The Connection Between Personnel Evaluation And School Evaluation

William I. Webster
Dallas Independent School District

The field of educational evaluation has developed and progressed since the early 1800's. Concentration, however, has been primarily upon school and program evaluation. Even the revitalization of the field that took place in the sixties and produced a number of alternative conceptualizations of evaluation (Cronbach, 1963; Scriven, 1967; Stake, 1967; Stufflebeam, 1966) was largely limited to program and school evaluation. Teacher evaluation usually was not considered specifically and, in many cases, was deliberately ignored as too politically volatile to be linked to the still developing field of educational evaluation. While program and school evaluation were becoming more outcomes-based, teacher evaluation was progressing generally in the direction of becoming less quantifiable and more process-oriented (Grote, 1992).

The failure to marry program evaluation and personnel evaluation may have been due in part to the belief that educational programs could be designed that were "teacher-proof". In the late 1960's there were a number of such programs that were designed and marketed. One of the most elaborate of these was Project PLAN that was developed at the American Institutes for Research in the Behavioral Sciences and funded primarily by Westinghouse Learning Corporation (Education, 1970). The philosophy was one of developing and evaluating the program in such a manner as to make it sufficiently "teacher-proof" that a viable teacher evaluation system would be unnecessary. Teachers were merely component parts of the system. Failures were blamed upon the system, not upon individual teacher inadequacies. Other examples of the approach were DITSTAR (Bereiter and Englemann, 1966) and the many and varied computer-assisted instructional systems that were developed during the late 1960's and early 1970's.

The DISD's Evaluation Department, founded in 1969, was built upon the premise that personnel evaluation was not a necessary part of program evaluation. The premise was assumed partly out of political expedience, in that a fledgling department could not be launched in the face of massive assaults by the teacher organizations. Everyone was more comfortable with the concept of evaluating programs and schools, not personnel. The author's oft-quoted paper on the organization and functions of research and evaluation units in large urban school districts made no mention of teacher evaluation as one of the functions of a district evaluation department (Webster, 1975). A survey of the best school district evaluation units in the United States, conducted in 1978 and again in 1983, did not identify a single district evaluation department, of the ninety surveyed, that had any responsibility for personnel evaluation (Webser, 1987).

After twenty-three years of program and school evaluation, much of which has been considered state-of-the-art, the author has come to the conclusion that assumptions were wrong. Teacher-proof programs cannot be designed. Competent teachers can make almost anything work, while incompetent ones can ruin even the most brilliant instructional design. Teacher evaluation systems must be outcomes-based and must be coordinated with ongoing program and school evaluation. Furthermore, evaluations of principals, support personnel, and administrators must be also outcomes-based, in most cases including the same outcome as the teacher system. The paper provides a brief summary of the state-of-the-art in program evaluation, a brief summary of the state-of-the-art in teacher evaluation, a general model for personnel evaluation, and a design for merging the two. The basic premise of the paper is that all educational evaluation, be it program, school, or personnel, rests on the evaluation of student needs.
Program Evaluation

Stufflebeam and Webster (1980) characterized evaluation endeavors into thirteen types of studies used to evaluate education. These approaches range from politically controlled studies, which the authors labeled pseudo-evaluation, to decision-oriented studies to consumer-based studies. Rather than devote a great amount of time and space to elaborating on the various approaches, Table 1, taken from the 1980 article, provides a comprehensive overview of the study types, advance organizers, purpose, source of questions, main questions, typical methods, pioneers, and developers. Evaluation practitioners in the public schools are constantly fighting, if they are serious, to conduct values-oriented or true evaluation, while school administrators are often attempting to promulgate politically controlled or public-relations-inspired studies. The ability of the head of the evaluation unit to force true evaluation studies often determines the prosperity and longevity of that unit, since Boards of Education most often want real evaluation studies, while many superintendents prefer politically controlled or public relation studies.

Having outlined the various approaches to evaluation, the paper will concentrate on the CIPP Model (Stufflebeam, 1969), since that is the model that has been primarily implemented by the evaluation department in Dallas. The section will provide an overview of the CIPP Model for program evaluation. Later in the paper, it will be expanded to include teacher evaluation.

The CIPP Model includes Context, Input, Process, and Product evaluation. Table 2, reproduced from a 1988 paper (Stufflebeam and Webster, 1988), presents an overview of the meaning of context, input, process, and product evaluation. The four types of studies are defined by their objectives, methods, and uses. These concepts are fully explained and illustrated in a book by Stufflebeam and Shinkfield (1985).

Figure 1 provides a schematic depicting ways the school improvement process functions within the parameters of a site-based decision-making model currently being implemented in Dallas. Each school receives an annual needs assessment specifying school levels on important outcome variables. The important outcomes of instruction are determined through District-wide assessments of all of the groups involved in the educational process. School program planning is implemented at the school level. Planning focuses on determining the best method to succeed from current levels of important outcomes to desired levels of those outcomes and culminates in the production of a strategic plan, the School Improvement Plan (SIP).

Specifically, once the needs assessment has identified needs, school staff must prioritize those needs and focus upon reducing the discrepancy between desired and existing outcomes by establishing goals for those needs that receive highest priority. Once priorities are established, schools must determine methods of resource utilization for accomplishing program goals.

The Dallas School Accountability Program is a two-part program. The District and School Improvement Plans are criterion-based. That is, the goals are absolute. They take current student achievement levels into consideration and project progress over a five-year period. School staffs work to achieve these goals. School and District Improvement Plans include goals in the following areas:

