

**EVALUATION FOR SCHOOL IMPROVEMENT:
TRY-OUT OF A COMPREHENSIVE
SCHOOL-BASED MODEL**

Joan L. Herman

CSE Report No. 274

Center for the Study of Evaluation
Graduate School of Education
University of California, Los Angeles

1987

The project presented, or reported herein, was performed pursuant to a Grant from the Office of Educational Research and Improvement/Department of Education (OERI/ED). However, the opinions expressed herein do not necessarily reflect the position or policy of the OERI/ED and no official endorsement by the OERI/ED should be inferred.

EVALUATION FOR SCHOOL IMPROVEMENT:
TRY-OUT OF A COMPREHENSIVE SCHOOL-BASED MODEL

Overview

"How well are we doing?" "How can we make things better?" School boards, administrators and educators constantly ask such questions. They are, however, difficult to answer. While districts often collect a great deal of data as part of their routine evaluation activities, such data is often poorly suited to illuminate these basic issues. Collected in the names of sound management and rational decisionmaking, the data instead often sits unused on bookshelves, in thick computer printouts, and in inaccessible computer files, with little or no significant impact on the process of education in districts, schools, or classrooms.

CRESST's Multilevel Evaluation Systems project seeks a more useful approach to evaluation by developing and implementing Baker's "top-down, bottom-up" evaluation model (Baker, 1983). The model calls for context sensitive information for principals and teachers to help them improve their instructional programs and policy sensitive information for superintendents, board members, and other administrators to inform their program planning and evaluation. More specifically, the project has the following objectives:

1. To develop and implement a model multipurpose evaluation system designed to facilitate educational decisionmaking and to support school improvement and renewal;
2. To develop and implement a core data base, drawing on a broad variety of quality indicators, that can serve the diverse decisionmaking needs of teachers, administrators and district policymakers;
3. To develop and implement a data management system that will provide student level, class level, grade level, school district, and inter-district summaries across selected measures included in the data base;
4. To extend our understanding of the production and use of information and its impact on educational innovation.

The project model draws on accumulated knowledge about what makes school effective, about what makes evaluative information useful to teachers and administrators; about what makes an information system useful in organizations; and on the power of currently available, low cost

microcomputer technology. In the sections below, the rationale underlying the project model is summarized briefly followed by a description of its implementation in five school districts. We end by considering the implications of the pilot project for the future design of school-based information systems.

Background

The model is premised on the assumptions that evaluation can be a valuable tool for improving schools, and that the collection, analysis, and distribution of information can stimulate and inform action to upgrade the quality of education. It assumes that evaluation information can have such an affect by facilitating better educational decisionmaking, improved instructional planning and more effective school management at all levels of the educational hierarchy. District and school administrators, for example, can use valid information about student achievement, among other indicators, to make judgments about their schools' performance, to evaluate the effectiveness of particular programs, to establish grade, school, or district wide priorities, to allocate resources wisely, and to spot curricular or other problems needing correction. Using information about student test performance, attitudes, preferences, etc. in combination with their own perceptions, teachers might more easily and effectively accomplish critical tasks such as assigning students to groups, diagnosing individual learning problems, monitoring student progress, assessing subject matter mastery, identifying students who need remediation or enrichment activities. Teachers and the principal working together could use information about school context, instructional processes and outcomes to analyze local problems and improve the effectiveness of their school programs. School board members and district leaders could likewise use such information to get a comprehensive, accurate picture of the quality of their schools and to target their improvement efforts accordingly.

But while evaluation information has this potential power, its impact has been quite modest (Alkin et al., 1979; Cohen & Garet, 1975; Patton, 1986). Why the discrepancy? The reasons appear to be many and varied. The source and nature of formal evaluation practice over the last two decades appears to be a major limiting factor. Much of this practice has led to the proliferation of standardized tests devoted to supplying the needs of legislators and administrators at the federal, state and local levels who wished to know how mandated programs were working and how schools were achieving. The people at the bottom -- teachers and local administrators -- have been seen as data providers rather than data users, as implementers of reform efforts rather than initiators of such efforts.

Teachers and local school administrators meanwhile have questioned the validity of these "top-down" evaluation efforts, arguing that required tests do not reflect what they are teaching and that some are inappropriate for particular groups of students (Herman & Dorr-Bremme, 1983). They claim further that the paperwork and bureaucratic burdens associated with mandated evaluation requirements intruded into, rather than supported, their own planning and improvement efforts. They have argued also that improvement of educational quality must be directed at local school sites where teachers and administrators directly interact with children. "Bottom-up" needs, in short, are not being well served by mandated evaluation and testing programs.

