DOCUMENT RESUME

ED 358 132 TM 019 915

AUTHOR Webster, William J.; Edwards, Marvin E.
TITLE An Accountability System for School Improvement.
PUB DATE Apr 93
NOTE 32p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993). For a related paper, see TM 019 913.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; *Accountability; Criterion Referenced Tests; *Educational Improvement; Educational Objectives; Effective Schools Research; Elementary Secondary Education; Incentives; Norm Referenced Tests; Outcomes of Education; Rewards; School Districts; *School Effectiveness; Student Characteristics; *Systems Development
IDENTIFIERS *Dallas Independent School District TX; Reform Efforts; *School Effectiveness Indices

ABSTRACT

The accountability system that is being implemented in the Dallas Independent School District (Texas) is a three-tier system. The first tier focuses at the school level. The second involves the District Improvement Plan, setting district accountability objectives. The third tier involves school effectiveness indices, which take into account important student background variables and provide information on how effective schools are with the students they serve. Accountability is operationalized in a criterion-referenced manner through an analysis of absolute outcomes relative to school and district performance on goals specified in the District Improvement Plan and the School Improvement Plan, and in a norm-referenced manner through school effectiveness indices. Schools and staffs are eligible for financial awards based on school performance on the effectiveness indices. The program does not reward individual competition between teachers within schools. To achieve the necessary improvement in student outcomes, teachers must work together. The program focuses on the important outcomes of schooling and affords all schools an opportunity to be distinguished, since the emphasis is not on absolute achievement, but on impact on the most students. The addition of the effectiveness indices makes the accountability system fair and valid. Six figures illustrate the discussion. (Contains 44 references.) (SLD)

An Accountability System For School Improvement [1]

William J. Webster and Marvin E. Edwards
Dallas Independent School District

As the nation progresses through the decade of the nineties, there is increased pressure from many segments of society for better educational accountability. This desire for accountability is often accompanied by societal skepticism of educators and the quality of the job that they are perceived to be doing. This perception has been fueled in recent years by the White House and the Department of Education and often used to support an educational agenda that pushes choice. The Texas Education Agency, as well as the education agencies of at least fifteen other states, have initiated programs that have increased focus on educational outcomes (Duttweiler and Ramos, 1966; Southern Regional Education Board, 1990). At the national level there is serious, though misguided, talk of a national achievement test (America 2000, 1991). In Dallas, a group of citizens appointed by the Board of Education developed a comprehensive plan for the Improvement of Dallas Schools (Commission for Educational Excellence, 1991). This plan called for rapid conversion from a school system to a system of schools and highlighted accountability as the linchpin for improvement.

The accountability system that is being implemented in the Dallas Independent School District (DISD), and is the subject of this paper, is a three tier system. The first tier focuses at the school level. Under the District's plan to move from a school system to a system of schools over a five year period, each school is held responsible and accountable for many aspects of its own operation. School Improvement Plans (SIP) are the vehicles through which this is accomplished. The second tier of the system involves

[1] Paper presented at the annual meeting of the American Educational Research Association, Atlanta, Georgia, April 12-16, 1993.

the District Improvement Plan (DIP). The DIP sets the desired levels on District accountability objectives and specifies how Central Office Divisions support the schools. The third tier involves school effectiveness indices. These indices take into consideration important student background variables and provide information on how effective schools are with the students that they serve.

One of the major concerns related to most accountability systems should be that of fairness. Educators who are caught up in the accountability movement have a right to know that the standards by which they are judged are fair. The system outlined in this paper attempts to incorporate fairness as defined by the Standards for Evaluation of Educational Programs, Projects, and Materials (Joint Committee on Standards for Educational Evaluation, 1981) and the Standards for Educational and Psychological Testing (AERA, APA, NCME, 1985). Where appropriate, this system will be compared to the accountability system being promulgated by the State of Texas and relative strengths and weaknesses enumerated. [2]

SCHOOL-CENTERED EDUCATION

The District's School-Centered Education Plan focuses control of most available resources and all instructional decisions at the local school level (Edwards, 1991). The only decisions that school level committees are not empowered to make are those involving the nature and magnitude of outcomes for which they are being held accountable. An extremely important step in the school improvement process is that of determining the important performance indicators that will inform educators, parents, and community members whether or not students are making satisfactory progress in the key developmental pathways that are critical for academic learning. These performance indicators are determined by an Accountability Task Force and influenced by the State's

---

[2] The Texas Education Agency's (TEA) accountability system is similar in many respects to other State systems. The TEA system is used for comparison purposes because the authors are very familiar with it.

Academic Excellence Indicator System. The accountability indicators are consistent across the three tiers of the accountability system.

The Accountability Task Force

The Accountability Task Force is a 27 member committee, appointed by the Board of Education, charged with the responsibility of overseeing the District's accountability system. The membership includes four elementary teachers, three middle school teachers, four high school teachers, four principals, four parents, five members of the business community, and three central office administrators. In addition, the various employee organizations each have an ex officio member on the task force. This task force deals with many aspects of the accountability system including methodology, testing, determining and weighting important performance variables, and determining the rules for financial awards that are related to the accountability system. The Accountability Task Force also hears any concerns or grievances relative to the accountability system.

The Comer Model

The DISD is implementing School-Centered Education through the Yale Child Study Center School Development Program (Comer, 1988). In describing the program, Comer states:

> In order to promote such change, mechanisms must be created that allow parents and staff to engage in a process in which they gain and apply child development systems, and individual behavior knowledge and skills to every aspect of a school program in a way and at a rate that is understandable and not threatening. Each successful activity outcome for staff, students, and parents encourages the staff to use these ways of working again, until the new way eventually replaces the old.
>
> In response to the conditions we found, working collaboratively with parents and staff, we gradually developed our present nine component (3 mechanisms, 3 operations, 3 guidelines) model: 1) a governance and management team representative of the parents, teachers, administrators and support staff; 2) a mental health or support staff team/ and 3) a parents program. The governance and management team carries out three critical

operations: the development of 4) a Comprehensive School Plan with specific goals in the social climate and academic areas; 5) staff/development activities based on building level goals in these areas; and 6) periodic assessment which allows the staff to adjust the program to meet identified needs and opportunities.

