ABSTRACT
        The Multilevel Evaluation Systems project was field
tested in grades 4 and 5 classrooms in five districts of the
University of Pennsylvania's School Council. Multilevel evaluation
coordinates multiple sources of information (about student
characteristics, school context, school and instructional processes,
and a range of student outcomes) with multiple reporting targets
(students, classes, schools, school boards, districts, state and
federal government) to facilitate data-based decision-making at all
levels. The field study attempted to set up user-based data
collection by first identifying needs and concerns common across
groups as well as concerns unique to each district. The core database
included variables addressing: (1) background information about
students; (2) information on student outcomes, both academic and
affective; (3) classroom processes; and (4) school content including
parent involvement. The types of reports desired by district
superintendents, school principals, and teachers were compared.
District leaders and principals favored analysis of trends over time,
whereas teachers wanted a detailed breakdown of student
characteristics and current achievement. All groups preferred that a
one-page synthesis of data be designed to meet their particular
needs. Steps for insuring quality control in data collection and for
providing support in data interpretation and use were recommended.
(LPG)

Center for Research on Evaluation,

Standards and Student Testing

Deliverable - March 1987

PROJECT: MULTILEVEL EVALUATION SYSTEMS

Evaluation for school improvement: Try-out of
a comprehensive school-based model

Project director: Joan Herman

Grant Number: G008690003

CENTER FOR THE STUDY OF EVALUATION

Graduate School of Education

University of California, Los Angeles

EVALUATION for SCHOOL IMPROVEMENT: TRY-OUT of a COMPREHENSIVE SCHOOL-BASED MODEL

Joan L. Herman
CRESST
UCLA Center for the Study of Evaluation

Overview

"How well are we doing?" "How can we make things better?" are questions school boards, administrators and educators are constantly asking. But while districts often collect a great deal of data as part of their routine evaluation activities, many feel that such data does not well answer their questions. Collected in the names of sound management and rational decisionmaking, the data instead sits unused on bookshelves, in thick computer printouts, and in often inaccessible computer files, with little or no significant impact on the process of education in districts, schools, or classrooms.

The Multilevel Evaluation Systems project seeks a more useful approach to evaluation. It seeks to develop and implement a "top-down, bottom up" evaluation model (Baker, 1983) that will provide context sensitive information for principals and teachers to help them improve their instructional programs while simultaneously providing superintendents, board members, and other administrators with information for policy decisions. 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 knowledge and its impact on educational innovation;

The project model draws on accumulated knowledge about what makes schools effective, about what makes evaluative information useful to teachers and administrators; about what makes information systems useful in organizations; and on the power of currently available, low cost microcomputer technology. In the sections which follow, the rationale underlying the project model

is summarized briefly and the technical approach and its results described. We end with consideration of problems which emerged and potential solution strategies.

Background

The model starts with the assumption that evaluation can be a valuable tool for improving schools, that the collection, analysis, and distribution of information can stimulate and inform action to upgrade the quality of education. It assumes that such 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 such tasks as assigning students to groups, diagnosing individual learning problems, monitoring student progress, assessing subject matter mastery, identifying students who need remediation or enrichment activities. A principal and teachers working together could use information about school context, instructional processes and outcomes to analyze local problems and improve the effectiveness 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 and Garet, 1975; Patton, 1986). Why the discrepancy? The reasons are many and varied. Chief among these has been the source and nature of formal evaluation practice over the last two decades. 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 and Dorr-Bremme, 1983). They claim further that the paperwork and bureaucratic burdens associated with mandated evaluation requirements intruded into, rather than

2

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 and 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. You cannot simply 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. The quality of school programs is only one of many factors which contribute to student test scores. Cultural, social, economic, demographic and motivation factors 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 were administered, and the results were to be integrated within a social/economic/community context, there would remain a further, serious deficiency in many previous evaluation conceptualizations. 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 focussing 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.

School context has been neglected not only as an source of explanatory hypotheses about why outcomes are as they are but also as an important intervening factor which influences 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 and 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 field outside of education provides clues
on some of the socio-organizational-political issues involved in
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.