Complementing these concerns were criticisms by some in the research community who also have questioned the value of standardized tests (Baker, 1983; Eisner, 1985; Sirotnik & Burstein, 1984). Criticized as providing a very limited view of educational quality, these tests, for the most part, examine student performance on only a narrow slice of the curriculum, emphasizing basic skills and giving little attention to learning in the content areas, higher-order reasoning skills, and the multiplicity of other academic, social, and vocational goals which schools are supposed to address.

Using "test scores only" to capture educational quality suffers from other validity problems as well. While the "How well are we doing" question provides impetus for much evaluation activity, answers framed solely in terms of test scores sometimes mask as much as they clarify. Contrary to what may be the belief of many policy makers, it is not possible to backward chain from a single test score to inferences about the overall quality of education in a state or district or at a particular school. Student test scores are the result of many factors, school quality among them. Cultural, social, economic, demographic and motivation factors also are clearly influential, but often ignored in giving districts or schools report cards. Inequities and invalidities result, crediting schools which serve advantaged populations and disadvantaging schools serving minority and poor students.

But even if credible testing instruments were available, more broadly-based tests administered, and the results integrated within a social/economic/community context, a serious deficiency in many previous evaluation conceptualizations would remain. Evaluation in support of school improvement at the local level should not be limited to the type of data typically collected: outcome data. Left undocumented by evaluations focusing only on outcomes are the processes and context features which create or contribute to those outcomes. Understanding these is

critical to directing an effective agenda for school improvement. Eisner speaks to this point:

"If we want to understand why we get what we get from our schools we need to pay attention not simply to the scores, but to the ways in which the game is played."

School context and process, thus, have not been used sufficiently as a source of explanatory hypotheses in routine evaluation practice (Sirotnik, 1984). They also have been neglected but as important intervening factors which influence how evaluation data themselves are interpreted and how they are used for school improvement and change (Sirotnik et al., 1985; Dorr-Bremme, 1984). Having technically sound, comprehensive data available does not assure that anyone will look at them, analyze them, discuss them, or take action stimulated by them. A growing literature on factors which influence evaluation utilization (Alkin et al., 1979, 1985; Bank & Williams, 1985), on factors which contribute to change and innovation in schools (Berman & McLaughlin, 1977; Sarason, 1982; Heckman et al., 1983) and on factors that affect the implementation of evaluation and information systems in fields outside of education (Lucas, 1975; Markus, 1981; Multinovich & Vlahovich, 1984) provides clues on knowledge utilization -- factors such as leadership support, ownership, perceived relevance, fit with routine practice, incentives, etc. which can be expected to influence whether evaluation information is acted upon and used to alter existing practices.

There are many reasons, in short, why evaluation has had only peripheral influence on teachers, principals and district personnel in their efforts to improve schools. To summarize: evaluation has been primarily linked with "top-down", "highly centralized improvement approaches which have not been sensitive to "bottom-up" needs; evaluation data has been based primarily on a narrow range of outcomes; evaluation often has ignored critical variables in the context and process of schooling; evaluations have not sufficiently considered the factors which would facilitate attention to findings and translation of findings into action.

Recognizing these limitations, some school districts are currently developing innovative evaluation systems that serve multiple users and their diverse information needs (Bank & Williams, 1984, 1985; Idstein, 1985; Dussault, 1985). Radical changes in evaluation methodology are emerging reflecting both the reality of our decentralized or "loosely coupled" educational system and the awesome power of computers.

Education comes down to what happens to students in classrooms and schools. Educational quality comes down to critical interactions between teachers and their students, behind the classroom door. Further, more so than in the past, schools and classrooms today encompass tremendous diversity in student population, in teacher skills, in curricular goals, in teaching strategies. Because of this diversity and because of the locus of instructional control, the school building, rather than more remote and larger administrative units, is the appropriate unit for solving many educational problems (Goodlad, 1983; Baker, 1983). Consequently, school personnel are among the appropriate designers and beneficiaries of improvement-oriented evaluation systems.

Individual schools however, often do not have sufficient resources, expertise, control, etc. to solve all their educational problems by themselves. Solutions frequently require initiative, direction, resources, and/or actions at higher administrative levels, levels which have legal responsibilities for governance, personnel, resource allocation, and policy formation, among other things. These realities suggest the desirability of a system of evaluation which could provide local schools with a rich, locally sensitive information base to aid their problem-solving and which could also provide appropriate aggregate information for decisionmaking at higher levels of the system. One potentially promising approach is a distributed information system which gives actors at various levels immediate access to a shared core of data but enables them to supplement and analyze it in response to their specific decision needs. The relatively low cost availability of powerful microcomputer technology makes possible a number of intriguing options for local site processing, data networking and a variety of lateral and horizontal linkages.