Several important guidelines and agreements are needed. Participants on the governance and management team 7) cannot paralyze the leader. On the other hand, the leader cannot use the group as a "rubber stamp." While the principal usually provides leadership to the governance and management group, 8) decisions are made by consensus to avoid "winner-loser" feelings and behavior. 9) A "no-fault," problem-solving approach is used by all of the working groups in the school, and eventually these attitudes permeate the thinking of most individuals.

The Comer model is a shared decision-making model. In implementing this model, a number of common beliefs are assumed (Edwards, 1993). These beliefs include:

- Individuals responsible for implementing decisions should have a voice in influencing those decisions.
- Decisions should be made at the lowest possible level.
- Teachers can and should play an important role in influencing decisions that affect the children they teach.
- The process can help schools make the most effective use of limited resources to deal with the educational needs of the students they serve.
- All members of a school community can work together to achieve common goals.
- Parents and community members have an important role in shaping the education of the community's children.
- Change is most likely to be effective and lasting when those who carry out the changes feel a sense of ownership and responsibility for the process. Staff morale is enhanced, parents are often more enthusiastic and easier to involve, and students seem to work harder.

The Comer model is operationalized through a number of school level committees. The number of committees at each school varies depending on the specific needs and organizational preferences of the school. There are, however, eight basic functions for which each school team is responsible. These functions include curriculum, instruction, assessment, parent and staff skills development, school community

socialization and interaction, public relations, evaluation, and modification. Committee membership includes representatives of all key groups involved in the school including teachers, support staff, parents, and school level administrative staff. At the high school level students are represented on one or more committees. All committees function under the auspices of a School-Community Council which itself consists of representatives of all key groups involved in the school, including the school principal.

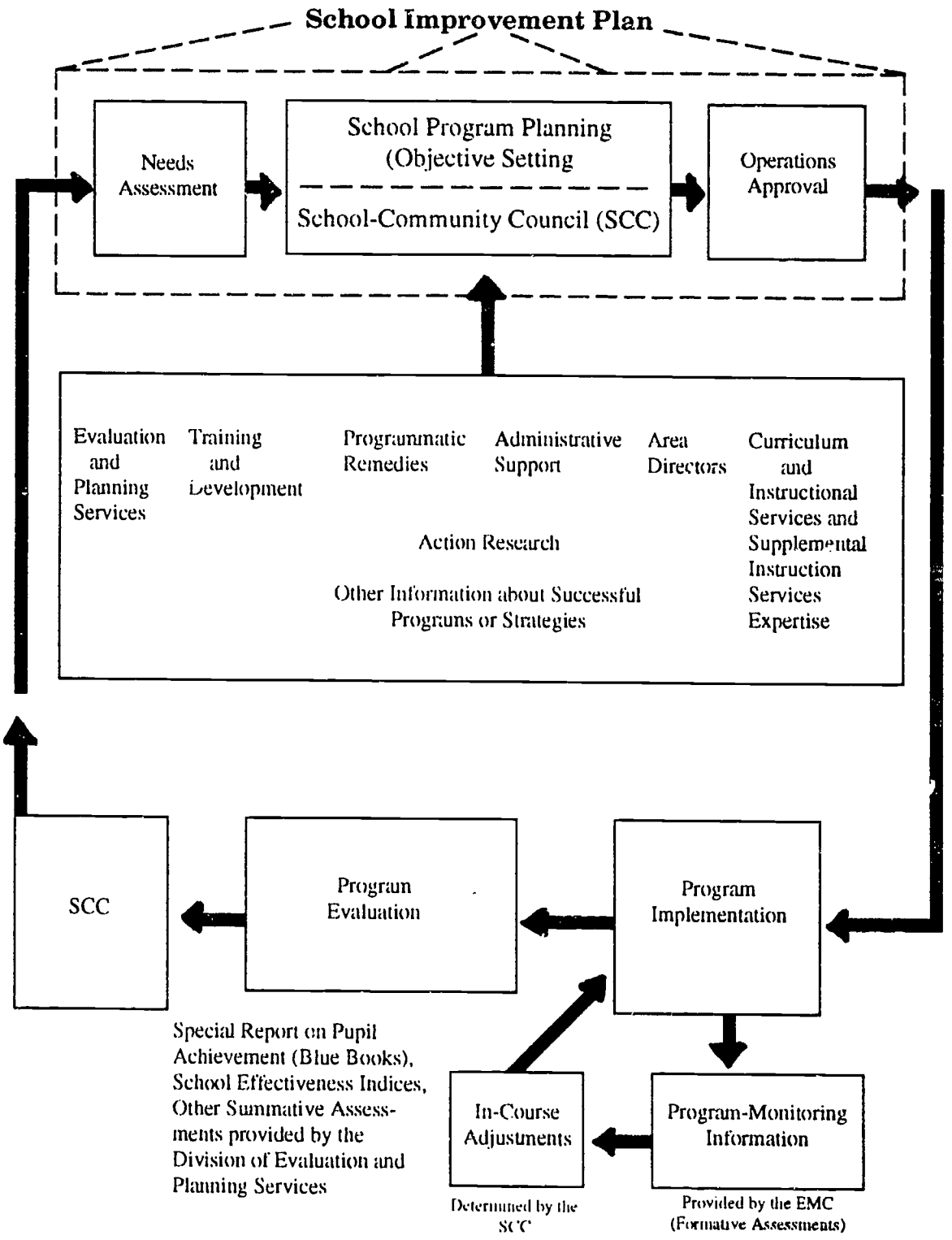
Regardless of the structure, the evaluation functions that are undertaken at the school level include the development of a School Improvement Plan; the interpretation of formative data for use in problem-solving, and of summative data for use in refocusing priorities, programs, and resources; the development of an implementation record of the various projects and programs within the school; and, the coordination of all school-based action research. Central Office research staff provide school personnel with training regarding how to accomplish many of the aforementioned tasks.