- Improve language arts skills (vocabulary, reading, oral competency, and writing skills, particularly TAAS objectives)
- Improve mathematics problem-solving, concept, and computational skills, particularly TAAS objectives
<table>
<thead>
<tr>
<th>TABLE 1-1</th>
<th>An Analysis of Political-Oriented Study Types (Paradigm-Evaluation)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Study Types</th>
<th>Definitions</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Political Orientation</td>
<td>Study Oriented Research Questions</td>
<td>Rely primarily on a positivist or negative view of an object from the perspective of...</td>
<td></td>
</tr>
<tr>
<td>Environment Evaluation Studies</td>
<td>Policy Oriented Framed Studies</td>
<td>Propagate an information search</td>
<td></td>
</tr>
<tr>
<td>Purpose</td>
<td>To acquire, manipulate, or increase a sphere of influence, power, or influence</td>
<td>To create a positive public image for an object</td>
<td></td>
</tr>
<tr>
<td>Scope of Question</td>
<td>Social interest groups</td>
<td>Public relations specialists and administrators</td>
<td></td>
</tr>
<tr>
<td>Main Question</td>
<td>What information would be best to possess or withhold in a proposed configuration</td>
<td>What information would be most helpful in securing public support?</td>
<td></td>
</tr>
<tr>
<td>Typical Methods</td>
<td>Content investigations and simulation studies</td>
<td>Based on surveys, experiments, and &quot;expert&quot; consultants</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 1-2</th>
<th>An Analysis of Question-Oriented Study Types (Quasi-Evaluation)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Questions Orientation</th>
<th>Study Types</th>
<th>Definitions</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quasi-Evaluation</td>
<td>Study Oriented Research Questions</td>
<td>Rely primarily on a positivist or negative view of an object from the perspective of...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environment Evaluation Studies</td>
<td>Policy Oriented Framed Studies</td>
<td>Propagate an information search</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purpose</td>
<td>To acquire, manipulate, or increase a sphere of influence, power, or influence</td>
<td>To create a positive public image for an object</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scope of Question</td>
<td>Social interest groups</td>
<td>Public relations specialists and administrators</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main Question</td>
<td>What information would be best to possess or withhold in a proposed configuration</td>
<td>What information would be most helpful in securing public support?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Typical Methods</td>
<td>Content investigations and simulation studies</td>
<td>Based on surveys, experiments, and &quot;expert&quot; consultants</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Legend:
- Political Orientation: Study Oriented Research Questions
- Environment Evaluation Studies: Policy Oriented Framed Studies
- Purpose: To acquire, manipulate, or increase a sphere of influence, power, or influence
- Scope of Question: Social interest groups
- Main Question: What information would be best to possess or withhold in a proposed configuration
- Typical Methods: Content investigations and simulation studies

Developers:
- Broom, Harransted, Mertall, and Michael, Popkins, and Perry
- Searle and Wachter, Kappas, and Glass
- Flanagan, Lord and Novak, Hymans, Theobald, and many more
- Kautzman
<table>
<thead>
<tr>
<th>Approaches</th>
<th>Value-Oriented (Use Evaluation)</th>
<th>Qualitative Evaluation</th>
<th>Content Analysis</th>
<th>Deconstruction of the Textual Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Type</td>
<td>Accreditation/ Certification Studies</td>
<td>Policy Studies</td>
<td>Decision-oriented Studies</td>
<td>Consumer-oriented Studies</td>
</tr>
<tr>
<td>Advance Organizational guidelines</td>
<td>Policy issues</td>
<td>Decision questions</td>
<td>Socio-political values and needs</td>
<td>Localized concerns and issues</td>
</tr>
<tr>
<td>Purpose</td>
<td>To determine whether institutions, programs and personnel should be approved to perform specific functions</td>
<td>To identify and assess the personal costs and benefits of continuing policies for a given jurisdiction or society</td>
<td>To provide a knowledge and value base for making and defining decisions</td>
<td>To foster understanding of alternative educational goals and services</td>
</tr>
<tr>
<td>Scope of Questions</td>
<td>Accreditation/ Certification agencies</td>
<td>Legislation, policy boards, and special interest groups</td>
<td>Decision-makers (administrators, teachers)</td>
<td>Society at large, consumers and the evaluator</td>
</tr>
<tr>
<td>Main Questions</td>
<td>Are institutions, programs, and personnel meeting minimal standards; how can they be improved?</td>
<td>Which of two or more competing policies will maximize the achievement of valued outcomes at a reasonable cost?</td>
<td>Which of several alternative consumable objects is the best buy, given their costs, the needs of the consumer, and the values of society at large?</td>
<td>What are the factors and conditions that distinguish an object from others?</td>
</tr>
<tr>
<td>Typical Methods</td>
<td>Self-study and sites visits by experts</td>
<td>Surveys, needs assessment, case studies, advocate triangles, comparison, and experimental design.</td>
<td>Checklists, needs assessment, goal- free evaluation, experimental and quasi-experimental design, health open-ended analysis, and case analysis</td>
<td>Case studies, advocacy reports, social inquiry, responsive evaluation</td>
</tr>
<tr>
<td>Pioneers</td>
<td>College Distance Examination Board (1921)</td>
<td>Rose</td>
<td>Crane</td>
<td>Stake</td>
</tr>
<tr>
<td>Development</td>
<td>Cooperative study of secondary school students (1935)</td>
<td>Gilmore, Jena, Niles, Owen, Wolf</td>
<td>Alton, Allbourn, Brink, Etta, Coba, Mercier, Ott, Roediger, Webster</td>
<td>Glass</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TABLE 2</td>
<td>Four Types of Evaluation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>--------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONTEXT</td>
<td>EVALUATION</td>
<td>INPUT</td>
<td>EVALUATION</td>
<td>PROCESS</td>
</tr>
<tr>
<td>OBJECTIVE</td>
<td>TO define the institutional context; to identify the target population and assess its needs; to identify opportunities for addressing the needs; to diagnose problems underlying the needs; and to judge whether proposed objectives are sufficiently responsive to the assessed needs.</td>
<td>To identify and assess system capabilities; alternative program strategies; procedural designs for implementing the strategy, budget, and schedule.</td>
<td>To identify and predict, in process, defects in the procedural design of its implementation; to provide information for programmed decision; also to record and judge procedural event and activities.</td>
<td>To collect descriptions and judgments of outcomes; to relate them to objectives and to context, input, and process; information; and to interpret their worth and merit.</td>
</tr>
<tr>
<td>METHOD</td>
<td>Systems analysis; survey; document review; interviews; experts; Delphi technique.</td>
<td>Literature search; visit to exemplary programs; advocate teams; and pilot trials, with the objective of inventorying and analyzing available human and material resources, solution strategies, and procedural designs for relevance, feasibility, and economy.</td>
<td>Monitoring the activity's potential procedural barriers and remaining alert to unanticipated costs; obtaining specified information for programmed decisions; describing the actual process; and continually interacting with and observing the activities of project staff.</td>
<td>Defining operationally and measuring outcome criteria; collecting judgments of outcomes from stakeholders; and performing qualitative and quantitative analyses.</td>
</tr>
<tr>
<td>RELATION TO DECISION-MAKING IN THE CHANGE PROCESS</td>
<td>For deciding on the setting to be served, the goals associated with meeting needs or using opportunities, and the objectives associated with solving problems (i.e., for planning needed changes); and to provide a basis for judging outcomes.</td>
<td>For selecting sources of support, solution strategies and procedural designs (i.e., for structuring change activities); providing a basis for judging implementation.</td>
<td>For implementing and refining the program design and procedure (i.e., for effecting program control), and providing a log of the actual process for later use in interpreting outcomes.</td>
<td>For deciding to continue, terminate, modify, or refocus a change activity; and presenting a clear record of effects (intended and unintended, positive and negative).</td>
</tr>
</tbody>
</table>
Figure 1. Schematic Depicting the School Improvement Process