An Improved Model

The limitations in current practice, the reality of the loosely coupled (Weick, 1978) educational system, and the availability of new technology all support the need for a new, top-down - bottom-up model of evaluation, one which provides quality data to aid the decisionmaking of policymakers and local school practitioners and one which provides a productive tool for improving the quality of schools. The background discussion and the problems it articulates foreshadow a number of features deemed critical for a such a valid and useful evaluation system:

1. it makes relevant information easily available to teachers, school administrators, and district and state policymakers to aid their decisionmaking;

2. it includes information on a range of school outcomes;
3. it includes information on school context and student characteristics to contextualize outcome and effectiveness analyses;
4. it includes information on school and instructional processes to elucidate and analyze local problems and accomplishment;
5. it links outcome information with instructional process and school context data to provide explanatory power for findings;
6. it enables efficient sharing of information within and across levels of the educational hierarchy, minimizing redundant, overlapping testing and evaluation requirements;
7. it includes externally fixed elements to assure sensitivity to information needs at the district and state levels and variable, locally selected elements and measures of interest to school professionals;
8. it encourages data collection, analysis, and use over time;
9. it builds on organizational and management strategies to facilitate system use including such things as:
 - locating responsibility for defining the system dually at the school and district levels
 - facilitating ownership and flexibility for local school uses
 - assuring leadership support at the district and school levels
 - attending to specific information and reporting needs to all groups
 - making the system user-friendly and easily accessible.

Attending to these critical features, the project model utilizes a comprehensive information base about student characteristics, school context, school and instructional process and a range of outcomes that can be analyzed, arrayed, and appropriately reported at various levels to facilitate decisionmaking at the classroom, school, district, and perhaps state levels and to satisfy

reporting requirements for special programs. (Figure 1 displays an overview of the model system.) The multilevel character of the system enables essentially the same set of data to be analyzed from the decision perspectives of a variety of users -- district policymakers looking at the performance and quality of the district as a whole, principals and teachers assessing effectiveness at the school level; teachers examining class-level outcomes and individual student strengths and weakness. While providing a common core data base available to all users, the model also reserves a place for interests and concerns unique to each unit. The information base at each level, in short, is comprised of fixed data elements (i.e., common for all sites) and variable elements (at the discretion of individual schools, etc.). Critical to the model is that its constituent data elements are collaboratively defined by intended user groups and its implementation managed to encourage ownership and promote use; further, to facilitate information use where education actually occurs, the system is school-based.

The next section describes a field test of this model in collaboration with five school districts in the Eastern United States.

Technical Approach

An important element in the technical approach was the organizational structure through which the project was to operate. The five participating school districts were a part of the University of Pennsylvania's School Council. The project was initiated at the request of the district superintendents and became a designated project of the Council. The Council's executive director served as project director responsible for facilitating and coordinating planning and implementation. Steering committees were constituted within each district to assure local representation and input into project planning and to locate responsibility for implementation within each district. Each steering committee included teacher, principal, and district administrator representatives as well as the district superintendent; superintendents were encouraged to designate one member as project coordinator for their district. The Center for Study of Evaluation at the University of California, Los Angeles was responsible for the original project conceptualization and for providing technical assistance in identifying data, instrumentation and analysis needs and for providing student, classroom, school, and district level data reports. The initial plan was to include two schools from each of the participating districts and two fourth and fifth grade classrooms at each participating school.