THE SCHOOL IMPROVEMENT PROCESS

Figure 1.0 provides a schematic depicting how the school improvement process functions within the parameters of site-based decision-making. Each school receives an annual needs assessment specifying school levels on important outcome variables. The important outcomes of instruction are determined through Districtwide assessments of all of the groups involved in the educational process. School program planning is implemented at the school level by the School Community Council. Planning focuses on 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 strategic plan, the School Improvement Plan.

Specifically, once the needs assessment has identified needs, school staff must prioritize those needs and focus on reducing the discrepancy between desired and existing

Figure 1.0 Schematic Depicting the School Improvement Process



Specifically, once the needs assessment has identified needs, school staff must prioritize those needs and focus on reducing the discrepancy between desired and existingoutcomes 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.

School-centered education does not assume 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 is ultimately 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 and the School Community Council. This 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 will probably 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 stems from an insufficient base of knowledge. Thus, educators are often in the position of having sufficient material resources but insufficient information resources. Once these decisions are made, the School Improvement Plan 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. This 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-Community Council informed of the extent to which program implementation conforms to specifications and, from an evaluation standpoint, guards against the evaluation of a fictitious event. 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 DISD evaluation personnel now must be implemented at the local level. This 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 over the past few years have been in the area of process evaluation.

Interim product evaluation provides periodic feedback to the School-Community Council relative to the attainment of specific sub-objectives 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 prortfolios of student work, performance testing, protocol analysis, and teacher-made tests; measures that are not available through systemwide data. Teacher Evaluation Consultants from each school are being trained in these techniques. [3] In cases where serious needs are

3 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 formative evaluation techniques to their campuses' school improvement plans. They will identify areas for school improvement, describe program activities, and periodically report information on program impact. This performance-based assessment thoroughly prepares consultants to design defensible evaluations, to measure program implementation, to identify appropriate instrumentation, to generate instrumentation similar to the *TAAS* through their microcomputers, to assess programs impact, and to compile and present reports for school improvement.

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 produced centrally 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). Expressed sequentially, action research requires 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 organizations' members in their respective work sites. In effect, members of the organization actively 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 implementation, usually one year, a series of summative product evaluation reports are prepared. These reports take the form of the Special Report on Pupil Achievement (REIS91-102), a school-level report that provides up to four years of disaggregated data on all relevant outcome and input variables and is used to determine whether or not schools met 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.

THE DISTRICT IMPROVEMENT PLAN

The District Improvement Plan (DIP) presents targets and corresponding strategic plans of action with a planning horizon of 1994-95. Since the District has a number of concerned audiences, the plan meets the accountability objectives and strategic planning requirements of the General Superintendent, the Board of Education, the Texas Education Agency, and the United States District Court. The DIP meets the four major requirements of a strategic planning system in that it receives input from all District departments and campuses, it sets accountability targets and minimum standards of performance for the District and each of its schools, it provides systemwide plans of action for meeting the major targets of the District, and, it specifies the methodology required for monitoring its implementation.

The DIP contains the strategic plans of each of the District's support divisions relative to their contributions to meeting each of the District's targets. It also contains the desired levels of District outcomes by 1995 and the intermediate steps necessary to get from 1991 levels to those desired outcomes. It is directly related to the SIPs in that outcome levels that are specified in each of the SIPs are those levels that will help the District reach its goals. The DIP sets the criterion level for desired outcomes. Goals are absolute. All schools could make them or no schools could make them, that is, target accomplishment is not determined by a norm group.

Targets

The DIP is organized around one systemwide enabling target. staff development, and ten outcome targets that focus directly on the District's priorities. Figure 2.0 shows the areas in which District targets exist for 1995.

Figure 2.0 DIP Target Areas, 1995

GOAL #	GOAL
1	Improve language arts skills (vocabulary, reading, oral competency, and writing skills
2	Improve mathematics problem-solving, concept, and computation skills
3	Increase parent/community involvement
4	Improve school climate and safety
5	Improve attendance (student and teacher)
6	Meet accreditation requirements/address citations
7	Increase promotion/course passing rate
8	Increase enrollment in advanced courses, diploma plans, and honors programs *
9	Increase graduation rate **
10	Increase college entrance test participation/performance **

* Applicable to middle and high schools only. ** Applicable to high schools only.

DIP Content

Each of the ten targeted outcomes has a strategic plan of action for meeting each target by 1994-95. The plans of action include the following elements:

- *Need* - a needs assessment summary describing the current status of the target.
- *Goal* - reference to the District's minimum accountability and accreditation's objectives or other standard of performance that will be met by implementing the plan.
- *Narrative of Strategy* - a summary of what will be done to address the target.
- *Waiver* - a specification of waivers required to implement the strategy.
- *Activities/Timelines/Divisions Responsible* - activities, corresponding timelines, and divisions responsible for meeting the District's targets.
- *Monitoring* - the methodology for directing, assessing, adjusting, and documenting formative activities to meet the goal.
- *Resource Implications* - a summary of the distribution (e.g. monies, personnel) changes required to implement the strategies. *(Implementation of the strategic plans of action is contingent upon meeting all resource*

requirements; in effect, the strategic plans of action will be null and void if any of their resource requirements are not met).

Authentic Assessment and Performance Testing

Schools are encouraged to use portfolios, protocol analysis, and other forms of authentic assessment in monitoring their programs. This information can then be used to provide evidence of accomplishment in instances where the more standard types of assessment fail to show progress. Performance testing is being built into the District's *Assessment of Course Performance (ACP)* test. The *ACP*s are final standardized examinations in 143 courses, grades 7-12. One hour is multiple choice while the other hour will be performance tests. These will be tests developed by the evaluation division and will have detailed scoring protocols. The performance portion of the tests will be scored by teachers with random scoring being done by the evaluation department. Performance tests will be developed across the next four years.

While it is not certain that the necessary reliability across scorers on the performance tests is attainable, it is important that the message be communicated to teachers that the kinds of skills and activities measured by performance tests are the kinds of skills and activities that the District wants them to teach their students. Early evidence on performance tests suggests that they are much more difficult than the average multiple choice tests (Dryden 1991). Figure 3.0 shows the formative and summative data currently available to the schools. Indicators that are collected centrally and provided to schools are specified with an "E". Formative indicators that should be part of a school's "action research" process are specified with a "C". State academic excellence indicators are asterisked while variables that are or will be outcome variables in the effectiveness indices are marked with a #.