SCHOOL IMPROVEMENT PLAN

Needs Assessment → School Program Planning (Objective Setting) → School-Community Council (SCC) → Operations Approval

Program and Planning Services → Training and Development → Programmatic Remedies → Administrative Support

Area Directors → Curriculum and Instructional Services and Supplemental Instructional Services Expertise → Action Research

Other Information about Successful Programs or Strategies

Interpretation by SCC → Program Evaluation → Program Implementation

Special Report on Pupil Achievement (Blue Books), School Effectiveness Indexes, Other Summative Assessments provided by the Division of Evaluation and Planning Services → In-Course Adjustments Determined by the SCC → Program-Monitoring Information
• Improve social studies skills, particularly TAAS objectives
• Improve science skills, particularly TAAS objectives
• Increase parent/community involvement
• Improve attendance (student and teacher)
• Facilitate dropout reduction and recovery a
• Increase promotion/course passing rate
• Increase enrollment in advanced courses, diploma plans, and honors programs 4
• Increase college entrance test participation/performance b

4 Applicable to middle and high schools only
5 Applicable to high schools only

Goals are set by determining the discrepancy between actual and desired levels of student accomplishment and projecting needed annual gain over a five-year period so that actual and desired performance become the same. The main problem with the approach is that the goals are not set empirically based on reasonable expectations, thus requiring the lowest achieving students to show the greatest amount of improvement. Those are the very students that have the most difficulty showing improvement and, therefore, the students whose schools should be rewarded upon significant improvement.

The second and most important part of the program, from the standpoint of improvement, is norm-referenced. That is, expectations are set based upon empirically determined predictions. The schools that are leading the others in improving student achievement set the expectations for the other schools. That is, expectations are set based upon best practice. School staffs know that the expectations are possible because the better schools are actually achieving them. This is the effectiveness-index portion of the accountability system. Effectiveness indices will be discussed in the next section of the paper.

Site-based management does not argue that local building staffs necessarily know how to solve all of their problems. It does, however, place decision-making responsibility and accountability at the local level. Central staff become resources to the schools whose function it is to provide viable alternatives to solving school problems. The principal ultimately is responsible and accountable for meeting the important objectives of instruction. Central staff is responsible and accountable for providing viable alternatives for consideration by school staff. Each school also has a School Community Council consisting of school staff and members of the community. The procedure is the input evaluation phase of the school improvement process and will only work if Central Office Divisions are competent and can supply the needed expertise. If the needed expertise does not reside in the appropriate Central Office Divisions, schools will not request needed services and the entire system probably will fail.

After the collection of relevant input information feeding a preliminary program-planning stage, the School-Community Council determines whether or not sufficient resources are available to make the desired changes. Quite often, sufficient resources are not available and some compromise is necessary. In many cases, the lack of resources is not limited to the realm of cost and political feasibility, but, rather, seems from an insufficient base of knowledge. Thus, educators are often in the position of having sufficient material resources but insufficient informational resources. Once these decisions are made, the SIP is complete.

The program implementation phase is then entered and the individual school staff is responsible for providing continuous formative feedback relative to program implementation. The feedback falls primarily into two categories — process evaluation and interim product evaluation. Process evaluation has three major objectives: (1) the detection or prediction of defects in procedural design or its implementation during program implementation stages, (2) the provision of information for programmed decisions, and (3) the maintenance of a record of the implementation procedure as it occurs (Stufflebeam, et al, 1971). Thus, process evaluation
information keeps the school staff and the School-Community Council informed of the extent to which program implementation conforms to specifications and, from an evaluation standpoint, guards against the evaluation of fictitious events. It also provides a record of implementation that can be cross-indexed to program effect.

Much of the process evaluation which was, at one time, implemented by DSED evaluation personnel now must be implemented at the local level. The procedure is consistent with the accountability emphasis that is currently the philosophy of District management and the community. Since process evaluation is extremely expensive, many of the cutbacks in research and evaluation activities during the early 1980's were in the area of process evaluation.

Interim product evaluation provides periodic feedback to the school staff and the School Community Council relative to the attainment of specific subobjectives during the implementation phase. Thus, process and interim product evaluation reports inform program management as to implementation and goal-attainment levels while program adjustments are still feasible. Much of the interim product evaluation can be done through portfolios of student work, performance testing, protocol analysis, and teacher-made tests, measures that are not available through systemwide data. Teacher Evaluation Constellants from each school are being trained in these techniques. In cases where serious needs are identified by interim product evaluation reports, tactical plans are developed, as supplements to the SIP, to meet these needs.

Local school staffs are also encouraged and trained to design, implement, and interpret action research studies. With the movement of the District to site-based management and the related reduction of Central Office Staff, it is impossible to supply school staffs with information centrally produced pertaining to their many and varied needs. Action research is a process for problem-solving that is designed and implemented at the local building level. It is a process of taking and studying action and its corresponding consequences so that more effective action may be taken (Lewin, 1946, Town 1973, Joynt and Haynes, 1992). Expressed sequentially, action research involves a continuous recycling through four steps: (1) identification of needs, (2) development of plans of action to address these needs, (3) execution of these plans of action, and (4) formative evaluation of these plans. In open organizations such as schools, the strength of action research lies in its implementation by the organization's members in their respective work sites. In effect, members of the organization actually learn while they study problems in contexts that they generally perceive as relevant and important. The results are used to supplement the more formal information available from the District's Evaluation Department.

Upon completion of a given cycle of program implemention, usually one year, a series of summative product evaluation reports are prepared. These reports take the form of the Special Report on Pupil Achievement, a school-level report that provides up to four years of disaggregated data upon all relevant outcome and input variables and is used to determine whether or not schools meet their SIP goals. School Effectiveness Indices, and program evaluation reports disaggregated by school. These reports, as well as relevant action research studies compiled by school staff, become the needs assessments for the next year's program adjustments.