The above analysis suggests some of the reasons 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   were   not
necessarily sensitive to "bottom-up" needs; evaluation data has
been derived primarily from tests of student achievement which
examine only a narrow range of outcomes; evaluation often ignores
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.

But there are possibilities for rethinking evaluation systems
so that they serve multiple users and their diverse information
needs. Some school districts are currently moving in this
direction (Williams & Bank, 1984, 1985; Idstein, 1985; Dussault,
1985). Radical changes in evaluation thinking are emerging which
reflect 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 in schools, schools and classrooms which encompass tremendous
diversity in student population, in teacher skills, in curricular
goals, in teaching strategies. Because of this diversity as well
as because actual control over instruction resides in the school
building, rather than in more remote and larger administrative
units, the appropriate unit for solving many educational problems
is the school (Goodlad, 1983; Baker, 1983). Consequently, school
personnel are among the appropriate beneficiaries of improvement-
oriented evaluation systems. But individual schools may not have
sufficient resources, expertise, control, etc. to solve all their
educational problems by themselves. The solutions often require
initiative, direction, resources, and/or actions at high
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 distributed system of evaluation which could
provide local schools with a rich, locally sensitive information
base to aid their problem-solving but which could also provide
appropriate aggregate information for decisionmaking at high
levels of the system.

4

The Project Model

Inherent in the foregoing analysis of problems in current evaluation practice are the roots of a more productive model for improving the quality of schools. What are its features? An ideal system:

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

2. enables efficient sharing of information within and across levels of the educational hierarchy, minimizing redundant, overlapping testing and evaluation requirements;

3. includes information on a range of school outcomes;

4. includes information on school context and student characteristics to contextualize outcome and effectiveness analyses;

5. includes information on school and instructional processes to elucidate and analyze local problems and accomplishment;

6. links outcome information with instructional process and school context data to provide explanatory power for findings

7. includes externally fixed elements to assure sensitivity to the information needs at the district and state levels and variable, locally selected elements and measures of interest to school professionals.

8. encourages data collection, analysis, and use over time

9. 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 of all groups

-making the system user-friendly and easily accessible

The project model, in short, features the use of a comprehensive information base about student characteristics,

5

school context, school and instructional process and a range of outcomes that can be analyzed, arrayed, and appropriately reported 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.) Critical to the model is that its constituent elements are collaboratively defined and its implementation managed to 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; he was responsible for facilitating and coordinating planning and implementation. Steering committees were constituted within each district to assure their 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 was encouraged designate one member of project coordinator for their district. CSE 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.

Utilizing this organizational structure, 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. Data Analysis and Reporting

