SCHOOL DISTRICT USE OF TESTING AND EVALUATION
FOR INSTRUCTIONAL DECISION MAKING: A BEGINNING

Adrianne Bank
Richard C. Williams

TABLE OF CONTENTS

INTRODUCTION

During this decade of the '80's, educational organizations -- like profit-making and governmental organizations -- are experiencing considerable stress. School systems are being asked to respond to demands for increased excellence in educational quality while their financial resources are declining. If educators are to secure the funds necessary to meet the diverse needs of a heterogeneous, mobile student population and to raise the low status and low salaries of teachers, and administrators, they are going to have to devise ways to more actively and effectively manage their instructional programs and then let the public know about it. An important ingredient in this process is information for teachers, principals, and the public about the effects of the instructional program and discussion concerning the implications for action.

School district offices currently obtain a great deal of information about student achievement. They also know some details about the operation of their instructional programs. Over the recent past the districts' human and technical capacity to collect data about students and programs has been enlarged by federal and local need-to-know reports concerning the outcomes of funded programs. Most large districts now have research and development offices. Many districts routinely administer standardized, norm-referenced, criterion-referenced tests, and proficiency tests. Many districts annually distribute parent, teacher, and student opinionnaires. The use to which this data is put, however, varies widely from district to district. Some districts largely report the data to external funding agencies; other districts utilize the data for internal district instructional decision making.

For the past three years, the authors have been conducting NIE spon-
sored field research at UCLA's Center for the Study of Evaluation in order
to understand what distinguishes districts who make internal, system-wide
instructionally-oriented use of student test scores or other evaluative
data from those districts who use mandated tests scores and evaluation
reports primarily to comply with federal and state requirements.

We are aware of the numerous technical difficulties related to devel-
oping test instruments and analyzing test scores as well as the other eval-
uative data from surveys, classroom observations, etc., in ways that en-
courage appropriate inferences for instructional policy making. We acknow-
ledge that even when these technical difficulties are overcome, generic
organizational characteristics of school districts as organizations make
the kind of information-based instructional management systems we have seen
in some districts relatively rare occurrences.

However, we have discovered some common conditions in the eight dis-
tricts we have studied that appear to contribute to their ability to forge
systemic linkages between testing and evaluation data on the one hand and
instruction on the other. We believe that similar facilitating conditions
may exist in other districts, or can be brought into existence, if the
motivation is sufficient. Our study districts are in what we have labeled
"phase three" in regard to data collection, analysis, and use, that is, the
integration of testing and evaluation with instructional decision making
We are not certain at this time if this third phase phenomenon will dis-
appear or whether it marks a beginning of a useful movement in district
instructional management.

BACKGROUND

In earlier, simpler -- phase one -- times, teachers used their own made-up tests to motivate students, to determine levels of student achievement in order to provide remedial or enrichment activities for them, and to assign required report card grades. Teachers, autonomous in their classrooms, were responsible for testing students, for evaluating their performance in other ways, and for teaching them. Districts did not require system-wide tests.

During the recent twenty-year history of high levels of federal and state involvement in the formulation and funding of educational programs, teachers continued to use self-developed tests, but district offices began to require additional tests or evaluative record keeping, either on their own initiative, or because of requirements accompanying externally funded categorical programs. Many districts now routinely require the administration of annual or semi-annual standardized tests at three or more grade levels, periodic criterion-referenced tests, proficiency tests, and state assessment tests. About half of the testing occurring in elementary schools in reading and math, and about one quarter of the testing occurring in high school classrooms is initiated by someone other than the classroom teacher (Burry, et al., 1982). Many observers, (David, 1978; Lyon, et al, 1978; Zucker, 1981) point out that, for the most part, test results are sent out of the district to fulfill reporting requirements. Whatever internal use is made of them is likely to be by administrators of the programs, rather than by teachers or principals. It is our finding as well as those of others (Kennedy, et al., 1981; Alkin, Daillak & White, 1981) that most districts do not relate their district-wide testing and

evaluation activities to their ongoing instructional programs. Those districts where district-wide tests are routinely administered or where evaluative record keeping is performed but where test, scores, and evaluative data are rarely scrutinized for policy or classroom purposes could be said to be in phase two with regard to testing and evaluation activities.

In a small, but growing, number of districts, school boards and administrators have found a way to utilize for internal decision making the information collected by their recently acquired research and evaluation personnel. These boards and administrators have decided, in response to their own, to teachers', and to the public's demands for better pupil achievement that test scores and other data can guide policy and classroom decisions as well as track student performance. These districts, due perhaps to particular circumstances, have developed what we are calling instructional information systems. These systems use information derived from a variety of data sources, e.g., student scores on one or more test formats, parent surveys, demographic trends, to analyze, and perhaps change, instructional programs, including texts, supplementary media or materials; supports or training for teachers; amount of time on specific subjects, etc.