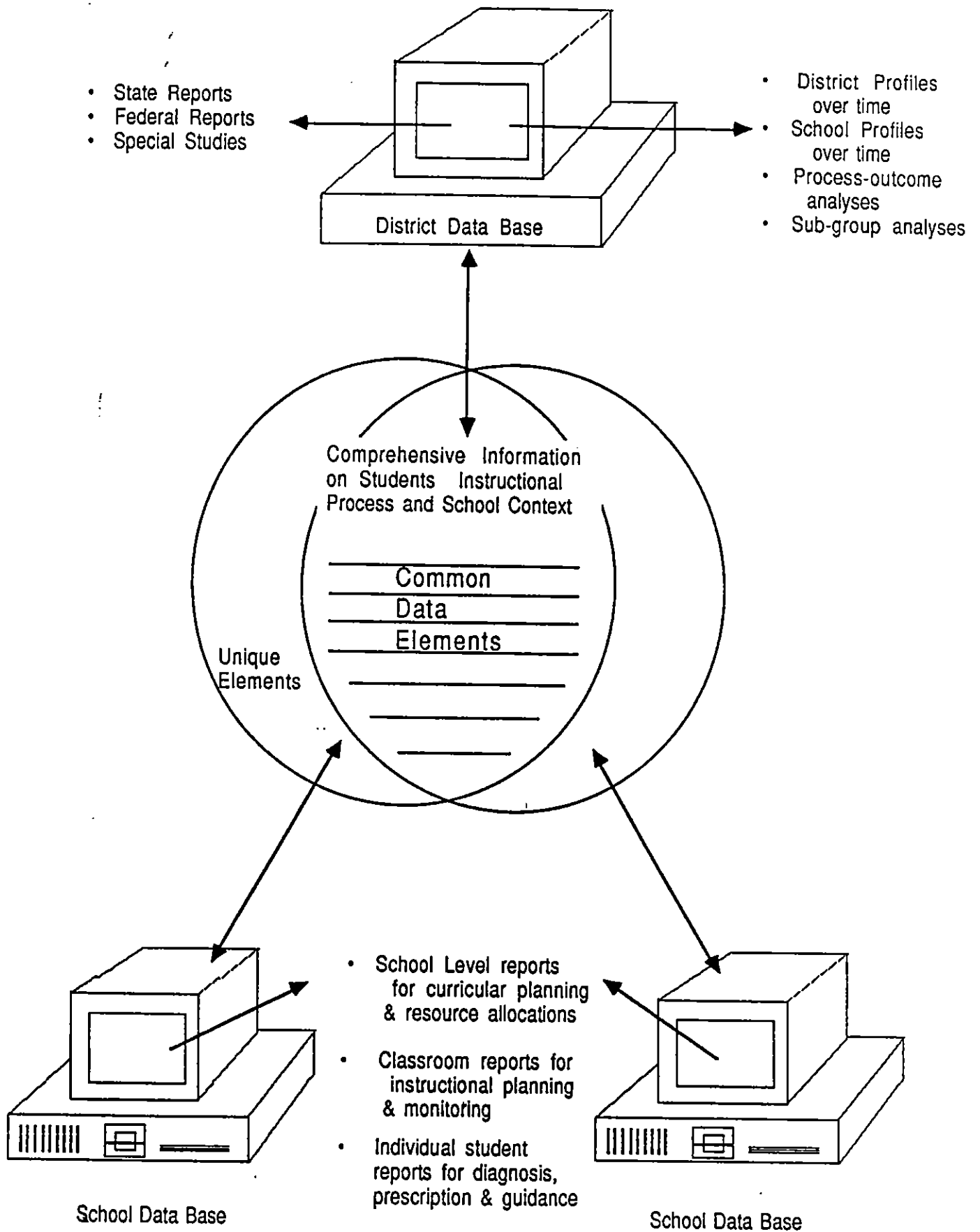


Figure 1

The technical approach proceeded in four general steps:

1. Deciding what needs the evaluation system should serve and the data that should be included within the core data base.
2. Determining data collection procedures.
3. Collection of data.
4. Determining data analysis and reporting needs.

Decisions in each of these areas were to guide the development and implementation of a user-friendly, microcomputer-based data management system to provide useful reports to teachers, principals, district administrators, superintendents and board members. (To enhance initial reporting flexibility and to avoid potentially costly reprogramming efforts, initial analysis were done on UCLA's mainframe computer.)

Essentially parallel processes were used to accomplish each of the above steps. Working meetings including participants from all five districts were convened to consider each decision area, to determine common priorities from among a range of given options, and to review progress and proposed products. Follow-up meetings in each individual district were used to verify consensus, to identify unique concerns and unique data requirements for each school/district, and to review instrumentation and reports. Data collection proceeded in two fourth grade and two fifth grade classrooms in each participating school; data collection included a combination of rostering archival data and administering specially developed student and teacher questionnaires. The results section below describes how the model was operationalized in the five districts, including the questions the information system was designed to elucidate, the types of indicators considered relevant, and the types of analyses and reports deemed useful.

Results

What needs and concerns should the evaluation system meet? While there was considerable diversity in the types of concerns expressed, several common questions emerged across the working groups. Their questions concerned the outcomes of schooling for students, the nature and effectiveness of the educational process, and the influence of the context in which instruction occurs. More specifically, their questions included:

Student Outcomes

- o How much growth do students show over time?
- o How does student performance compare to that of similar students in other districts?