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,

6
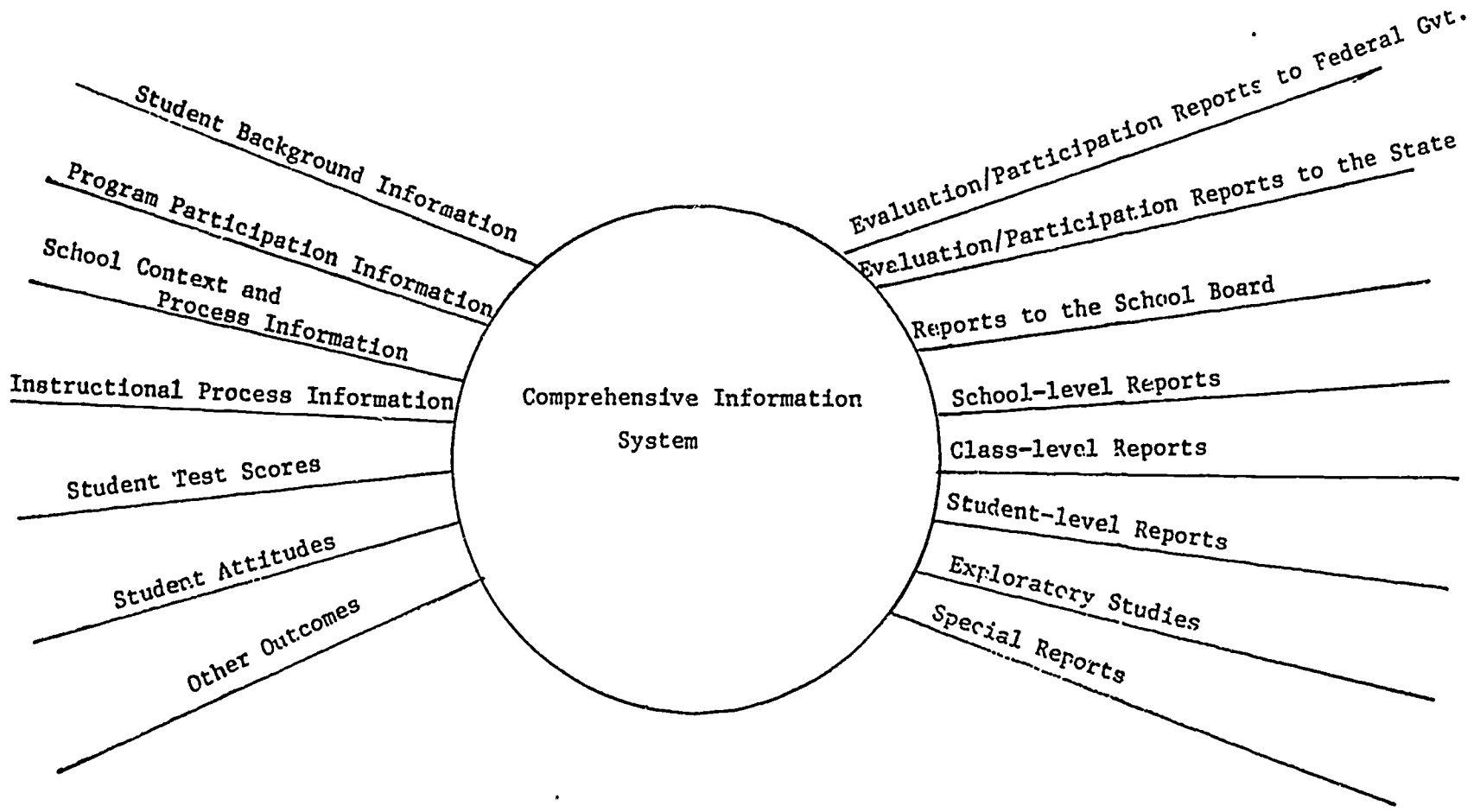
OVERVIEW OF MODEL SYSTEM

Figure 1

10

11

principals, district administrators, superintendents and board members. (To enhance initial reporting flexibility and to avoid potentially costly reprogramming efforts, initial analyses 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 requirements, and to review instrumentation and reports. Data collection proceeded in two fourth grade and two fifth grade classrooms in each participating school; data collect. a included a combination of rostering archival data, administering a commercially published student attitude measure, and administering specially developed student and teacher questionnaires.

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. These 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

8

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

9

<u>Classroom</u> <u>Processes</u>
Use of instructional time (TQ,SQ)
Expectations of achievement (SQ)
Amount of homework (SQ,TQ)
Use of individualized instruction (TQ)
Use of instructional resources and materials (TQ)
Student instructional preferences (materials and activities)

<u>School</u> <u>Content</u>
School climate (SQ):   Perceptions of physical plant
                        Perceptions of principal
                        Perceptions of teachers
                        Perceptions of other students
Parent participation (TQ,SQ)
Frequency of parent help (SQ)
Parent support for school (TQ,SQ)
Parent knowledge about school (TQ)

    <u>What</u> <u>kinds</u> <u>of</u> <u>analyses</u> <u>and</u> <u>reports</u> <u>are</u> <u>desired?</u> Presented
with a range of options, the various user groups identified which
would be most useful and helpful in their planning and
decisionmaking. An interesting tension emerged between simple,
visually appealing displays which could help users better grasp
trends and patterns with regard to particular variables and a
desire to see "every at once" on a single page or on a single
screen. Thus though 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 desire 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. Specific requests by role group were as follows:

