ED 311 047   SP 031 541

AUTHOR   Furtwengler, Carol B.
TITLE   Evaluating Career Ladder/Incentive Programs. Sumary
              of a Report.
INSTITUTION   Southern Regional Education Board, Atlanta, Ga.
PUB DATE   89
NOTE   17p.; Complete report is available from SREB Career
              Ladder Clearinghouse, 592 Tenth Street, N.W.,
              Atlanta, GA 30381-5790 ($4.00).
PUB TYPE   Reports - Descriptive (141)

EDRS PRICE   MF01/PC01 Plus Postage.
DESCRIPTORS   *Career Ladders; Elementary Secondary Education;
              Evaluation Criteria; *Evaluation Methods;
              *Incentives; Merit Pay; *Motivation Techniques;
              *Professional Development; Program Evaluation
IDENTIFIERS   Arizona; North Carolina; South Carolina; Utah

ABSTRACT
        This report summary focuses on third-party
evaluations of career ladder/incentive pay programs in four states:
North Carolina, South Carolina, Arizona, and Utah. The summary
addresses programs that contain the concepts included in
performance-based ladders, job-enlargement ladders, and
professional-development ladders. The first section presents
highlights from four major third-party evaluations of career
ladder/incentive programs that have been completed or are underway at
this time. The second section offers considerations that should
precede a third-party evaluation and suggestions for the major steps
included in the process. (JD)

EVALUATING CAREER LADDER/INCENTIVE PROGRAMS

Carol B. Furtwengler

Have the career ladder and other incentive programs for teachers made a difference in public education? Many states, legislatures, and educators want to know if public dollars are being well spent on these reform efforts. This report focuses on third-party evaluations of career ladder/incentive pay programs in four states and provides guidelines for evaluating incentive programs. North Carolina and South Carolina are both member states of the Southern Regional Education Board; Arizona and Utah are also included.

Many answers to questions about whether a program is working can be obtained by conducting a third-party evaluation. A comprehensive program evaluation uses information from many sources and combines the information to draw conclusions on the major questions that need to be answered about the program's implementation, operation, and effects and to make recommendations about the program's continuation and/or expansion.

Third-party evaluation involves persons from outside the system who can objectively assess the program through unbiased and appropriate research techniques to determine its effectiveness.

At least three general types of career ladder/incentive programs can be identified, although their characteristics are not mutually exclusive.

o   <u>Performance-based ladders</u> are those where progress up the rungs is based upon evidence of increased competence at progressively more difficult and/or complex levels of professional performance.

o   <u>Job-enlargement ladders</u> are based on differentiated job roles and responsibilities that serve the needs of students and the school beyond the teacher's own classroom. These job-enlargement activities include such duties as supervising beginning teachers, serving as a teacher representative to the administrative staff, and/or developing new or updated curriculum.

o   <u>Professional-development ladders</u> determine advancement based upon the completion of qualifying staff development activities, coursework, and/or advanced degrees, similar to ways that teachers are now paid.

It is not uncommon to find a career ladder/incentive program based upon a combination of two, or even all three, of these concepts.

Carol Furtwengler, former Assistant Commissioner of Education in Tennessee during development of the Tennessee Career Ladder Program, and now a private consultant, led a team of researchers who conducted a recent evaluation of North Carolina's Career Development Program.

This report addresses programs that contain the concepts included in performance-based ladders, job-enlargement ladders, and professional-development ladders. The first section presents highlights from four major third-party evaluations of career ladder/incentive programs that have been completed or are underway at this time. The second section offers considerations that should precede a third-party evaluation and suggestions for the major steps included in the process.

Evaluations of Four Career Ladder/Incentive Programs

Third-party evaluations of incentive programs in the states of North Carolina, South Carolina, Arizona, and Utah have been completed recently. These studies have been selected because they were not limited to opinion survey research--that is, perceptions of what those involved believe is happening. Perceptions do not present a complete picture, although they are an important part of an evaluation process. A comprehensive program evaluation examines "output" measures, such as improvement in teachers' performances, improvement in students' achievement, statistical information, and program records as well as information received from opinion surveys and personal interviews.