Process

- o Are resources effectively allocated and used?
- o What instructional practices contribute to quality education?
- o Are educational programs challenging and appropriate in their levels of expectation for students?

Context

- o Can school climate contribute to quality student performance?
- o What's the role of student background in their performance?

Concerns unique to each district focused on academic performance in specific subject matter areas, the effectiveness of particular instructional practices, the special needs of students from particular backgrounds, and the influence of contextual features specific to the district.

What indicators might help illuminate these questions? Starting with an initial pool of potential indicators identified on the basis of the literature, a core list of priorities was identified for student outcomes, instructional process, school context, and student demographic characteristics. Highly ranked elements across all five districts were student outcomes as indicated by standardized achievement test scores (reading, math, language) as well as affective outcomes such as attitudes toward school and academic self-concept. A broad range of student characteristics were viewed as important, including identification information such as sex, ethnic background, years at current school, and program designation (e.g., Chapter I, Special Education, Gifted). Highly ranked instructional practices included primary learning goals and objectives, instructional time, and expectations for achievement and class conduct. Important contextual features included quality of worklife (for teachers, school staff, and administrators), school climate, and parent involvement. In addition, each district designated specific

elements within each category as important based on their unique situation, improvement priorities, and concerns.

Following screening for measurement feasibility and political consequences, consensus was reached that the following data elements would comprise the core database system:

Background Information About Students

Age
Grade level
Sex
Ethnic background
Time at current school
Time in district
Attendance/absence rate
Socio-economic status
Language status
Special program participation

Information on Student Outcomes

Reading achievement
Math achievement
Attitude toward reading, including liking, perceived importance, self-confidence
Attitude toward Math, including liking, perceived importance, self-confidence
Attitude toward school, including motivation, academic self concept, sense of control, instructional mastery

Classroom Processes

Use of instructional time
Expectations of achievement
Amount of homework
Use of individualized instruction
Use of instructional resources and materials
Student instructional preferences (materials and activities)

School Content

School climate: Perceptions of physical plant
 Perceptions of principal
 Perceptions of teachers
 Perceptions of other students
Parent participation
Frequency of parent help
Parent support for school
Parent knowledge about school

What kinds of analysis and reports are desired?

Presented with a variety of options, users appeared torn between simple visual displays which graphically highlighted

trends or group patterns on one or a few variables and their desire to see "everything at once" on a single page or on a single screen. Thus, although almost everyone in the group found graphics more appealing than numbers, they also wanted rosters that would enable them to see all scores at once. In general, as one might expect, district superintendents were more interested than teachers in looking at trends over time and were more sophisticated in their ability to analyze the data in depth and in their ability to understand more complex displays (e.g., analyses of score distributions over time). Teachers, in keeping with their responsibilities, were more satisfied with simple bar charts which enabled them to analyze their classes at single point in time. Both reporting formats and preferred types of analyses, in other words, differed for the various user groups. Based on initial preferences, the following reports by user group appeared to be desirable:

District Superintendents

Student achievement in reading and mathematics and their attitudes for the district as a whole and for each school, including longitudinal tracking of the same cohort over several years; and cross-sectional analysis of the same grade levels over time. They were interested in displays which would give them a sense of the mean as well as the score distribution, (e.g., box plots) and wanted to be able to examine the performance of all schools within their district on a single graph. They also wanted to be able to see and track over time the proportion of students in their district scoring in each national quartile;

Group comparisons (by grade) of student achievement in reading and mathematics by SES (high, medium, low), by sex, by ethnicity, by special program, by regularity of school attendance (absent less than 10 days, between 10 and 20 days, 20 or more days annually), and by years in current school (new vs. longer term resident students);

Overall school climate by school;

Scattergrams for any significant relationships found between any of the instructional or school context variables and student achievement and attitudes;

District profile and school profiles rostering all outcomes, school climate, and demographic variables.

School Principals

Student achievement in reading and in math over time by student; by class; by grade for their school; by special program participation for their school; and by student demographic characteristics;

Student attitudes by grade;

Selected instructional process and school context variables, including expectations for achievement, amount of parent support and amount of homework by student; by class, and by grade;

Relationships, if any, between time and achievement, parent participation and achievement, expectations and achievement and between attitudes and achievement.

Teachers

Roster of individual students to include all student background characteristics except SES; all outcomes; parent support/help with schoolwork; instructional preferences, and perceptions of the school climate;

Breakdowns of their class by grade level; ethnicity; attendance rates; special program status; each outcome; each instructional process and school context variable;

School by grade level breakdowns by ethnicity; absence rates; language status; special program participation; sex.

