have worked in diverse settings. The format should be closer to *Popular Mechanics* than *Psychometrika*. The present measurement and research journals simply don't appear to be appropriate.

When the Chicago Board of Education officials circulate an RFP for an evaluation project, the mailing list is around one hundred names. This list includes only some of those in the Chicago area who look upon themselves as "evaluators" for at least part of their professional time. Across the country, the total list of people who have functioned as evaluators for schools or projects must number in the thousands. There must be some very practical suggestions we can give one another. Somehow the professional measurement organizations should take a more active role in opening communication lines among these many evaluators.

Evaluation is not research. Once the conditions have been established, the researcher does not interact with with the experiment with thoughts toward changing conditions to insure significance. And evaluation is not equivalent to measurement. Measurement applies to the devices used and their validation—the tests, interviews, performance tasks, or questionnaires used. Evaluation is very much an interactive process where the technical expertise must interact with the political realities of the system, as well as the idiosyncratic personalities involved. From my perspective, the presence of an external evaluator has a very positive impact on both project and funding agency. In such a setting, the evaluator role is interesting, challenging, and worthwhile.

NCME NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION

Second class postage paid at Palo Alto, California

Office of Evaluation Services
Michigan State University
East Lansing, Michigan 48823

A series of special reports of the National Council on Measurement in Education