Some states have preferred to have their incentive programs reviewed by professionals from outside the system. An outside program evaluation provides more objectivity than relying on program evaluations done by persons who have been involved with, or have already formed opinions about, program development and implementation. Third-party evaluations, thus, are playing an important role in assessing the general effectiveness of these programs.

North Carolina's Pilot Career Development Program

North Carolina is in the final year of its pilot Career Development Program, which began in 1985. The program was designed to attract and retain the best personnel in teaching and school administration and "to improve the quality of classroom instruction, to increase the attractiveness of teaching, and to encourage the recognition and retention of high-quality teachers." The program is a performance-based ladder with job-enlargement opportunities. A nine-month, third-party evaluation of the 15 pilot sites was completed recently. Legislation that outlined the purposes of the evaluation stated that it ". . .shall be designed to study the impact of the School Career Development Program on: improved teacher performance, employee retention and recruitment, employee satisfaction, overall school improvement, enhanced learning environment, students' attitude towards school, and community support of the programs."

This third-party evaluation is of interest because of its multiple data-source approach to collecting information. Six separate evaluation activities were carried out and the

information was combined within major areas to draw conclusions about the effects of the program and to make recommendations for program improvement and expansion. The major areas included: teacher evaluation, administrator evaluation, the appeals process, the staff development programs, extra-duty/extra-pay activities, state agency's role, and future directions for the program.

The comparison of program implementation and operation at the 15 pilot sites revealed that many similarities exist, but differences also were apparent. In many cases, the differences could be attributed to the size of the school system. For example, the evaluation revealed that a large number of the appeals occurring were in larger systems, a natural outcome of system size. Therefore, initial concerns about the number of appeals were not necessarily warranted.

The second activity was a comprehensive opinion survey of all personnel eligible to participate in the Career Development Program at each pilot site. Local school board members also were surveyed. The results of the opinion survey (89 percent response rate) revealed that overall mean scores about the program ranged from slightly positive to positive. Questions focused on four general areas: the program and the evaluation systems, the observer/evaluators who carry out the evaluations, the staff development programs, and the attainment of program goals. In general, teachers were slightly positive while other groups of school personnel and school board members were much stronger in their support of the program. On a five-point scale (1 - strongly disagree and 5 - strongly agree), the mean scores responses on selected questions from the various respondent groups were:

I agree with the results of my year-end evaluation.
Teachers 3.7; Principals 3.7; Central Office/Others 4.0

My system has fair and reasonable procedures for the Career Development Program.
Teachers 3.6; Principals 4.3; Central Office/Others 3.9; School Board Members 4.4

The Career Development Program should be expanded to other North Carolina school systems.
Teachers 3.3; Principals 4.1; Central Office/Others 3.7; School Board Members 4.0

The Career Development Program has encouraged teachers to improve their teaching techniques.
Teachers 3.4; Principals 4.2; Central Office/Others 3.8; School Board Members 3.9

The third evaluation activity included a review of more than 300 randomly selected evaluation records for teachers, principals, and assistant principals over a three-year period. Teachers with experience as observer/evaluators were trained to read and analyze evaluation records using the North Carolina Teacher Performance Appraisal Instrument. Results

of the activity revealed that teacher performance was improving over time, as shown by the ratings of the independent readers. When the principals' summative ratings of teachers given each year were analyzed over the three years, improvement in teacher performance was also noted. Information on Professional Development Plans, a part of the evaluation system, revealed that 44 percent of the teachers showed improvement in their identified areas. While the study reports that teacher performance is improving, a comparison of the readers' ratings and the principals' ratings revealed that significant differences existed in the ratings assigned to teachers. The principals' ratings were consistently higher than the independent readers' ratings.