Design Consideration

The above preferences provide a blueprint for analysis, without regard to the appropriateness, technical quality, or confidentiality of particular data sources. For example, teachers wanted individual responses about students' attitudes and school climate (including perceptions of the teacher). Yet it is questionable whether student attitude measures are sufficiently reliable at the individual level to warrant that level of diagnosis and attention; it is likewise moot whether students will answer honestly about their perceptions of the teacher if they know that their teacher will have direct and easy access to their responses. Similar questions arise with regard to teachers' or principals' responses to sensitive school issues. (This, in fact, was the reason why "quality of work life" was deleted from the original set of system elements.)

The reporting priorities articulated above also are generally mute about what constitutes appropriate and meaningful summary statistics for reporting various process and outcome indicators. These are partially technical decisions based on the nature of the assessment devices employed but user preferences are equally important if utility is to be maximized -- i.e., what kinds of summaries are perceived as most familiar, easily understandable and/or meaningful? With norm-referenced assessments, for instance, a wide variety of derived scores are possible, e.g., percentile scores, NCE's Grade equivalents, stanines, quartiles; and each type of score can be characterized in a variety of ways: mean, median, percentage of students scoring above, below, and/or within a particular score range, e.g., mean percentile scores, percentage of students scoring above grade level, percentage of students scoring in the highest quartile compared to the national norm group. Further, what constitutes meaningful cut-off point for reporting score distributions will vary depending on the local context and priorities. For example, an inner city school might want to examine the percentage of students scoring at or above the national average as an indicator of effectiveness, while the cut-off point for a more advantaged suburban environment might be quite different.

The choice of meaningful cut-off scores, in other words, is an interpretation issue that needs to be resolved during analysis. Regardless of whether the measures are more criterion-referenced or nationally normed, the nature of the local distribution and human judgment will need to be taken into account. In our study, many of the student questionnaire items, including the attitudes toward reading and mathematics scales, used Likert type scales that generally represent the range from very negative to very positive. How should mean scores from such measures be interpreted? What represents a positive response, a negative response, a neutral response, particularly given the nature of self-report measures? Is there a cut-off point above which or below which scores deserve special scrutiny? Similarly with the interpretation of attendance data. What is satisfactory attendance? What level indicates a potentially significant problem?

The interest across all groups in an "everything at once on a single page" roster that might provide an overall picture of quality and performance and at the same time enable users to detect potential trouble spots gives rise to additional scaling and interpretation concerns. How do users compare performance across various indicators, particularly when some are norm-referenced, some are criterion-referenced, and others reflect different scales? An intuitive solution was used to solve the problem. To counteract evaluation's negative image, the reports were

designed to emphasize the positive; we chose group summary indicators that would be constituted as "percent responding positively." What counted as "responding positively" was defined by the measure as interpreted by local users: for norm-referenced achievement measures, it meant scoring at least one-half year above grade level; for the norm-referenced attitude measure, it meant scoring at or above the 70 percentile; for point scale. Additional work needs to be conducted to arrive at more elegant, technically grounded solutions, but the point to emphasize is that users wanted and needed some kind of common scale against which they could interpret all the data.

Usage considerations. As users examined the reports, a number of observations were apparent. (Figures 2-11 in the appendix display some sample reports.) First and foremost, teachers and principals generally were uncomfortable in dealing with numbers and needed considerable support in understanding them. This was not necessarily a problem with the reports themselves but rather speaks to the extensive orientation/training that educators may need prior to or accompanying system use. What do the different scores and statistics mean? How should they be interpreted? What's a productive strategy for delving into the data? Further, this apparent anxiety about numbers and dealing with data meant that displays need to be labelled as clearly and as completely as possible and short-hand titles or abbreviations avoided. To help guide naive users' inquiries, it may also be helpful to frame displays in terms of the question(s) that the data can help answer.

The technical naivete of the potential users brings with it also the problem of guarding against the misuse/misinterpretation of the data. For example, in one district report, students' test score performance was compared by ethnic group. In several cases, there was only a couple of students representing a particular group and any conclusions would be unfounded and erroneous. Rather than assuming that users will know when particular analyses are inappropriate, it may be better to program the system to suppress analyses under given conditions. This parallels the suggestion made earlier regarding suppressing access to data that may violate privacy or standards of technical quality for particular levels of use. A similar issue relates to data access. Who shall have access to what data? Are there political or other reasons to restrict access to particular data elements or particular levels of analysis? What safeguards need to be provided and how?