---

1 Evaluation consultants are teachers who are trained by the Division of Evaluation and Planning Services to provide evaluation and data interpretation services at the school level. Throughout the school year, the consultants participate in performance-based assessments so that they may learn to apply formalized evaluation techniques to their campus's school improvement plans. They will identify areas for school improvement, describe program activities, and periodically report information on program impact. The performance-based assessment thoroughly prepares consultants to design defensible evaluations, i.e. measure program implementation, to identify appropriate instrumentation, to generate instrument similar to the TASS through their microcomputers, to assess programs impact, and to compile and present reports for school improvement. The concept has not received overwhelming support because the schools would rather receive direct help from the District's Evaluation and Planning Services Division.
The careful reader will note that, at no point, does teacher evaluation enter the process. In fact, over 93% of the District’s teachers are rated as “exceeds expectations” or “clearly outstanding”, with the remaining 7% being predominantly rated “satisfactory” (Buchanan, 1993). District studies have shown no relationship between student achievement and teacher evaluation (Chadbourne, 1992). The condition is one which the District hopes to remedy soon. Of course, since it would not be appropriate to revise the teacher evaluation system to reflect outcomes and leave the administrator and support personnel evaluation systems in their current state of measuring nothing reliably, those systems also need to be revised to reflect District instructional emphases.

Effectiveness Indices

The heart of the District’s accountability system features value-added methodology and is known as school effectiveness indices. These indices may be computed in a number of different ways (Bryk and Lee, 1989; Frantz and Solganik, 1993; May, 1997; Mendro and Webster, 1997; Mitchell, 1993; Ortiz and Almaguer, 1993; Raudenbush, 1993; Reynolds and Walberg, 1992; Sanders and Horn, 1993; Webbed, Mendro, and Almaguer, 1993-94; Webster and Olson, 1988.) All of them, however, are designed to establish some type of measure of improvement as the basis for evaluating schools and/or programs.

It is important to stress that the effectiveness indices are norm-referenced. That is, in order for one school to rank high, another must rank low. If the District has a year in which marked improvement occurs, ranking low on the effectiveness indices is not nearly as disastrous as ranking low on those same indices when the District has a mediocre year. Thus, school rankings based on the effectiveness indices must be interpreted in light of improvement, or lack of improvement, on the goals specified in the SIP.

Most indices are computed on an individual student basis. There are some school level variables that are also included (promotion rate, graduation rate, dropout rate, SAT-and PSAT-percent tested, percent in accelerated courses, and percent in advanced diploma plant). All other variables are analyzed on a student-by-student basis. All indices are designed to reward improvement rather than absolute outcome levels.

The school effectiveness methodology defines a school’s effectiveness as being associated with exceptional measured performance above or below that which would be expected across the entire District. When a school’s population of students departs markedly from its own preestablished trend or from the more general trend of similar students throughout the District, the departure is attributed to school effect. The problem of measuring a school’s effect, then, becomes one of establishing the student levels of accomplishment on the various important outcome variables, setting expected levels of performance based on these, and determining the extent to which its students, on the average, exceed or fall short of expectation. The procedures involve regression analysis to compute prediction models by grade level or by school for each outcome variable, independent of school identification, and then using these equations within schools to obtain mean gains over expectations. A major feature of the approach also involves assigning relative weights to each of the outcomes. Once weighted levels of performance have been determined, the methodology provides an indicator of how well a school performs relative to other schools throughout the District. A step-by-step description of the process follows, along with associated reasons for each step.

Step 1. Appoint or elect an Accountability Task Force. The Task Force should have representatives from all constituencies involved with the school district and be charged with the responsibility of overseeing the accountability system. The Accountability Task Force deals with many aspects of the accountability system including methodology, testing, determining and weighting important performance variables, and determining the roles for financial awards related to the system.
Why? To the degree that educational goals vary, school effectiveness has limited meaning. Therefore, the first step in developing methodology to identify effective schools must be the development of a community consensus as to the important goals of schools.

Step 2. Determine eligible students for each criterion variable. Eligible students include all students who:
- are enrolled continuously in a specific school from the end of the first six weeks,
- have the necessary preobservation data in the DISD and postobservation data for the appropriate school year in that specific school, and
- are eligible for the testing program according to the DISD Systemwide Testing Policy (see the testing variables).

All eligible students will be included in the cohort longitudinal analysis. Thus, in order to be included as a member of a given school's cohort, a student must be enrolled in that school by the end of the first six weeks, have the necessary preobservation data, and be tested in that school in accordance with DISD policy through the systemwide testing program. Students who transfer out of a school and back into that school over a short period of time will be included in that school's cohort. The definition of a "short period of time" is subject to change based on empirical data. Schools that, in the opinion of the Accountability Task Force, attempt to manipulate their continuously enrolled student population will be disqualified from the Awards Program.

Why? Schools are held accountable only for the outcome levels of students who have been exposed to that school's instructional program. That is, schools are held accountable only for their continuously enrolled students. Since all equations are based upon growth, and therefore require a baseline, preobservation data must be available.

Step 3. Regress each of the student-level predictors and each of the criterion variables on the background variables. (Background variables include main effects and all possible interactions for student-level variables: gender, ethnicity, Limited English Proficiency status, socioeconomic status, and main effects for the school-level variables - overcrowdedness and mobility statistics.) Obtain residuals for each student-level criterion and predictor variable.

Why? The procedure removes the variance attributed to important background variables over which the schools have no control. This "levels the playing field" and addresses practitioners' concerns about the impact of background variables on outcomes for which they are held accountable. Background variables generally account for 15-20 percent of the variance in student achievement and other outcomes.

Step 4. Divide the predictor space into 256 arrays. Standardize the residuals for both the predictor and criterion variables.

Why? The procedure assures that schools which have unusual numbers of students in certain areas of the predictor space will not be ranked based upon differential variance in different arrays. Assures array normality.
Step 5. Use standardized residuals to develop the next level of equations (both predictor and criterion variables are residuals).

Why? After accounting for variance due to the background variables, there is still considerable variation due to individual student ability levels. Accounting for the variation in the next level assures that schools are not penalized or rewarded for the ability levels of the students who walk in their door rather than the school's own effect on these students.

Step 6. Carefully examine the data. Utilize an all-possible regressions procedure to arrive at the best models for predicting the various residualized criterion variables from the best-residualized predictors. Through examination, preserve the maximum number of students with the maximum amount of variance for accountability. Scrutinize the percentage of loss in the number of students versus the percentage of gain in predictability. Make sure the relationships are linear. If not, use an appropriate transformation. Obtain residuals for each student on each outcome variable.