The fourth evaluation activity was a district progress assessment to determine whether any of the broader goals identified for the program evaluation (improved learning environment, students' attitudes toward school, and overall school improvement) were being met. Information was collected on seven variables from the 15 pilot sites and 15 "matched" sites-- that is, selected systems that were not in the Career Development Program. Variables included items such as student attendance, estimated student dropout rate, and teacher applicants. No significant differences between pilot sites and matched sites were revealed for any of the variables examined. Evidence from the other activities indicated that teaching and learning were improving. This was confirmed by responses to the opinion surveys and site interview questions and was also evidenced in the performance of teachers as shown in the evaluation records. A study done by the State Department of Public Instruction also indicated that student achievement appeared to be greater in the pilot sites than other matched sites. The evidence was inconclusive, however, on whether the program was affecting student dropout rates, student attendance, and other areas related to overall school improvement.

The fifth activity involved conducting interviews at each of the 15 pilot sites. Over 300 individuals, representing various school personnel groups, were interviewed. Strong support in favor of the program was found among most of the interviewees.

The sixth evaluation activity was designed to obtain information from interviews with deans of colleges of education to determine their knowledge and attitudes toward the Career Development Program and their perceptions of how the Career Development Program has affected teacher training programs.

When the information from the six evaluation activities was analyzed and compared, it was apparent that teaching and learning were improving in the pilot sites in North Carolina. No evidence was found, however, during the study between the pilot sites and non-pilot sites, when other variables related to school effectiveness, such as student attendance or estimated student drop-out rates, were analyzed.

6

The evaluation report offered recommendations for program improvement and expansion. Among them were: 1) the early admittance of out-of-state teachers to the program when they enter North Carolina school systems, 2) principals and peer observer/evaluators meeting together to determine teachers' summative ratings at the end of the year (consensus decision-making), 3) an increased emphasis on Professional Development Plans, and 4) the statewide expansion of the program in a well-planned manner.

South Carolina's Teacher Incentive Program (TIP)

The South Carolina Improvement Act of 1984 mandated the development of a Teacher Incentive Program (TIP) to reward teachers for superior performance and productivity. Two other incentives programs--a Principal Incentive Program and a School Incentive Program--were also instituted. These programs also had third-party evaluations, but this discussion is limited to teacher incentive programs.

The Teacher Incentive Program originally pilot-tested three models--the Bonus Model, the Career Ladder Model, and the Campus/Individual Model. The Career Ladder Model was eliminated from the approved models at the conclusion of the pilot phase because of its emphasis on extra work (job-enlargement) rather than on performance.

The Bonus Model and the Campus/Individual Model emphasize assessment of classroom performance and student achievement, rather than more pay for more work, and are, therefore, performance-based ladders. The self-improvement activities (or professional development components of the ladder) of these models have been modified in ways that reduce their original emphasis. The incentive program now focuses predominately on teacher performance and student outcome measures.

Consultants have conducted three third-party evaluations of South Carolina's Teacher Incentive Program. The 1986-87 study addressed four major areas: program implementation, understanding and acceptance of the program, program resource allocation, and guidelines and procedures for model implementation. The study included interviews at all sites, a randomly administered teacher questionnaire, a principal questionnaire, and a review and analysis of statistical information received from the State Department of Education. While a majority of teachers in all models reported that they had a clear understanding of the program and that the instrument used for classroom observation (evaluation) was appropriate, a majority of the teachers responding to the questionnaire did not feel that the TIP had encouraged teachers to improve their teaching techniques or that their concept of a "superior" teacher was consistent with the award requirements of their district's TIP.

Evaluations of TIP were also conducted in 1987-88 and 1988-89. The third and most recent findings report: 1) 24 percent of initial participants withdrew from the program due to the time and additional work required; 2) 78 percent of the remaining participants received

an incentive award; 3) a slight relationship was found between student achievement gains and teachers who received awards in the Campus Model Programs; and 4) variance was found in the survey responses of superintendents, principals, teachers, and board members, with those receiving awards being most positive and those who were disqualified being the least positive. Four major recommendations of the most recent evaluation study include:
1) continuing statewide implementation of the Campus Model and Bonus Model programs;
2) expanding future program evaluations, both in scope and financial resources; 3) close monitoring by the State Department of Education of the 50 school systems that are implementing TIP for the first time during 1988-89, and 4) collecting, monitoring, and reporting annual TIP participation and award statistics by the Department of Education .