Another observation relates to the continuing tension between individualized reporting options and ease of report access. It was clear with the "at a glance" rosters, for example, that different users representing the same role group wanted different data elements included on the form

(it is not possible to include everything on a single page or screen); as another example, there were many individual differences in preferred graphic displays and tolerance for numbers of elements displayed. A reasonable compromise may be to provide standard reporting options for easy access, but enable more dedicated or more computer-comfortable users an option to design their own analysis forms.

Finally, it appears that the types of reports desired by the different levels of users may need to vary not only in the level of analysis but in the sophistication of the display. Superintendents continued to be interested in stem and leaf plots and other displays which gave them a sense of the score distributions while teachers were desirous of more simplified pictures. To avoid endless arrays of menu selections, it may be more effective to branch the program by user group and customize the reports to each groups' needs; reports may also need to be customized for each individual district. In any event, additional interactive work is needed with each user group to be more sensitive to their preferences, interests and concerns.

Summary and Conclusions

The field test of a prototype multilevel evaluation model in five school districts produced a number of important lessons for future project design. First and foremost, teachers, principals, district administrators, and school board members were interested in getting better information about the quality of their schools and interested in a broad array of information to aid their decisionmaking. They were enthusiastic about both broadening available data beyond standardized tests and being involved in the decision process. There also was substantial agreement across the various groups on the types of indicators and data that would be most beneficial.

However, data-based decision making is a new concept for most teachers and principals, and although familiar to district administrators and policymakers, they have little experience with its many possible iterations. The amount of support intended users need in envisioning a comprehensive system and how its data might be used to help them to accomplish their responsibilities should not be underestimated. For example, users needed far more orientation to the model concept, to the potential role of data in teaching, school and district decisionmaking and policy needed, and to specific, concrete examples of use prior to trying to articulate their own information needs or subsequent analysis and reporting needs.

Further, and related to the first point, because a data-based information system represented a new idea and an innovation in the ways schools and the personnel within them

typically operate, its implementation required sustained attention to the organizational and socio-political factors which facilitate change. The process of implementation was designed to promote user ownership in the system by trying to build the system around user needs and getting their input and reactions at each step; further we tried to foster district ownership and responsibility for the project by establishing steering committees within each district and requesting that one person be designated as coordinator for within-district operations. In addition, because the superintendents were enthusiastic about the project and their districts' participation in it, and because principals volunteered their schools for the project, we assumed that critical leadership support would be forthcoming as would sustained interest and attention to the project. We assumed that each district could be relatively self-sustaining and manage its own process without extensive intervention or support from the project coordinator. These assumptions, unfortunately, turned out to be partially erroneous. Bringing teachers, principals and other administrators in for several central planning meetings was not sufficient to build their ownership; considerably more interaction was required. Although steering committees were implemented and responsibilities assigned, the locus of the project apparently was perceived by some districts as outside their district -- potentially a function of the fact that participants had difficulty envisioning exactly what the final product was going to look like or what it was going to do for them, or how it fit within their work routines. In addition, crises emerged in some districts which eclipsed the salience and importance of the project and the attention it was accorded by school leadership. Time delays in the project further eroded support. The bottom line was that project activities were perhaps viewed as more peripheral than central to participants, and their project commitment and memory needed further bolstering. Future implementation will need to pay greater attention to the organizational structures and incentives supporting the project and to facilitating group process both within and across projects.

Quality control also emerged as an important problem area. Project participants in the main are unschooled in the technical requirements for rigorous data collection and coding; as a result, things which we as researchers take as self-evident (and provided directions for), e.g., the need to carefully designate student ID numbers and/or teacher ID numbers and/or school ID numbers on all completed instruments, did not receive the care we had naively anticipated. Early and repeated checks for data quality, in short, need to be built into the system. At a minimum, districts needed more precise and prescriptive directions for handling data and assignment of ID numbers; in our directions, we tried to be responsive to individual differences in district practices by providing flexible

guidelines. Our good intentions, however, ended up doing the districts a disservice; more prescriptive rules would have been easier to follow. In addition, any data entry process should routinely check for out of range values and for consistency and accuracy of ID numbers.

Fourth, while data about school and instructional process are critical in a sound evaluation system, the feasibility of collecting data that is sensitive to intended uses bears further scrutiny. It is moot whether easily collected self-report data are sufficiently precise to support school and class level planning or process-outcome analyses. However, while more indepth observational approaches are possible, their time, resource and commitment requirements raise difficult cost-benefit questions.