Why? The procedure assures that one has the best models to predict growth, or lack of same, and establish individual predictions for each student. According to Bryk and Raudenbush (1992), good data analysis begins with a careful examination of the univariate frequency distribution of each variable that may be employed in a subsequent multivariate analysis. Examination of the shape and scale of each variable provides a check on the quality of the data, identifies outlying observations, and may suggest a need for a variable transformation. The next step in model building involves exploration of the bivariate relationships. Plots of two continuous variables can identify possible nonlinear relationships and can identify discrepant cases that could arise from some erroneous observations.

Note: Only one year of historical data are used. That is, a stepwise regression approach is used on the residuals of multiple predictors so that, in most cases, satisfactory prediction is achieved without having to go back more than one year. This maintains the degrees of freedom associated with the equations since, in an urban district, each additional year of data used significantly reduces the degrees of freedom associated with the equations.

Step 7. Divide the predictor space into 256 arrays. Standardize the residuals for each of the criterion variables.

Why? The procedure assures that schools derive no particular advantage by starting with high-scoring or low-scoring students or with students at a particular point in the predictor space. While the equations in Step 6 established individual expectations for each student based on that student’s performance on the predictor(s) of interest and determined that lower scoring students have lower predicted scores and higher scoring students have higher predicted scores, the step assures that the schools will not be advantaged or disadvantaged by differential variance in the predictor space at different points along the regression line.

Step 8. Associate individual student residuals with the schools. Obtain mean residuals on each of the criterion variables.

Why? A gross estimate of school effect is produced. The more positive the mean residuals, the more effective the school. That is, the more the school's students exceeded
The effectiveness indices, computed at the school level, have obvious application to be used as part of a personnel evaluation system. The District is currently exploring the feasibility of utilizing a three-level hierarchical linear model to estimate teacher effects and school effects. Hierarchical linear modeling models the hierarchical structure found in most educational data and holds promise for further reducing the error term associated with prediction and for solving the degrees of freedom problem associated with the small numbers that often occur when computing effectiveness indices at the class or teacher level. Current thinking suggests that this methodology might be best applied after Step 5 in the previously outlined procedure.

**A Unified Personnel Evaluation Model**

It is important to develop a personnel evaluation model that includes provisions for the evaluation of all personnel. That is, the evaluation system should not hold teachers to stringent criteria while allowing administrator evaluation to continue to be loose and amorphous. In addition, where possible, personnel evaluation should be double-nested. That is, the evaluation of the principal should be coupled to the evaluation of his or her teachers just as the evaluation of the superintendent should be coupled to the evaluation of principals. To the extent possible, many of the same outcomes should be used for both sets of evaluations.

Sconce and Helm (1990, 1991, 1992) developed a seven-step model for the evaluation of teacher personnel in education while Stufflebeam posited a twelve-step model for the evaluation of school administrators (Stufflebeam, 1994). Both systems are grounded in the Professional Standards for Personnel Evaluation. Scriven (1994) suggests a common model for the evaluation of all personnel that incorporates the major characteristics of the aforementioned two approaches. The five-step model calls for:

---

**Step 5.** Multiply the mean residuals associated with the schools by $\overline{y}$.  

**Why?** Intuitively speaking, the procedure weights the results by the size of the school. That is, larger schools are given more weight for their accomplishments because it is harder to move larger groups of students. Statistically speaking, the procedure equals the variance of the different school means.

**Step 10.** Restandardize the distribution of mean residuals on each criterion variable to a mean of 0 and a standard deviation of 10.  

**Why?** After Step 9, distributions of mean residuals no longer have identical means and variances. Before Step 11 can be accomplished (weighting and combining variables), all distributions must have equal means and variances.

**Step 11.** Multiply mean standardized residuals by the appropriate weights assigned by the Accountability Task Force and aggregate for each school.  

**Why?** The procedure allows the variables to be weighted on the basis of perceived importance by District constituents, i.e., the Accountability Task Force.
1. Needs Assessment. The step includes Stronge’s identification of system needs and identification of personnel duties as well as Stufflebeam’s review of student outcome data, staff performance, and system needs plus the update of job descriptions and definition of priorities as well as the review and strengthening of each position’s work plan and schedule of activities.

2. Evaluation of Staff. The step includes Stronge’s selection of performance indicators, setting of performance standards, documentation of job performance, and the evaluation of performance. Stufflebeam’s definition of evaluation users and selection of performance indicators and weights, definition of performance standards, documentation of performance and accomplishments, collection of stakeholder input, and assessment of the work environment are included in the step. Considerations of worth and merit also enter into the step.

3. Evaluation of the Superintendent. The evaluation of the superintendent should be related closely to the evaluation of his/her staff. A superintendent should not be permitted to thrive if his/her staff are incompetent. Formative feedback should be provided at all levels of the organization and summative evaluation compiled on all personnel.

4. Professional Development. The step includes Stronge’s improve-and-maintain professional services and Stufflebeam’s improve professional service. Personnel at all levels of the organization must have the opportunity to improve, meaning that the personnel evaluation system must be linked to the personnel development system and that meaningful personnel development must occur.

5. Recommendations and Actions Based Upon The Above. If personnel development doesn’t produce the desired product, action must be taken to eliminate the problem. If the action meets the standards of the various professional organizations, it will pass the test of due process.

Scriven (1994) advances an approach to teacher evaluation based upon the duties of the teacher. He classifies duties into five areas: knowledge of subject matter, instructional competence, assessment competence, professionalism, and other duties to the school and community. CREATE (Center for Research on Educational Accountability and Teacher Evaluation), through its Dallas Project, will expand the concept of duties-based evaluation to administrative as well as teaching personnel.

Table 4 cross-indexes the five duty-areas with data collection strategies for documenting each. Student outcomes are a major part of documentation. The approach outlined in Table 4 can be used in evaluating all teaching and administrative personnel.

Judgments would be collected in two ways. From parents, peers, and community members, judgments would be collected from an open-ended survey instrument which would solicit one example of good practice, one example of bad practice, and an overall judgment as to the merit of the person. Teachers would then also be rated by their students (grades 7-12), as would principals by their teachers. These ratings would be on a Likert scale.