Arizona Career Ladder Pilot Test Program

Arizona's program is designed as a teacher incentive program to improve the profession and to increase student achievement. Local pilot systems were given complete control over the program design, implementation, and measurement of program outcomes. The 15 pilot programs have performance-based ladders with provisions for job-enlargement activities.

When the Arizona legislature passed a bill in 1985 for establishing a career ladder program, it also established a Center for Excellence in Education at Northern Arizona University to carry out evaluation research related to the pilot programs for a five-year period. According to Packard and Dereshiwsky, the project researchers, the purposes of the evaluation are to determine if:

1. Education is positively influenced when teachers are paid for performance, rather than on years of experience;

2. The program improves recruitment, retention, and motivation of high-quality teachers;

3. The program will develop and improve teacher performance in the classroom;

4. The program will, in fact, improve student achievement.

Twelve research variables are being studied at the Center for Excellence in Education to determine the career ladder program's effectiveness. Each of these factors--for example, legislative guidelines, organizational climate, readiness of districts, change factors, how teachers are evaluated--will be described, researched, and evaluated as part of the project. The main focus, however, is on whether the career ladder program is affecting improved teacher and student performance. Information is collected from a network committee comprised of members from the 15 pilot sites and from information received at each of the sites through interviews and other data-gathering techniques.

8

Various approaches are being taken in studying the 12 research variables, such as case study, opinion survey, and measures of student achievement gains. Reports of teachers who are assuming leadership roles are described in a case study format. *A Perception Assessment Scale* was developed in 1986 to survey persons most directly affected by the career ladder program. Data are being compiled showing the information from various sources for career ladder and non-career ladder respondents. Student achievement data from the pilot sites are being studied. Each of the 15 pilot sites selects the type of analyses it wants to use to determine student achievement gains and to determine teacher performance.

Although findings from this research study will not be published until October 1989, the researchers state that a tremendous diversity exists among the 15 pilot sites and their readiness to implement the reform program. The report of findings will include recommendations for continuation of the program.

Utah's Career Ladder Program

In 1984, Utah implemented a career ladder program with five parts: extended contract year, job enlargement, performance bonus, career-ladder levels, and incentive funding for teacher shortage areas.

In their request for a study of the program, the State Department of Education stated that the goal of the project was to gather data from various groups of people involved with the program to determine:

1) The effectiveness of the various career ladder program components;

2) The overall impact of the program on Utah's public education system;

3) The impact on teachers, principals, and district administrators who are working in the system;

4) The impact on the instructional process as it pertains to students; and

5) The impact on current and future reform/restructuring of Utah's public education system.

The Far West Laboratory conducted the study and reports the use of five data-collection activities for the evaluation of the Utah Career Ladder Program. These activities included:

1) A content analysis of district Career Ladder Plans from the 1985-86 through the 1987-88 school years;

2) A telephone survey of the state's 40 superintendents and school board presidents;

3) A mail survey of all principals in the state (68 percent response rate) and a random sample of 1,500 teachers (63 percent response rate);

4) A fiscal analysis of teacher salary distribution in 10 districts employing two-thirds of the state's teachers; and

5) Case studies of the Career Ladder's implementation in 12 districts representing different sizes, geographical locations, and variations in funding patterns.

The study reported that the greatest effect of Utah's Career Ladder Program was the focus of teachers and principals on more frequent and effective evaluations of teachers. Broad support was found for the program, and there was an expansion of teacher professionalism through the extended contract year and the job-enlargement parts. The extended day program allows teachers to plan, develop curriculum, and improve their skills, while the job-enlargement opportunities allow teachers to become mentors, grade level chairpersons, or curriculum specialists. The program also provides a powerful mechanism for school improvement activities by creating a more positive climate for learning and for carrying out school and district curriculum planning and management activities. Variation was found, however, in the program's implementation in the school districts.