Finally, we are left with an overall strategy question about the optimal approach to system development and implementation. The project reported here attempted a "top-down, bottom-up" approach to the development process, merging our own top-down vision of what the project might look like and accomplish with the bottom-up needs of the various users groups. Neither set of requirements were initially fully specified and this caused tensions and impediments throughout the development process. Rather than combining the two approaches, it perhaps would have been better to begin with one or the other: e.g., start with a fully flushed out version of an information system and the sets of questions and problems it could address, and then modify/adjust the system to accommodate bottom-up needs; that is, start top-down with an imposed order, but then let local users adapt to their context. Another approach would be to start bottom-up with explorations of the problems and decisions that particular user groups are faced with and work interactively with them to discover the ways in which data can help them and the reports and displays that are of greatest use. Which of these is the more effective approach is an empirical question worthy of future study.

REFERENCES

- Alkin, M., Daillak, R., & White, P. (1979). Using evaluations: Does evaluation make a difference? Beverly Hills, CA: Sage Publications.
- Alkin, M., Jacobson, P., Burry, J., Ruskus, J., White, P., & Kent, L. (1985). A guide for evaluation decision makers. Beverly Hills, CA: Sage Publications.
- Baker, E.L. (1983). Evaluating educational quality: A rational design. Invited address, Educational Policy and Management, University of Oregon.
- Bank, A., & Williams, R.C. (1984). The search for consequences: Assessing the impact of district instructional information systems. Evaluation and Program Planning, 6 (3).
- Bank, A., & Williams, R.C. (1985). From program evaluation to instructional information systems. Studies in Educational Evaluation, 11 (2).
- Berman, P., & McLaughlin, M.W. (1977). Federal programs supporting educational change Vol III: Implementing and sustaining innovations. R-1589/8-HEW, Santa Monica, CA: Rand Corporation.
- Cohen, D.K., & Garet, M. (1975). Reforming educational policy with applied social research. Harvard Educational Review, 45(1), 17-43.
- Dorr-Bremme, D. (1984). Contextual influences in developing a school-based comprehensive information system. In L. Burstein, & K. Sirotnik, Comprehensive information systems for local school improvement: A reality-test in secondary schools. Report to NIE. Los Angeles: UCLA Center for the Study of Evaluation.
- Dussault, N.F. (1985). Integrating instructional information systems into the instructional process. Presented at the invitational conference "Information Systems and School Improvement: Inventing the Future," UCLA Center for the Study of Evaluation, Los Angeles.
- Eisner, E. (1985). The art of educational evaluation: A personal view. Philadelphia, PA: The Falmer Press
- Goodlad, J.I. (1983). A place called school. New York: McGraw Hill.

- Heckman, P.E., Oakes, J., & Sirotnik, K.A. (1983). Expanding the concepts of renewal and change. Educational Leadership, 40, pp.26-32.
- Herman, J., & Dorr-Bremme, D. (1983). Uses of testing in the schools: A national profile. New directions for testing and measurement, No. 19. San Francisco: Jossey-Bass, pp. 7-17.
- Idstein, P. (1985). Let's create the future but let's keep it a secret. Presented at the invitational conference "Information Systems and School Improvement: Inventing the Future," UCLA Center for the Study of Evaluation, Los Angeles.
- Lucas, H.C. (1975). Why information systems fail. New York: Columbia University Press.
- Markus, M.L. (1981). Systems, Objectives, Solutions. Implementation politics: Top management support and user involvement. Amsterdam: North Holland Publishing Co.
- Multinovich, J.S., & Vlahovich, V. (1984). A strategy for a successful MIS/DDS implementation. Journal of Systems Management, 35, pp. 8-15.
- Patton, M.Q. (1986). Utilization-focused evaluation. Second Edition. Beverly Hills, CA: Sage Publications.
- Sarason, S.B. (1982). The culture of the school and the problem of change. Second Edition. Boston: Allyn & Bacon.
- Sirotnik, K.A. (1984). An outcome-free conception of schooling: Implications for school-based inquiry and information systems. Educational Evaluation and Policy Analysis, 6(3), 226-239.
- Sirotnik, K.A., & Burstein, L. (1984). Making sense out of comprehensive school-based information systems. Los Angeles: UCLA Center for the Study of Evaluation.
- Sirotnik, K.A., Dorr-Bremme, D., & Burstein, L. (1985). Social vs clinical perspectives in the use of information: Implications for school-based information systems. Systemic Evaluation Project Deliverable to NIE. Los Angeles: UCLA Center for the Study of Evaluation.
- Weick, K.E. (1978). Educational organizations as loosely coupled systems. Administrative Science Quarterly, 21, pp. 1-19.