Figure 3.0 Formative and Summative Indicators Available to DISD Schools

Indicators	Goal(s) Impacted	Date Available
* *TAAS* Results Disaggregated by Demographic Variables **	1, 2, 6	Fall (Grades 3, 7, 11)
Demographic Variable include Gender, Ethnicity, Free or Reduced Lunch and LEP (E)	1, 2, 6	Spring, Grades 4, 8, 10) #
ITBS Results Disaggregated by Demographic Variables (E)	1, 2, 6	Spring (Grades K-2) #
**NAPT* Results Disaggregated by Demographic Variables (E)	1, 2, 6	Spring (Grades 3-11) #
**ACP* Results Disaggregated by Teacher and Skills (E)	1, 2, 6, 7	Fall & Spring, (Grades 7-12)#
Reconstituted *TAAS, ITBS, NAPT,* Data (Class Lists and Skills Analyses) (E)	1, 2, 7	End of fourth week of school
Disaggregated Test Data by Program (Chapter 1, Reading Improvement, Bilingual, etc.) by School (E)	1, 2, 6	Fall
Portfolios of Student Work (C)	1, 2	Local Option
Performance Testing (C,E)	1, 2	Local Option/ACP #
Protocol Analysis (C)	1, 2	Local Option
Teacher-Made Tests (C)	1, 2	Local Determination
Teacher Satisfaction with Teaching, Ranking of Importance of Educational Goals, Perception of Teacher Influence, and Degree of Seriousness of Schoolwide Issues (E)	1, 2, 4, 5, 7, 9	Winter (all grades)
Student-to-Volunteer Ratio (E,C)	3	Fall
Volunteer Hours-to-Students (E,C)	3	Fall
Parental Involvement Log (C)	3	Local Option
Parent School Expectations, Perception of School Climate, Needs, Involvement/Participation (E)	3, 4	Winter (all grades)
* Student and Teacher Attendance (E,C)	1, 2, 5	Each six-week period #
Teacher Grade Distributions (E,C)	1, 2, 6, 7, 9	Each six-week period
School Effectiveness Indices (E)	1, 2, 5, 9, 10	September
School Effectiveness Indices Disaggregated by Student Group (E)	1, 2, 5, 9, 10	September
Student Satisfaction with Learning, Academic Self-Concept, Family Emphasis on Education, Cohesion	1, 2, 4	Winter (grades 4-12)
Teacher Climate Survey (E) (8 scales)	4	Provided on request by EPS

Figure 3.0 Cont.

Student Climate Survey, Grades 4-12 (E)	4	Provided on request by EPS
Principal Perceptions of Effectiveness of Training Services, Time on Task, School-wide Issues, Decentralization (E)	4	Winter (all grades)
Sociograms of Informal Interaction (lunch, recess, faculty meetings, etc.) (C)	4	Local Option
School-Community Council Survey (E)	4	Fall and Spring
Assistance and Consultation Team (ACT) Surveys (global issues, case management, training on mental health principles) (E)	4	Fall and Spring
Measures of Mobility and Stability (E)	5	Fall
Percent Eligible Tested versus Average Daily Attendance (E)	5	Fall
Monitoring of Local School Accreditation Remedies (C)	6	Fall
Monitoring of Implementation of Local School Programs (C)	7	Local Option
Monitoring of Instructional Delivery (C)	1, 2, 4, 6, 7	Local Determination
Student Retention Rate (E)	7	Fall #
* Student Enrollment in Advanced Courses (E,C)	8	Fall, Spring #
* Student Enrollment in Honors (E,C)	8	Fall, Spring #
* Student Enrollment in Diploma Plans (E,C)	8	Fall, Spring #
Survey of Student Course Interest (Grades 7-12) (E)	8, 9	Provided on request by EPS
* Dropout Rate (E)	9	December #
* Graduation Rate (E)	9	Fall #
* *SAT/ACT* Participation Rates (E,C)	10	Fall #
* *SAT/ACT* Scores (E)	10	Fall #
TASP Results (E)	10	Provided by the State
Graduate Follow-Up (E)	19	Fall #
Student Post-Graduate Pursuits (E)	8, 9, 10	Fall #
PSAT Participation Rates (E)	10	Fall #
PSAT Scores (E)	10	Fall #

* An Academic Excellence Indicator

** *TAAS* is the *Texas Assessment of Academic Skills*, a State-administered criterion-referenced test. *ITBS* is the *Iowa Tests of Basic Skills*. *NAPT* is the *Norm-referenced Assessment Program for Texas*, a Texas version of the *ITBS*. ACPs *are 143 criterion-referenced course exams, grades 7-12.*

Obviously, a great deal of training must occur if school staffs are to utilize available data and objectively collect and interpret additional data for aid in improving

their schools. Training modules for school staffs are currently being developed in keeping and scoring student portfolios of work, designing and scoring performance tests, conducting protocol analysis, developing teacher-made tests, interpreting and using data, and designing and conducting action research.

Accountability without information for diagnosis and improvement is of limited utility. In designing an accountability system, it is important to analyze data needs at each point in the organization. Data needs at the teacher level should be identified and those data aggregated upward and summarized to meet information demands at each successive level of the organization. It is essential that the system provide teachers with the information necessary to improve instruction. Without instructional improvement, accountability alone cannot improve a school system.

Figure 4.0 shows an example of the operationalization of the DIP targets. Each school receives its own data on each of these targets and is responsible for achieving its targeted outcomes. The targets are criterion-referenced in the sense that the schools have absolute goals and can concentrate resources on attempting to achieve those goals.