Policy recommendations included the need for a long-range commitment to the program and the need for continued technical assistance from the Utah State Department of Education. The evaluators also recommended that the job-enlargement component be a short-term, temporary feature of the program to permit school districts to be flexible and innovative and to identify differentiated roles and responsibilities on the Career Ladder. Another major recommendation was to keep the performance bonus relatively small and make it a symbol of good teaching that could be earned by more than just a minority of teachers. The researchers felt that rewarding only a few teachers fosters competition and resentment and creates dissension rather than encouraging good teaching.

Summary of Four Approaches to Evaluation

The third-party evaluations just described approach the task in similar, yet different, ways. In defining the purposes of the evaluation studies, Arizona and North Carolina wanted to determine if teachers were being recruited, retained, and rewarded by the incentive programs. South Carolina's study requested more information about program implementation and the fiscal impact of the program. All studies addressed the improvement of teacher performance, and South Carolina, Arizona, and Utah addressed the programs' effect on student achievement gains. (The North Carolina Department of Public Instruction had already completed a study related to student achievement in their pilot districts.) Although some purposes of the studies are not precisely stated in the objectives, each was concerned about the perceptions of program participants and about the effects of the programs on overall school improvement.

10

The ways in which the evaluations are carried out also reveal similarities and differences. Arizona has established an ongoing Research Center to collect information on the pilot programs as implementation proceeds. South Carolina has commissioned studies each year rather than waiting until the completion of its full pilot period, and recommendations have altered the South Carolina Teacher Incentive Program during the course of its pilot implementation. North Carolina's evaluation was a detailed study of actual pilot program operation and results conducted near the end of the four-year pilot period. All pilot sites and participants were included in the study. This information allowed for the formulation of recommendations for improvements and for program expansion. Utah's evaluation was a broad-based policy implementation study. Because Utah's program was implemented statewide, several of the activities focused on randomly selected respondents or school districts. Emphasis was given to the role of the Career Ladder Program--whether the program had created effective school practices. The fiscal impact on teachers' salaries was also examined. Each study used several sources of information to draw conclusions and make recommendations.

Essential Elements to Consider for a Third-Party Evaluation

Four evaluative studies of career ladder/incentive pay programs provide guidance on the essential elements involved in preparing for a third-party evaluation. As state agencies reflect on their reform programs and begin to assess their impact, it is important that they be cognizant of the benefits of a third-party program evaluation. This section provides direction for initial considerations for a third-party evaluation and the major steps involved in the evaluation process.

Initial Considerations for Evaluations

Recognize change. In assessing reform programs, it is necessary to understand the change process and the types of stress and anxiety created. It takes several years, and adjustments by all parties involved, before the final outcomes are known. One should not expect 100 percent acceptance of an educational reform program during its early years. Does a program have to be acceptable to a certain percentage of those involved to be successful, or is it more important to examine whether the program is producing the desired effect and meeting the goals set for it? After a reasonable period of time, one can expect to find out whether any substantive changes in the areas targeted for improvement are occurring from the effects of the reform program and to weigh this against the "stress" or acceptance level of involved personnel to decide on program continuation.

Determine the goals, objectives, and outcomes. Most career ladder/incentive programs have associated goals that are prescribed, or being sought, by the initial legislation. These may include attracting, retaining, and rewarding of teachers; improving teacher performance; providing staff development programs; or improving student learning. The study must be defined in terms of what information is needed to determine if goals are reached and to make decisions about the program's future.

Do you want to know--

If the initial goals and objectives are being met?

How the evaluation process is working?

About particular aspects of the program, such as staff development, extra-duties, the appeals process?

How the program is affecting the individual teacher, principal, student, school system, or state?

Decide on the use of ongoing or summative evaluation. There are two major approaches to evaluation. Formative evaluation is to gather information during the developmental period and to make needed changes for improvement before a program is fully operational. Summative evaluation makes an overall judgment about a program's effectiveness once it has been implemented. Often, formative evaluation allows changes to be made that improve the program's chances of ultimate success during implementation. On the other hand, changes made during the implementation process often can hinder making decisions about the program's effectiveness because of the constant change, which makes it difficult to determine cause and effect relationships.