Objective data would involve the collection of outcome data on each individual. The collection is the piece that student achievement and other outcome data would be compiled. Rather than use absolute achievement data, effectiveness indices would be used to control for different student background and ability levels. A number of different outcomes would be utilized. The unit of analysis for the teacher would be the classroom, while the school would be the unit of analysis for the superintendent.

13
All personnel would be encouraged to keep portfolios of their work. These portfolios would provide additional evidence about their performance of the duties of their position. Teachers, for example, could keep the results of student performance tests administered in their classrooms as evidence of student performance that would augment the objective observational data.

If the first three data collection methods yielded poor results, a structured interview would be undertaken to determine if, and to what extent, weaknesses existed in the requisite knowledge and skills necessary to perform the duties of the position. The results of the interview, plus student skills profiles of strengths and weaknesses gleaned from the skills analyses provided through the testing program, would then be used to channel the teacher or principal into meaningful staff development.

A truly comprehensive evaluation system must also hold central administrative staff responsible for the quality of their contributions to the schools. Thus, the quality of services under the input section of Figure 1 must be evaluated regarding their contributions to improving the quality of education as measured by the same outcomes for which the teachers are accountable. Central staff would be evaluated using the methods and duties outlined in Table 4. All levels of personnel evaluation would go through the same steps outlined earlier in the paper.

<table>
<thead>
<tr>
<th>Duties</th>
<th>Judgment</th>
<th>Objective Data</th>
<th>Portfolio</th>
<th>Interview to Determine Requisite Knowledge and Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge of subject matter</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Instructional Competence</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Assessment Competence</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Professionalism</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Duties to School and Community</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Teacher Evaluation**

A comprehensive personnel evaluation system must start with legitimate evaluation of the classroom teacher. The primary purpose of teacher evaluation is to safeguard and improve the quality of student instruction. Millman and Darling-Hammond (1990) provided a list of functions of teacher evaluation which included preservice evaluation of teachers, licensure and certification of teachers, teacher selection, assistance and assessment for beginning teachers, evaluation for professional development beyond minimum competence, evaluation for tenure and dismissal, and, teacher evaluation for school improvement.

Despite the importance of the aforementioned functions of teacher evaluation, the traditional model of teacher evaluation which is based upon presumably systematic classroom observations and continued contact with each teacher throughout the school year to develop a "preponderance of evidence" is still prevalent today. The Texas Teacher Appraisal System is based on the traditional model of teacher evaluation. Bloom, et al., (1981) defined formative, summative, and diagnostic evaluations based upon a taxonomy of thinking skills developed into educational and, subsequently, evaluative objectives. Hunter (1983) stresses a management-by-objectives approach to coaching, supervision, and evaluation of teachers based upon multiple observations.
of teacher behavior. Iwanicki (1993) in attempting to improve upon the prevalent model, suggests the use of portfolios to achieve a cooperative evaluation process. Student outcomes are not one of the standard criteria in the process.

Interest in performance-based or outcomes-based teacher evaluation dates all the way back to 15th century Italy where a teacher's salary was dependent upon his/her students' performance (Texas Education Agency, 1988). Despite the long-term interest, the use of student achievement data to evaluate teachers is highly controversial. Student achievement data seem to provide the most direct evidence of teacher effect. While classroom observations and ratings by principals, teachers, and students measure a teacher's behavior on the job, student achievement data relate directly to the outcomes of that behavior. Most current uses of student achievement data to evaluate teachers are in the realm of incentive programs. That is, student achievement data are not generally used in the mainstream teacher evaluation system (Bachrach, Lipsky, and Shedd, 1984).

In 1983, a national survey of merit-pay programs reported that nine school districts in seven states (Arizona, New Hampshire, North Carolina, Oklahoma, South Dakota, Texas, Utah) used student test scores as evaluative criteria in determining merit pay for classroom teachers (Cilhoue & Protheroe, 1983). In all but two of the districts (Dallas and Rockston), student achievement served as the only evidence of teacher performance. As of January, 1990, student achievement was a criterion of teacher performance in one-third of all statewide teacher incentive/school incentive/career-ladder programs. These programs have been fully implemented in four states (Florida, South Carolina, Tennessee, Utah), are at pilot stages in four states (Arizona, Kentucky, Maine, South Carolina), and are under development in four states (Alabama, Florida, Georgia, Texas). A school incentive program based upon student achievement is also under consideration in Alaska, and several career-ladder or merit pay programs based upon student performance have been implemented by local districts (e.g., Campbell County and Davielle, Virginia).

The issue really is not one of whether or not student achievement data should be used in teacher evaluation, but rather, entails a methodological debate over ways to operationalize and implement such a system. Unfortunately, the preponderance of literature in the field concentrates upon reasons student achievement data cannot be used for teacher evaluation rather than upon credible ways to use it. Some of the concerns raised in the literature include:

- the development of procedures to account for the difficulty in measuring the long-term development of skills which may not be measured in year-to-year growth patterns (TBA, 1988).

- the assessment of diverse areas of achievement which do not have readily available standardized tests is an area of concern when dealing with non-academic areas teachers.

- programs which pull out students for remediation, programs which involve team-teaching, and programs with extensive use of instructional aids inhibit the estimation of an individual teacher's contribution to improved student achievement.

- norm-referenced standardized test sample broad subject domains and are unlikely to match closely the curriculum in particular classrooms at particular times (Hearst, 1956).

- well-established, broadly applicable, and accepted achievement measures are not available in all the relevant areas of learning (Banc, 1985).

- standardized achievement tests are unlikely to reflect the full range of instructional goals in their subject areas. Norm-referenced tests tend to ignore the higher-order
skills. Therefore it is likely that products of superior teaching are not measured adequately or completely by standardized achievement tests (Bano, 1985).

- what the student brings to the classroom in terms of ability, home and peer influence, motivation and other influences is very powerful in affecting academic achievement at the end of the year (Inwitsicki, 1986).

- the statistical methods used to control for non-teacher factors cannot take into account all of the relevant factors. More importantly, the methods will be incomprehensible to those being evaluated and difficult to defend in public (Bano, 1985).

- non-statistical models for controlling non-teacher factors are easier to explain, but cannot take into account most of the necessary circumstances (Bano, 1985).