Figure 4.0 An Example Of An Accountability and Accreditation Profile *

OUTCOME VARIABLES					
GRADUATION RATE (5 Yr. %) (ACCRED. GRDTN RT)	52	54	57	59	61
% ADA - STUDENTS (CLI. ATTNDNC, ACCRED)	90.3	91.3	92.1	92.9	93.6
% ADA - TEACHERS (CLIMATE, ATTENDANCE)	96.5	97.5	98.0	98.0	98.0
% FROSH ADV GRAD PLANS (ACCRED. ENR ADV PLANS)	37	40	43	46	49
% SENIORS TAKING SAT/ACT (ACCRED. COL TSTS)	57	59	61	63	65
% SR > 700 SAT (16 ACT) (R, W, M, ACCRED, COL TSTS)	72	73	75	76	77
% SR > 1000 SAT (21 ACT) (R, W, M, ACCRED, COL TSTS)	27	31	34	37	41
% SR > 1300 SAT (27 ACT) (R, W, M, ACCRED, COL TSTS)	4	9	13	18	22
% GRADUATES CONT EDUC (PRMTN/GRDTN RT, COL TSTS)	48	51	53	55	58
% SRV LEP >= 40 R&L POST (R, W)	5	10	14	19	23

Figure 4.0 Cont.

% PASSING ALL COURSES (PRMTN RT)	53	55	58	60	62
% IN HONORS/AP/PRE-HNRS (ACCRED. ENR ADV PLANS)	22	26	30	33	36
DROPOUT RATE (%) (COMM. CLI. ACCRED. GRDTN RT)	7.0	6.5	6.0	5.5	5.0
% PASSING TAAS 09	20	28	35	42	48
% PASSING CURRICULUM REFERENCED TEST					
Biology	60.5	62.5	64.4	66.4	68.3
Chemistry I	65.0	66.5	68.0	69.5	71.0
Physics	70.9	71.9	72.9	73.9	74.9
U.S. History	63.6	65.2	66.9	68.5	70.2
Economics	68.1	69.3	70.5	71.7	72.9
Trigonometry	72.4	73.4	74.4	75.4	76.4
English I	69.1	70.2	71.3	72.4	73.5
NORM REFERENCED TESTS **					
TAP 10%≥25					
Reading	82	75	75	75	75
TAP 10 Median					
Reading	62	63	64	65	65
TAP 10%≥75					
Reading	33	25	25	25	25

* The profiles include many more variables. This figure is for illustratory purposes.

** Because the State changes the norm-referenced test every year, District goals in this area are to mirror the national norm group.

SCHOOL EFFECTIVENESS INDICES

The final tier of the accountability system is the most important from the standpoint of defining and rewarding outstanding schools. Inherent in the task of identifying outstanding schools are two complex issues:

- how to define effectiveness, and
- how to develop a model to assess effectiveness

In an attempt to provide a better definition of effectiveness and respond to the narrowly focused concern of earlier effective schools research, Murnane (1987), David (1987), and others have been proponents for developing an expanded number of outcome indicators. In addition, Oakes (1989), David (1987), and Cohen (1986) have argued the importance of incorporating input and process/context indicators as important aspects of better accountability mechanisms.

Possible input indicators often include school enrollment, socioeconomic/ethnic composition, proportion of limited-English-speaking children, enrollments in categorical programs, staff characteristics, and financial resources. Process indicators describe what is being taught, the way it is being taught, and include consensus on school goals, instructional leadership, opportunity to learn, school climate, staff development, and collegial interaction among teachers. Outcome indicators are usually related to capturing the results of school on students or providing information about other definitions of "good schooling", and may include student academic performance, teacher and student attendance rates, dropout and completion rates, performance of students at the next level of schooling, parent and student satisfaction, percent completing advanced courses, college attendance, and individual school goals (David, 1987; Oakes. 1989; Olson & Webster, 1990; Pollard, 1987; Shavelson, McDonald, Oakes, & Carey, 1987).

The Academic Excellence Indicator System

The Texas Education Agency, like many other State education agencies, has its own accountability system. This system is called the Academic Excellence Indicator System and includes the variables that are asterisked in Figure 3.0. It reports data on Districts in both a cross-sectional and cross-sectionally longitudinal manner and purportedly allows for the comparison of Districts to "like" Districts and to the State as a whole. This system has many flaws.

If one overlooks the flaws in basic measurement that are often present in State testing programs, flaws which extend all the way from unreliable tests to tests that are not scaled yet used to make quasilongitudinal comparisons, the technique of comparing schools based on unadjusted outcome measures usually adversely affects schools with population demographics that differ from the norm. This fact was graphically illustrated relative to ethnic background and *SAT* scores in a recent article by Richard Jaeger (1992). The non-statistical technique of comparing schools with similar characteristics is one

solution for cases involving a limited number of grouping characteristics, however, this approach has serious limitations when there is consistent one-directional variance on the grouping characteristics within group.

To illustrate this point, examine the group wherein the DISD was classified in the recent Academic Excellence Indicator Report published by the Texas Education Agency. The DISD was 15.9% White, the comparison group was 20.9% White. The DISD was 45.5% African-American, the comparison group was 38.8% African-American. The DISD was 66.5% poor, the comparison group was 55.5% poor. The DISD was 19.3% LEP, the comparison group was 17.4% LEP. Thus, on every important variable, the DISD had the group that performed most poorly on the *TAAS* statewide, independent of the District from which they were drawn. Is it any surprise that DISD scores were generally below those of the comparison group and the State? Yet, when those scores were adjusted for only the ethnic background of students, DISD performed at about State levels (Webster, 1991).

In short, for accountability purposes, the only fair and equitable method of comparison among and between schools or districts is one that statistically adjusts the outcome variables by the important inputs that relate to those outcomes but are not under the control of the schools. To incorporate a large number of input, process, and outcome variables in a fair and unbiased manner, an appropriate statistical method is multiple regression analysis (Barro, 1976; Felter and Carlson, 1985; Kirst, 1986; Klitgaard and Hall, 1973; MacKenzie, 1983; Saka, 1989). As a simplified illustration, the mean score for an outcome measure such as achievement is predicted after considering such input variables as gender, ethnicity, and socioeconomic level. The equation becomes more accurate if one or more estimates of previous achievement level are included. The difference between predicted and actual achievement, a residual or adjusted score, can then be interpreted as a comparison with other statistically similar schools, and as the school's own effect on achievement. It is important to note that a longitudinal data base is

necessary for these types of studies since cohorts must be used in the analyses. The characteristics of such a data base are detailed in Webster and Schumacher (1973).