    <u>District</u> <u>Superintendents</u> wanted reports on:

        Student achievement in reading and mathematics and
        their attitudes over time for the district as a whole
        and for each school, including longitudinal tracking
        of the same cohort over several years; tracking of the
        performance 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,
        and wanted to be able to examine the performance of all
        schools in their district on a single graph.   The also
        wanted to be able to see and track over time the
        proportion of students 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

10

school attendance (absent less than ten 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 wanted reports on:

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 wanted reports on:

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.

The above preferences provide a blueprint for analysis,
without regard to the appropriateness, technical quality, or
confidentiality of particular responses. For example, teachers
want individual responses with regard to 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 and whether students will answer honestly

11

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 preferences articulated above also are generally silent
with regard to the types of scores upon which they should be
conducted. Except for the district superintendents who were
direct in their requests for score distributions and changes in
percentile ranges, users did not mention, and were not asked
about, the types of scores and cut-off points which they would
find meaningful. For example, the student questionnaire items,
including the attitude toward reading and mathematics items, used
Likert-type scales that generally represented the negative to
positive range. How should mean scores from such measures be
interpreted? Is there a cut-off point above which or below which
scores deserve special scrutiny? Based on experience with self-
report measures, it was decided that mean scores at or above 3.8
on a five point scale would be considered as significantly
positive, and percentages of students responding at or above this
level would serve as a summary indicator of response. Other
decisions clearly were and are possible.

Even with norm-referenced measures, the choice of meaningful
score categories remains. For example, on a school profile which
seeks to give information on all indicators at a glance, what
single indicator should be used to characterize students'
performance on a standardized reading test? The mean percentile
score? The percentage of students scoring at or above a certain
percentile or stanine? And if the latter, what is a meaningful
cut-off probably will differ in a traditionally low scoring
school versus one serving a very advantaged community.

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, it was
decided that the reports would emphasize the positive and it was
further decided that on group summaries, summary indicators would
be constituted to represent "percent responding positively."
What counted as "responding positively" was defined by the
measure: 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 questionnaire items, it meant mean responses
above 3.8 on a five 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.

How did users react to the analyses and reports? As users examined the reports, a number of observations were apparent. First and foremost were that the 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 naivity 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 display which gave them a sense of the score distributions while teachers were desirous of more simplified

13

pictures. To avoid endless arrays of menu selections, it may be more effective to branch the program by user groups and customize the reports to each groups needs; reports may also need to be semi-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, and as they gain experience in using data they may be better able to articulate those 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, data-based decisionmaking 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 they 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.

Second, 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; considerable more interaction apparently was required. Although steering committees were implemented and responsibilities assigned, the locus of the project apparently was perceived in some districts as outside their district --potentially a function of the fact that participants has difficulty envisioning exactly what the

14

18

final product was going to look like or what it was going to do for them. 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.

Third, quality control emerged as an important problem. 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 in-depth observational approaches as possible, their time, resource and commitment requirements raise difficult cost-benefit questions.

Given the complexity and relatively limited resources of the project, it may have been overly ambitious to try to develop and implement a user-based system in five districts simultaneously. The number of accommodations that needed to be made to arrive at a common set of data elements for the multiplicity of users across all districts perhaps distanced the system too far from any single user's or groups needs and perhaps mitigated against feelings of ownership and control. In retrospect, too, each district needed more individualized support and help in customizing the data collection and coding requirements to their context, e.g., in assigning student id numbers, in creating tape specifications, in communicating to the data analysts the meaning of the particular unique coding schemes. In the interests of efficiency and conservation of time and staff resources, we attempted to make everything as uniform as possible across

15

19

districts; these efficiencies turned out to be costly for project effectiveness. Having the primary technical/data expertise available primarily at long distance also proved to be an ineffective strategy; sites needed easier access to technical assistance and more frequent feedback

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.