Arizona and South Carolina used formative evaluation procedures to assess their programs. The evaluation process is ongoing in Arizona and annual in South Carolina. Arizona's study gathers program information each year, and the 15 pilot sites use the research project to change and adapt their processes during the implementation period. The evaluation has summative features because recommerdations have not been done annually, but will be reported after several years of gathering data. South Carolina commissions annual evaluations and uses the information to make program changes during the implementation process.

Utah and North Carolina have evaluated their programs in a summative manner--waiting until the projects are nearing completion and then evaluating their strengths and areas for improvement. This type of evaluation is more final in nature, but it also allows for feedback on the success the program had in meeting its goals. Thus, it includes formative features as well.

Each new program should be studied by examining its goals and objectives, reviewing the current status of implementation, and identifying the information needs of the policymakers. Both formative and summative evaluation have desirable features and *are not mutually exclusive.* The decision on formative and summative evaluation should be based on several factors. First, a decision must be made about what information is needed and then the available program evaluation funds must be used wisely. Second, what types of changes are expected to occur within a year's time? If an evaluation is done during the first or second year of implementation of a pilot program, information describing what is occurring can be gathered and, perhaps, some suggestions or recommendations for improvement can be made. However, if the intent is to have a program implemented on a pilot basis and then see if it is accomplishing its goals, summative approaches may be a better way of approaching the task. This way, changes do not occur during implementation, and the effects of the program can be studied over several years.

Design the evaluation study. There are at least two approaches for the design
of third-party evalua ions. The first is to formulate a request for proposal (RFP),
which defines in advance the goals to be addressed, the specific audiences to
participate in the study, the methods for conducting the evaluation process, the
time frames, and the products desired (outcomes). Third-party potential contractors
then respond to the RFP.

The second approach is to meet in advance with potential contractors, outline the
general needs of the evaluation study, and let the contractors propose their
different approaches for conducting the study. Often this "pre-proposal" evaluation
plan development involves funding the contractors for their time and effort. If
this method is selected, it is important to check the legal requirements for the
state bid process and make sure potential contractors have equal opportunity to
qualify for the developmental work.

When writing RFPs, it is possible to overlook important audiences to be involved in
the evaluation process or to overlook possible methods for conducting the study.
For example, in the 1986-87 evaluation, South Carolina surveyed only teachers. The
contractor's report recommended that other educators be included in future
evaluations. If it is decided in advance to develop an RFP for the third-party
evaluation, include a flexible specification that allows the bidders to propose
other innovative methods for carrying out the research design.

Determine the budget. The quality and comprehensiveness of any third-party
program evaluation is determined, in part, by the amount of funding that is
available to complete the study. A third-party evaluation is only as good as its
design, and the design is related to the amount of money that is budgeted to
complete the study.

Once the goals, objectives, and outcomes of the study have been defined, then they
must be considered within the context of available funds. Certainly, being
financially prudent is of importance. When millions of dollars are being spent on
career ladder/incentive programs, however, spending money for a quality evaluation
is fiscally sound.

Factors for Third-Party Evaluations

Selecting the third-party evaluator. The selection of a third-party evaluator
should be based upon several criteria, including: 1) the knowledge and skills
represented by the evaluation team; 2) the presentation of a proposal with clear
goals and objectives, which outlines the methods, activities, needed resources, time
frames, outcomes, and budget requirements; and 3) samples of previous work and/or
recommendations from previous contract administrators for whom the evaluator has
worked.

Setting clear administrative procedures. Anyone considering commissioning a
third-party evaluation would be wise to have written administrative procedures
available for the evaluator as part of the contract. Who is the contract
administrator and, if that person is not available, who can be contacted? Who is
actually *in charge*? What procedures exist for correspondence with local school

systems? With the State Department of Education? What are the procedures for review of instruments and data collection devices? What information currently exists and how can access to it be made? Who is in charge of the state data base and what procedures must be followed to obtain information from it? What information is confidential in your state? What type of open-records law exists? Must specific formats be followed for submitting interim reports? Is any special format required for the final report? It is essential that clear administrative procedures be established before the evaluation process begins.

Time should be spent with the contract administrator and third-party evaluator reviewing the required resources needed to carry out the evaluation. Questions should be clarified, such as:

Who will be contacted?