- attempting to use any one of a number of regression-based techniques at the teacher level creates a rather subtle problem related to the statistical concept of "degrees of freedom." In general, the number of degrees of freedom upon which a statistical procedure is based depends on the sample size (N) and the number of sample statistics (i.e., variables in multiple regression). The sample size (i.e., number of students) for a teacher is relatively small to start with. However, the useable sample size becomes even smaller because development of the regression equations requires existing test scores for each student for at least two successive years. As an example, a second-grade teacher may have a class of 22 students, but may only have test scores from the first grade for 11 of those students. Since degrees of freedom also depends on the number of variables in the multiple regression equation, a regression equation with four (4) variables would leave seven (7) degrees of freedom. The stability of a projected regression line is primarily dependent on the number of degrees of freedom. Seven is generally not enough for stable estimates. As a general rule of thumb, forty students per variable has been recommended as a minimum number upon which to base a projected regression line.

Nontechnical concerns include the concern that objectives that are not measured by the tests will be omitted by teachers. That other duties such as playground supervision and school committee work may be slighted, and that, with each teacher being rated separately, the collegiality necessary to build good instructional teams within a building may be damaged.

Most of the methodological issues raised above can be resolved. (1) Longitudinal growth curves, or alternatively, relationships based upon two years of data, can be formulated upon important outcome variables. In the case of relationships based upon two years of data, replication is necessary to assure greater reliability. (2) Criterion-referenced tests can be developed and used to assess diverse areas of achievement. (3) In cases where there are pull-out or send-in programs, team teaching, or instructional aides, data can be provided at the team level rather than at the individual teacher level. (4) Measures in addition to norm-referenced tests can be used. (5) Constituents are primarily interested in basic skills. To the extent that measures are needed in music, art, physical education, etc., they can be developed. (6) Criterion-referenced tests can be used to measure higher-order thinking skills. In addition, performance testing can be used as one outcome variable with the outcomes being weighted by the reliability of the instruments. (7) What the student brings to the classroom in terms of background variables can be statistically controlled. These variables typically account for 9-20% of the variance in student achievement (Webster, Mendro, and Almquist, 1993). (8) It has been the author's experience that gender, ethnicity, limited English proficiency status, and free or reduced-lunch status, plus their interactions, account for most of the variance that can be attributed to background variables. They are easy to explain and defend. (9) Non-statistical models for controlling non-teacher factors are misleading and should not be used (Webster and Edwards, 1993). (10) The degrees
of freedom problem is real in this one must worry about the stability of the regression line when it is applied to one teacher. The actual student models are developed Districtwide and then applied to individual teachers so that the actual regression lines are extremely stable. At the teacher level, replication over several years is the best safeguard against errors because of small sample size.

In terms of the non-technical concerns, it is not proposed that student test scores be the only thing used in the teacher evaluation system. The system can be expanded to include other duties (see Scribner, 1994). In the area of collegiality, the system is being developed within the context of an overall school incentive system. Teachers will not be rewarded on an individual basis but rather as part of the school team. Teacher evaluation data will be used by the principal in evaluating the teacher. Teacher evaluation data aggregated at the school and district level would be used to evaluate administrative staff.

The Teacher Evaluation System

The Texas Teacher Appraisal System, used for teacher evaluation throughout the State of Texas, has five domains upon which teachers are rated. It is an instrument that prohibits reliable teacher evaluation. The five domains are Instructional Strategies, Classroom Management and Organization, Presentation of Subject Matter, Learning Environment, and Professional Growth and Responsibilities. At no point does student achievement enter the equation. There are thirteen different sub-areas under the five domains. The system must be replaced by a system that emphasizes duties as outlined in Table 4. The output of the effectiveness indices at the classroom level would be used to determine impact of instruction on student achievement. Since the indices are standardized to a mean of 50 and a standard deviation of 10, the statistics are easily interpretable by all concerned. The system will be established so that the indices can be replicated each year and action will not be taken until the second year of system operation. Even then, student outcomes will be just one variable in the evaluation system (as outlined in Table 4). All available norm-referenced and criterion-referenced test scores will be used in this analysis.

Table 5 gives an example of a teacher effectiveness report for a secondary Language Arts teacher. Included, by class, are the effectiveness indices on the Texas Assessment of Academic Skills (TAAS) Reading and Writing subtests, as well as absolute scores. Also shown are school and District statistics upon each of these variables. In addition, course grade distributions are shown as well as similar scores and effectiveness indices on the Texas Assessment of Proficiency (TAP) and the English I Assessment of Course Performance (ACP). All of these data would be collected under the objective data heading in Table 4.

The data in Table 5 are simulated. They are for an English I teacher. The teacher is a relatively effective teacher. His/her students achieved gains on all of the cognitive measures and consistently ranked above 50.0 on the effectiveness indices. His/her class average daily attendance is good and he/she had only two absence. On the ACP, the standardized final examination for the course, his/her students did extremely well achieving an effectiveness index of 56.9. The recommendation would be to leave the teacher alone to teach. A solid profile, such as the above mentioned, would weigh heavily in the evaluation of the teacher. Judgments by peers, students, parents, etc., would weigh less heavily, unless some question of moral turpitude was brought forward. The structured interview and examination of the portfolio would not be necessary.

If specific skill areas were identified as weak, the system would then produce class profiles of student skill areas which would be linked to staff development in an attempt to remediate weaknesses. The process, along with the results of the portfolio analysis and the structured interview, would provide a blueprint for personnel development.
How do all of the procedures outlined fit into program evaluation? Figure 2 provides a flow chart of the teacher evaluation and improvement system, which would operate in tandem with the program evaluation system specified in Figure 1 and would be part of the context evaluation data base that feeds program design. The system would provide evaluative data at the teacher, school, District, and program level and would provide the information necessary for continuous improvement. Thus, the same system could be used for the evaluation and improvement of personnel at every level in the district.

The personnel evaluation system, like the program evaluation system, begins with an annual needs assessment. The needs assessment includes the identification of system needs and personnel duties as well as a review of student outcome data. Once needs have been identified, personnel goals must be established that focus upon reducing the discrepancy between desired and existing outcome by establishing goals for those needs that receive the highest priority. Planning focuses upon determining the best method to proceed from current levels of important outcomes to desired levels of those outcomes and culminates in the production of a professional growth plan.

Once the professional growth plan has been developed, the question of whether or not appropriate training is available is raised. If appropriate training is available, that is, training that has been previously evaluated and found to be effective, it is prescribed. The individual is then exposed to the training. If satisfactory performance results, the individual is cycled back into the system. If satisfactory performance does not result, the individual is terminated with cause. The number of iterations through the cycle before termination is determined primarily by considerations of worth.