The Anatomy of Effectiveness Indices

The school effectiveness methodology, as implemented in the DISD, 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, this 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 levels of performance based on these expectations, 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 equations by grade level for each outcome variable independent of school identification and then using those equations within schools to obtain mean gains over expectations. Relative weights are assigned to the outcomes by the Accountability Task Force. 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. To a great extent, the same targets that were used in the SIP and DIP processes were used as outcome variables in the school effectiveness indices. Thus schools work on improving target variables in an absolute sense through their SIPs and are judged in terms of a normative rank through the effectiveness indices.

School performance on the effectiveness indices is considered in terms of overall District patterns on the important outcome variables. If the District experiences a year of greatly increased achievement, individual school ranks on the effectiveness indices are not so important as long as improvement is shown. The emphasis of the methodology is

on the valid identification of effective schools, not on explaining their effectiveness through mathematical models such as path analysis or hierarchical linear modeling. Once effective schools are reliably and validly identified, detailed studies can be done of the process variables that contributed to their effectiveness.

The first step in developing the effectiveness methodology involved what educational practitioners have called "leveling the playing field". The Accountability Task Force was extremely concerned that all schools, regardless of the students that they served, had an opportunity to rank high on the effectiveness indices if they improved. Thus, the first step in developing the equations was to eliminate the variance in outcomes accounted for by ethnicity, gender, socioeconomic status, and limited English proficiency status. To accomplish this each outcome and predictor variable was regressed on the set of background variables and their interactions to produce a set of residuals for each of the predictor and outcome variables. (Webster, Mendro, and Almaguer, 1992).

A stepwise regression approach was then used on the residuals of both the outcome and predictor variables. Equations were developed utilizing individual students rather than school means. Satisfactory prediction was achieved in all cases without having to go back more than one year. This maintained the degrees of freedom associated with the equations. A previous model that was utilized by the District in 1984 used a variant of time-series analysis, but since this model required at least three years of historical data, it suffered from severe subject mortality due to a high student mobility rate (Webster and Olson, 1984). For 1992-93, simulations are being run adding mobility and an overcrowdedness statistic to the fairness variables and using all possible dichotomous and continuous predictor variables in the equations. This approach, which eliminates the need to assume parallel regression lines, will be used if the problem of multicolinearity does not present itself (Aiken and West, 1991; Webster, Mendro, and Almaguer, 1993).

Once each student's standardized residual values on each variable were computed, the next step involved determining mean residuals for individual campuses. These mean residuals were then standardized within each distribution. To account for differences in school size, the means for each school were then standardized by computing the standard deviation of the mean for each school mean. Outcome variables were then weighted based on a weighting schema determined by the Accountability Task Force. Figure 5.0 shows the 1991-92 weighting system devised by the Task Force. The weighting schema emphasized norm-referenced achievement at the elementary grades and placed increasing emphasis on curriculum-referenced tests for measures of middle and high school achievement.

Finally, the Accountability Task Force assigned weights for each outcome variable were multiplied by the standardized mean residual for each school and summed across variables to obtain a final weighted mean for each school. The school means were then ranked to produce an ordering of effective schools.

Figure 5.0
Weighting of Criterion Measures by the Accountability Task Force*
1991-92

Elementary Criteria		Middle School Criteria		High School Criteria	
Criterion	Weight	Criterion	Weight	Criterion	Weight
NAPT-Vocabulary	2/Grade	*ACP*	4/Grade	*ACP*	6/Grade
NAPT-Reading	3/Grade	*NAPT*-Vocabulary	2/Grade	*NAPT*-Reading	3/Grade
NAPT-Language	2/Grade	*NAPT*-Reading	4/Grade	*NAPT*-Language	3/Grade
NAPT-Mathematics	2/Grade	*NAPT*-Language	2/Grade	*NAPT*-Mathematics	3/Grade
Attendance	1/Grade	*NAPT*-Mathematics	3/Grade	*SAT/ACT*	4/School
Promotion Rate	1/School	Attendance	1/Grade	*SAT/ACT*	4/School
				Graduation Rate	4/School

* The weight columns show weight in the first part and whether the weight is applied to the variable at each grade or once for the entire school in the second part.

For 1992-93 and future years, the number and nature of outcome variables have been expanded. Figure 6 shows the weights and nature of variables for 1992-93. For 1993-94, the *ACP*s will include performance tests. When the results of performance tests are utilized in the effectiveness indices, the results of all tests will be weighted by their reliability. Also for 1993-94, dropout rate, student enrollment in accelerated courses and related *ACP* scores, high school student enrollment in Advanced Diploma Plans, post-graduate enrollment in college or business schools, percent tested on the *Preliminary Scholastic Aptitude Test (PSAT)* and student scores on same, and post-graduate career pursuits will be added to the outcome variables (Webster, Mendro, and Almaguer, 1993).

Dallas Independent School District schools and their staffs were eligible for cash awards for 1991-92 performance based on the school effectiveness methodology under the District's School Performance Improvement Awards Program. In September of 1992, 2.4 million dollars was distributed to effective schools and their employees. Half of the 2.4 million dollars was budgeted by the District, the other half came from the community. To qualify schools had to exceed prediction on the effectiveness indices, test 95% of their eligible students, and outgain the national norm group in at least fifty percent of their cohorts. Once a school was selected as an award winner, the school received $2000 for its activity fund, each member of its professional staff received $1000, and each member of its support staff received $500. This program in continuing in 1992-93.