What instruments and procedures will be used?

What instruments and correspondence need to be previewed before their use in the field?

Evaluation is often a threatening process to those undergoing it. Removing stumbling blocks and smoothing the way in advance makes the process more comfortable for everyone involved.

Gaining access to accurate and reliable information. One of the biggest problem areas for a third-party evaluator is attempting to collect accurate and reliable information. Good evaluators will base their research design on the goals and objectives set for the program. However, the goals often are broad and difficult to measure. In other cases, the information to measure goal attainment is not available. If programs are designed to attract people to the profession, does *anyone, anywhere,* know how many new out-of-state applicants have applied to teach in these new programs? Does anyone know if an increase has occurred in teacher applicants at the local school level? And, if so, can these figures be examined to determine if there are significant changes in the numbers?

The same types of questions are raised to determine if the program is retaining teachers. Data may exist on how many teachers have left positions, but the reasons they have left are not known. Did they take maternity leave? Move to another locale where they are still teaching? Or, did they *leave* the profession?

Accurate and reliable data to measure overall school improvement and effectiveness are difficult to obtain in most states. Statewide data, or even school system data, usually are not available to document the number of student tardinesses or the number of discipline referrals in schools. State data reflecting student suspension and expulsion are often erratic. For example, if one wants to know if fewer students are leaving or dropping out of school, you will probably learn it is almost impossible to determine. Student achievement statistics for comparison purposes are usually limited to statewide testing programs. Although it is possible to analyze these data sets, they normally represent a narrow band of curriculum areas.

Evaluation results must be based on accurate and reliable data. Many states, however, institute program goals and objectives and do not concurrently institute data collection (or data base systems) that can be used to determine if the goals and objectives are being reached. A third-party evaluator can only report "what

is." If relevant data are not available, analysis is not possible; if the available data are questionable, the results are not worth the effort. When program goals are set, be sure that procedures for generating and collecting data are put in place to track the achievement of the goals.

Using multiple measures. The third-party evaluations described earlier were selected because they used several sources of information to measure attainment of program goals and objectives and other effects of program implementation. A major weakness in several studies of reform programs is that conclusions are drawn only on information that pertains to perceptions of program effectiveness. Opinions may reflect reaction to one part of the program rather than the total program or to an individual's lack of success in the program. Consideration must be given to the opinions of all audiences involved in a reform program, but other _ata sources that can reveal program effects in a more definitive manner also must be considered.

Perception information from surveys and interviews is useful, but also include "output" information, such as student achievement data, improvement of teaching as reflected by actual evaluation records, documentation of changes that are occurring in the deployment and use of personnel and in the school organization, and the staff development opportunities.

Carrying out the evaluation activities. Collecting accurate and reliable data is essential to the evaluation process. The selection and design of instruments must be related to the goals and objectives being evaluated. In setting the time frames for the evaluative study, provide an ample period for the development and pilot-testing of data collection instruments. Rushing these preparatory steps can result in faulty instruments and can mean a final report lacking cohesiveness. All instruments and procedures should be technically sound and defensible in terms of providing valid and reliable information. Specific instructions, consistently applied, should be utilized in all data collection activities. Pilot tests, sampling techniques, and directions for administration of all data-collection instruments should be well-documented and explained as part of the final evaluation report. In most instances, anonymity of the respondents must be guaranteed in order to obtain fair and impartial opinions.

Analyzing and interpreting the information. If data collection techniques are well-planned and carried out, the analysis and interpretation of the data are easier tasks. All data should be reviewed for accuracy and for the correct application of statistical or qualitative analyses. Information drawn from different sources should be compared to determine its validity and reliability in drawing conclusions. For instance, if opinion surveys are administered, differences found among respondent groups should be compared to findings from other data sources--such as personal interviews, student achievement results, and results of teachers' evaluations. Rationales should be provided for every conclusion or recommendation and should reflect information gleaned from several data sources. The use of only one source of information may provide support for conclusions or recommendations. However, the more "output" data available, or consistency found among several sources of information, the sounder the conclusions.