If appropriate training is not available and if resources are available, input evaluation is undertaken to find or develop a promising training strategy. Process and product evaluation of the training strategy is crucial since the system cannot hold personnel responsible for improvement if real help isn't provided. All of the evaluation must, of course, be conducted according to the Professional Standards for Personnel Evaluation (propriety, utility, accuracy, feasibility).

A system designed along the lines of the system outlined in this paper is being developed through a cooperative project between CREATE and the Dallas Independent School District. It will meet a great amount of opposition from teacher groups but absolutely is essential if the school District is going to reach its goal of educating all children. Programs alone are not sufficient. Competent personnel are crucial. Good teachers and administrators should be rewarded and emulated, mediocre teachers and administrators should be developed and improved, and incompetent teachers and administrators must be improved or eliminated.

The most difficult tasks include operationally defining personnel duties and providing valid quantitative and qualitative measures of success and failure. While, as the literature suggests, these tasks are difficult, they are not impossible, and absolutely must be accomplished if we are to improve significantly American education.

Summary and Conclusions

This paper has discussed models for the evaluation of programs, schools, and personnel. Since the purpose of the series of papers contained in this symposium is to develop a common model of evaluation, it is useful to examine the common characteristics of the program/school evaluation model outlined in Figure 1 and the teacher evaluation model specified in Figure 2.
<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Pre</th>
<th>Post</th>
<th>Effect</th>
<th>N</th>
<th>Pre</th>
<th>Post</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21</td>
<td>68</td>
<td>75</td>
<td>51.68</td>
<td>26</td>
<td>73.3</td>
<td>79.7</td>
<td>53.2</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>71</td>
<td>73</td>
<td>50.00</td>
<td>25</td>
<td>74.6</td>
<td>81.8</td>
<td>58.3</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>59</td>
<td>67</td>
<td>54.79</td>
<td>26</td>
<td>60.5</td>
<td>69.6</td>
<td>51.9</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>64</td>
<td>78</td>
<td>48.63</td>
<td>26</td>
<td>63.7</td>
<td>72.6</td>
<td>52.2</td>
</tr>
<tr>
<td>Total</td>
<td>81</td>
<td>85</td>
<td>69</td>
<td>51.28</td>
<td>103</td>
<td>68.0</td>
<td>75.7</td>
<td>56.9</td>
</tr>
<tr>
<td>Other Teachers in School</td>
<td>175</td>
<td>61</td>
<td>60</td>
<td>49.73</td>
<td>225</td>
<td>62.0</td>
<td>64.3</td>
<td>48.7</td>
</tr>
<tr>
<td>District</td>
<td>6219</td>
<td>47</td>
<td>52</td>
<td>50.00</td>
<td>6600</td>
<td>63.6</td>
<td>86.7</td>
<td>50.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>Pre</th>
<th>Post</th>
<th>Effect</th>
<th>Pre</th>
<th>Post</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>57.1</td>
<td>59.2</td>
<td>49.79</td>
<td>*</td>
<td>86</td>
<td>53.71</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>58.0</td>
<td>55.1</td>
<td>48.19</td>
<td>*</td>
<td>83</td>
<td>52.68</td>
</tr>
<tr>
<td>3</td>
<td>23</td>
<td>46.7</td>
<td>50.2</td>
<td>53.28</td>
<td>*</td>
<td>79</td>
<td>51.19</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>47.5</td>
<td>47.6</td>
<td>46.10</td>
<td>*</td>
<td>62</td>
<td>47.17</td>
</tr>
<tr>
<td>Total</td>
<td>83</td>
<td>52.3</td>
<td>54.0</td>
<td>49.35</td>
<td>*</td>
<td>77</td>
<td>52.19</td>
</tr>
<tr>
<td>Other Teachers in School</td>
<td>181</td>
<td>50.2</td>
<td>50.8</td>
<td>47.85</td>
<td>*</td>
<td>71</td>
<td>48.16</td>
</tr>
<tr>
<td>District</td>
<td>6308</td>
<td>46.1</td>
<td>50.7</td>
<td>50.00</td>
<td>*</td>
<td>70</td>
<td>50.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade Distribution (%)</th>
<th>Class 20A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>Pre 18-30</td>
</tr>
<tr>
<td>1</td>
<td>98.3</td>
</tr>
<tr>
<td>2</td>
<td>96.5</td>
</tr>
<tr>
<td>3</td>
<td>95.2</td>
</tr>
<tr>
<td>5</td>
<td>98.2</td>
</tr>
<tr>
<td>Total</td>
<td>98.3</td>
</tr>
<tr>
<td>Other Teachers in School</td>
<td>98.1</td>
</tr>
<tr>
<td>District</td>
<td>97.3</td>
</tr>
</tbody>
</table>

*Percent scores not available

Teacher Absences (Days)
Teacher 2
Teachers in School 6 (median)
District 6 (median)
Figure 2. A Flow Chart of the Teacher Evaluation and Improvement Process

1. Needs assessment
2. Annual personnel evaluation
3. Objective Data Satisfactory?
   - Yes
   - No
     - Weakness from Judgments
       - Yes
       - No: Effective staff development (Professional Growth Plan)
     - Yes: Definition of the problem through interviews, classroom observations, judgments, profiles of student performance, portfolio analysis (Professional Growth Plan)
4. Appropriate training available?
   - Yes
   - No: Resources available?
     - Yes: Input evaluation
     - No: Promising training strategy needed?
6. Implementation of special training
7. Development and testing
8. Satisfactory program performance?
   - Yes
   - No: Project worthy of further effort?
     - Yes
     - No: Process and product evaluation
9. Training provided

Abort
Both approaches are grounded in the CIPP Model. That is, both include context, input, process, and product evaluation. Both begin with needs assessment (context evaluation). In order for the needs assessments to be accomplished, both must consider standards, duties, and important outcomes. Both are therefore grounded in an evaluation of student needs. Both also focus on improvement and accountability.

The principle difference between the two models is that the program evaluation model is focused at the program or school level while the teacher evaluation model is focused at the individual teacher level. It is also important to note that the teacher evaluation model, in fact a common model for the evaluation of all personnel, is an extremely important piece of program and school evaluation. Outside of these differences, the models are very similar and can be conceptualized as a common model for evaluation.
References


Texas Education Agency. (1988). *A Study to Determine the Most Effective Means of Implementing Career-Teacher Assignments That Are Made on the Basis of Student Achievement in Addition to Other Bases Required by Law*. Austin, Texas: Texas Education Agency, ERIC ED 309 143.