Figure 6.0
Weighting of Criterion Measures by the Accountability Task Force*
1992-93

Grade	1	2	3	4	5	6	7	8	9	10	11	12
ITBS												
Reading	3	3	•	•	•	•	•	•	•	•	•	•
Language	2	2	•	•	•	•	•	•	•	•	•	•
Vocabulary	2	2	•	•	•	•	•	•	•	•	•	•
Math	2	2	•	•	•	•	•	•	•	•	•	•
NAPT												
Reading	•	•	3	3	3	3	4	4	3	3	3	•

Figure 6.0, Continued

Language	•	•	2	2	2	2	2	2	3	3	3	•
Math	•	•	2	2	2	2	3	3	3	3	3	•
Promotion	1 per school						1		•	•	•	•
Attendance	1	1	1	1	1	1	1	1	1	1	1	1
TAAS												
Reading	•	•	•	3	•	•	•	3	•	4	•	•
Writing	•	•	•	3	•	•	•	3	•	4	•	•
Math	•	•	•	2	•	•	•	2	•	4	•	•
LEP TEST	3	3	3	3	3	3	•	•	•	•	•	•
ACP												
Language Arts	•	•	•	•	•	•	2	2	2	2	2	2
Math	•	•	•	•	•	•	2	2	2	2	2	2
Social Studies	•	•	•	•	•	•	2	2	2	2	2	2
Science	•	•	•	•	•	•	2	2	2	2	2	2
ESOL	•	•	•	•	•	•	2	2	2	2	2	2
Reading	•	•	•	•	•	•	2	2	2	•	•	•
World Language	•	•	•	•	•	•	•	•	1	1	1	1
Graduation Rate	•	•	•	•	•	•	•	•	5 per school			
SAT/ACT % Tested	•	•	•	•	•	•	•	•	•	•	5	
SAT/ACT Score	•	•	•	•	•	•	•	•	•	•	4	

Teacher Effectiveness Indices

Since the teacher is the principal deliverer of instruction to students, it is essential that a method for attributing student outcomes to teachers be developed. Of the numerous methods available for teacher evaluation, student outcome data provide an attractive option for evaluating teacher effectiveness as they are basically objective in nature. One only needs to examine the distributions of teacher evaluations and student achievement in the average urban District to quickly realize that something is wrong with the current system. A number of researchers have enumerated many factors that inhibit the reliable and valid use of student achievement data as an evaluation of teacher performance (Bentz, 1988; Dutweiler and Ramos-Camel, 1986; Grobe, 1992; Haertel, 1986; Koehler, 1985; Murnane and Cohen, 1986; Redfield, 1987).

Some of the most troublesome of these factors include:

- Standard measurement instruments are not available for many courses and subject areas. (In Dallas, about 57% of high school course sections and 70% of middle school course sections have *ACP*s.)
- Reliable performance measures are non-existent.
- Because of team teaching, pull-out and send-in programs, and other special programs it is difficult to isolate an individual teacher's effect on individual students. (In Dallas, supplementary teachers account for 30% of the teaching force).
- What the student brings to the classroom in terms of ability, home and peer influence, motivation, etc., is very powerful in affecting school outcomes. Attempting to adjust student outcomes based on student inputs at the teacher level creates serious degrees of freedom problems, that is, the resulting estimates are not stable.

The preferred solution to this dilemma is to provide the principal, as the instructional leader and Chief Executive Officer of the school, with a large and diverse set of explanatory variable information to help him or her with the evaluation of teachers. Information includes class improvements on outcome measures over the previous year, class item and skills analyses, class background information, and allows for teacher-generated information such as protocol analysis. Also provided are standards which communicate progress being made with similar students across the District. The emphasis is currently on diagnosis and improvement although the information will eventually be used as part of the teacher evaluation system. There are currently no plans for rewarding individual teachers based on some type of effectiveness index since one of the major strengths of the school effectiveness methodology is the staff collegiality that it reinforces. Such collegiality is important in restructuring around Comer's school-centered education model.

SUMMARY

This paper has described a three tier accountability system. District goals and desired outcomes are established through a Districtwide planning process and operationalized through the District Improvement Plan. Each school's role in helping the District to meet its goals is determined through a School Community Council which ensures involvement at the local campus level. Accountability is operationalized in a criterion-referenced manner through an analysis of absolute outcomes relative to school and District performance on goals specified in the District Improvement Plan and the School Improvement Plans, and in a norm-referenced manner through school effectiveness indices. Schools and their staffs are eligible for financial awards based on school performance on the effectiveness indices.

Besides providing an objective procedure for identifying effective schools, the program has a number of practical advantages. First, and most important, it is designed to foster teamwork among school staffs within schools. In order to achieve the necessary improvements in student outcomes, school staffs must work together in a coordinated effort. The program does not reward individual competition among teachers within schools.

Second, the program focuses attention on the important outcomes of schooling. The Accountability Task Force, as well as many other groups associated with the schools, are discussing what it is that the schools are about. The process of weighting the outcome variables, a procedure that is done annually, gives many divergent groups the opportunity to share their views relative to the purposes and importance of schooling. While the accountability system alone will not improve instruction, the curriculum and instructional delivery processes that must be changed to impact the defined outcomes will.

Third, the procedures described afford all schools an opportunity to be distinguished in the awards independently of their student populations status on the achievement continuum. The emphasis is on effectiveness with the students who come in

the door, not absolute outcome levels. The techniques reward those schools that impact the most students the most positively (Webster, Mendro, and Almaguer, 1993).

Many District and State accountability systems include District and School Improvement Plans that encompass absolute goals. The addition of effectiveness indices make the accountability system valid and fair. Among the advantages of this type of approach are that each school's performance is not judged by simple examination of raw outcome variables, but instead by comparing its student outcome levels with empirically determined expectations based on individual student histories; that schools derive no particular advantages by starting with high-scoring or low scoring students of any particular ethnic or economic group; that schools are only held accountable for the outcome levels of continuously enrolled students, that is students who have been exposed to their instructional program; that adequate time for test make-up is allowed and schools must test 95% of their eligible students; and, a Task Force representing all of the important groups that have a stake in schooling determines the important outcomes of schooling and their respective weights in the equations.

References

Aiken, L. S. and West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions.* Newburg Park: Sage

American, 2000: An Education Strategy (1991). *Sourcebook.* Washington, D. C: U. S. Department of Education

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1985). *Standards for Educational and Psychological Testing*. Washington, D. C: American Psychological Association

Bano, S. M. (1985). *The Logic of Teacher Incentives.* Washington, D.C: National Association of State Boards of Education.

Berk, R. A. (1984). The Use of Student Achievement Test Scores As Criteria for Allocation of Teacher Merit Pay. *National Conference on Merit Pay For Teachers.* Sarasota, March, 1984.