Preparing the final report. The contract for the study should provide specifications for the final report. Acceptable formats include: 1) prominent background sections that describe the program under evaluation, 2) the specific methods that were used, 3) the reason for the selection of these methods, 4) application and results obtained from the methods. Within the report, major issues should ' e outlined and rationales should be provided for the conclusions and recommendations pertaining to each area.

The final report is the documentation of a credible evaluation process. Therefore, it should be written to be understood by the various audiences it addresses, but it should also contain the technical information necessary to understand the procedures and research techniques employed.

Disseminating results. Once an evaluation report is delivered to those who commissioned it, the report should be made available to the various audiences interested in its findings and recommendations. Several issues are important to the dissemination of the third-party evaluation report. The first is that it should be available in several formats for various audiences. The second issue is that it should be released in a timely manner. Decision makers and personnel involved with the future direction of the program need time to assimilate the recommendations and to consider the feasibility of accepting changes or making the adaptations needed for program improvement, expansion, or elimination.

The third issue is the most important and involves not only the actual dissemination of the report, but how the results of the report are used. *The critical part of an evaluation study is the follow-through that occurs after the examination of a program's strengths and areas for improvement have been reported.* The purpose of an evaluation study is to use the information to make decisions to either eliminate the program, refine the program, and/or expand the program.

Conclusion

Interest in career ladder/incentive programs has increased greatly during the 1980s. Because these programs have instituted major changes in the way teachers are rewarded for doing good or additional work, the programs have been surrounded by controversy and uncertainty as to whether they were accomplishing what they originally intended to accomplish. A good practice for determining the effectiveness of career/ladder incentive programs is by a comprehensive third-party evaluation.

Each of the evaluations discussed in this report had unique elements, but in all cases, multiple methods were used to obtain objective information. Perhaps the most critical factor is that the evaluations gathered opinion and perceptual data, and also combined these results with other more specific information, such as outcomes of teacher evaluations, gains in student achievement, and reviews and analyses of technical and statistical information related to program implementation and operation.

While results from Arizona's evaluation process are not yet available, the other three studies have found positive outcomes. Utah's study reports that the program "...is powerfully and positively changing both the teaching profession and the ways schools are

16

organized to teach students." North Carolina's study found that teaching and learning are improving in the career ladder pilot sites. All of the reports have provided recommendations for continuing and expanding the program. The studies have also suggested that good practices in certain sites should be expanded statewide and recommendations have been made for improving program operation.

Third-party evaluation requires deliberate planning by states and other policymaking personnel to determine the framework and desired outcomes of the study. Only through competent and vigorous evaluation, including the use of valid and reliable evaluation information from several sources, can the effects of these important reform programs be determined.

Selected References

Amlser, M. et. al. (1988) An Evaluation of the Utah Career Ladder System. San Francisco: Far West Laboratory for Educational Research and Development.

Cornett, L. (1988, December). Is "Paying for Performance" Changing Schools? The SREB Career Ladder Clearinghouse Report 1988. Atlanta, Georgia: Southern Regional Education Board.

Furtwengler, C. et. al. (1988). Evaluation of North Carolina's School Career Development Pilot Program. Brentwood, Tennessee: Research and Service Institute, Inc.

Furtwengler, C. and French, R. (1989 in press). Teacher Incentive Programs: Policy, Procedures, Practices and Performance Assessment. Washington, DC: Office of Educational Research and Improvement.

MGT Consultants. (1989, March). An Evaluation of the South Carolina Teacher Incentive Program 1987-88 Pilot-Test Following the Payment of Incentive Awards and an Initial Assessment of Statewide Implementation of the Teacher Incentive Program in 1988-89.

Packard, R. and Dereshiwsky, M. (1988, November 29). Research and Evaluation: A focused Design to Improve Teacher Development and Student Achievement for Enhanced School Effectiveness. Document presented to the Joint Legislative Committee on Career Ladders, Arizona State Capitol.

A copy of the complete report, "Evaluating Career Ladder/Incentive Programs: Accountability for Reform Efforts," by Carol B. Furtwengler, is available from SREB Career Ladder Clearinghouse for $4.00.

17