Berk, R. A. (1988). *Fifty Reasons Why Student Achievement Gain Does Not Mean Teacher Effectiveness.* Journal of Personnel Evaluation in Education, 1, 345-363.

Cohen, M. (1986). Designing State Education Assessment Systems. *Study Group on the National Study of Student Achievement.*

Comer, J. P. (1988). Educating Poor Minority Children. *Scientific American*, 259, 5, 42-48.

Commission for Educational Excellence (1991). *Final Report.* Dallas: DISD, June, 1991.

David, J. (1987). Improving Education with Locally Developed Indicators. *Center for Policy Research in Education*, New Brunswick: Rutgers University.

Dryden, M. (1991). *Evaluation of the 1990-91 South and West Dallas Learning Centers.* Dallas, Texas: Division of Evaluation and Planning Services, DISD.

Duttweiler, P. C. and Ramos-Cancel, M. L. (1986). *Perspectives on Performance-Based Incentive Plans*. Austin, Texas: Southwest Educational Laboratory (ERIC ED272511).

Edwards, M. E. (1991). *School-centered Education: A Plan for Creating a System of Schools in the DISD*. Dallas, Texas: DISD

Edwards, M. E. (1993). *School-centered Education: The Dallas Model.* Dallas, Texas: DISD.

Felter, M. (1989). A Method for the Construction of Differentiated School Norms. *American Educational Research Association.* San Francisco, California (ERIC ED312302).

Felter, M. and Carlson, D. (1985). Identification of Exemplary Schools on a Large Scale. In Austin and Gerber (eds.), *Research on Exemplary Schools.* New York: Academic Press, 83-96.

Grobe, R. P. (1992). *Using Student -Achievement Data for Teacher Evaluation.* Dallas, Texas: Division of Evaluation and Planning Services, DISD.

Haertel, E. (1986). The Valid Use of Student Performance Measures for Teacher Evaluation. *Education Evaluation and Policy Analysis,* 8, 45-60.

Haertel, E. (1990). Performance Tests, Simulations, and Other Methods, in J. Millman and L. Darling-Hammond (eds.), *The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers.* Newbury Park, Ca.: Sage, 278-294.

Institutional Research Office (1992). *Special Reports on Pupil Achievement*, REIS92-102, 1-200, Dallas, Texas: DISD.

Jaeger, R. M. (1992). Weak Measurement Serving Presumptive Policy. *Kappan*, 74, 2, 118-128.

Kirst, M. (1986). New Directions for State Education Data Systems. *Education and Urban Society*, 18, 2, 343-357.

Klitgaard, R. E. and Hall, G. R. (1973). *A Statistical Search for Unusually Effective Schools*. Santa Monica, Ca.: Rand Corporation.

Koehler, V. (1985). National Overview: Career Ladder and Merit-Pay Plans, in C. Clark (ed.), *Proceedings of the Preventive Law Institute on Career Ladder/Merit Pay.* Austin, TX.: Southwest Educational Development Laboratory (ERIC ED272511).

Lewin, K. (1946). Action Research and Minority Problems. *Journal of Social Issues*, 2, 58-73.

MacKenzie, D. (1983). School Effectiveness Research: A Synthesis and Assessment. In P. Duttweiler (ed.), Educational Productivity and School Effectiveness. Austin, Texas: Southwest Educational Development t Laboratory.

Murnane, R. S. (1991). The Case for Performance-based Licensing. *Kappan*, 73, 2, 137-142.

Murnane, R. S. and Cohen, D. K. (1986). Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and Few Survive. *Harvard Educational Review*, 56-1-17.

Nicoll, R. (1989). *School Accountability in the Mt. Diablo Unified School District,* American Educational Research Association, San Francisco, California.

Oakes, J. (1989). What Education Indicators? The Case for Assessing the School Context. *Educational Evaluation and Policy Analysis*, 11, 181-199.

Olson, G. H. and Webster, W. J. (1986). *Measuring School Effectiveness: A Three-year Study.* American Educational Research Association, San Francisco, California.

Pollard, J. (1987). *Viewpoints from Selected States on Accreditation and Accountability.* Austin, Texas: Southwest Educational Development Laboratory.

Redfield, D. (1987). *Expected Student Achievement as a Potential Factor for Assessing Teacher Effectiveness.* ERIC ED 290764.

Saka, T. (1989). *Indicators of School Effectiveness: Which Are the Most Valid and What Impacts Upon Them?* American Educational Research Association, San Francisco, California (ERIC ED306277).

Shavelson, R. J., McDonnell, L. M., Oakes, J., and Carey, W. (1987). *Indicator Systems for Monitoring Mathematics and Science Education.* Santa Monica, California: Rand Corporation.

Southern Regional Education Board (1986). Incentive Programs for Teachers and Administrators: How are they Doing? *Career Ladder Clearinghouse.* Atlanta, Georgia: Southern Regional Education Board.

Stufflebeam, D. L., et.al. *Educational Evaluation and Decision-making.* Itasca, Illinois: Peacock, 1971.

The Joint Committee on Standards for Education Evaluation (1981). *Standards for the Evaluation of Educational Programs, Projects, and Materials.* New York: McGraw-Hill.

Town, S. W. (1973). Action Research and Social Policy. *Sociological Review.* 12, 4, 128-137.

Webster, W. J. (1991). An Analysis of Available Student Achievement Data in the Dallas Independent School District, *Executive Summaries of Evaluation Reports,* DISD, 1-44.

Webster, W. J., Mendro, R. L., and Almaguer, T. (1992). *Measuring the Effects of Schooling: Expanded School Effectiveness Indices.* American Educational Research Association, San Francisco, California.

Webster, W. J., Mendro, R. L., and Almaguer, T. (1993). *Effectiveness Indices: The Major Component of an Equitable Accountability System.* American Educational Research Association, Atlanta, Georgia.

Webster, W. J. and Olson, G. H. (1988). *A Quantitative Procedure for the Identification of Effective Schools.* Journal of Experimental Education, 56, 213-219.

Webster, W. J. and Schumacher, C. C. (1973). A Unified Strategy For Systemwide Research And Evaluation. *Educational Technology*, 13, 5, 68-72.