This "phase three" linkage of testing, evaluation, and instruction is characterized by district acknowledgement of system-wide responsibility for student learning and by district confidence that test scores and other evaluation data can be used as planning and assessment tools.

The two questions addressed by this paper are: Why have some, but not other school districts moved towards the development of system-wide

linkages between testing and/or evaluation with instruction? What do existing linkages, that is, instructional information systems, look like?

WHY SOME DISTRICTS AND NOT OTHERS?

We have identified several generic characteristics of school districts as complex organizations that tend to keep most of them at phase two; that is, lots of test data collected at district behest, little test data used at district direction to improve the instructional program.

These characteristics predispose district central offices towards short-term reactions to their ever-present and constantly changing problems rather than towards sustained long-term efforts to improve instruction in a cumulative and systemic manner. They also predispose central office research and testing units towards autonomous, in contrast to coordinated functioning. Furthermore, individual teachers or schools assume, implicitly or explicitly, delegated authority for most classroom decisions. Central office responsibility for instruction is limited to inconsequential support and supervision activities. These generic characteristics include:

° rapidly changing environmental conditions. Fluctuations in district budget levels; changes in student population characteristics and distribution within the district; reductions in staffing all make it difficult for district administrators to find the time, resources, or constancy to develop a coordinated instructional improvement effort monitored and guided by information from tests and evaluations.

° boundary permeability. There is a great deal of external societal regulation of districts and schools. Specific laws and

regulations, as well as changing and sometimes conflicting community expectations, may dispose district administrators to attend first and foremost to those tasks for which they are legally and politically accountable even if those tasks do not most directly contribute to instructional excellence (Zucker, 1981).

°   goal diffuseness. The pluralism and diversity within our society and within any given district may make it difficult for districts to adopt educational goals specific enough to tune tests, texts, instruction to them. In order to avoid offense, educational goals are often stated in ambiguous language.

°   weak technical core. Some observers have noted that education's weak technical core -- that is, the small number of credible research-validated cause-effect relationships between student instructional methods and learning outcomes -- encourages educators to use and justify a variety of instructional methods. Arguments have been made (Meyer & Rowan, 1977) that revelation of this weak technical core may be detrimental to a district's public image and therefore administrators leave such decisions to teachers who have the imprimatur of being credentialed.

°   loose coupling. A widely held view of the internal operations of public school districts is that the instructional components of the organization, e.g., curriculum department, in-service training, individual classrooms, are loosely coupled (Meyer & Rowan, 1977; Weick, 1976). It follows that the linkage between policy and management and between management and operations is likely to be

weak. Loose coupling within a district may be more pronounced in the complex areas, such as instruction, which are protected from public view than in simpler, but more visible areas like the enforcement of policies relating to corporal punishment. Instructionally, schools and classrooms may be thought of as federations or zones of influences (Lortie, 1975), rather than as places where board or central office policies necessarily are carried out by teachers in classrooms.

In light of these generic conditions, what are the situation-specific elements present in those districts which have developed information-based instructional management systems? The elements can be conceptualized as the interaction between a benign external environment (Pincus & Williams, 1977) and the organization's capacity to respond.

In the districts we studied the thrusts from their many environments, for example, federal, state, local, media, academia, parents, were aligned in the same direction. Conjointly, these pressures pushed the district towards making instructional changes to increase student achievement as reflected in test scores. In addition, it seemed to us that the political, social, and economic situations in our study districts were somewhat less turbulent, somewhat more stable than in other districts where crises seemed continually chronic.

Our study districts were able to respond to the direction suggested by the pressures from the environment because of: the presence of influential "idea champions," and of a stable core of central office personnel, staff orientation towards comprehensive problem analysis, and an administrative capacity for dealing with ambiguity and delay (Bank & Williams, 1981).

Idea champions were defined as individuals in key administrative and policy positions who firmly believed in some variation of the following position: that classroom instruction could be conceptualized in terms of student learning outcomes, that tests of student learning outcomes and evaluations of instructional activities could be used to relate instruction to learning, and that district policies and procedures should be attentive to this linkage.

Stable core staff meant that a critical mass of teacher and administrative supporters of these ideas had been around for a while. In districts with the most advanced information systems it had taken more than five years for the linkages to develop and mature. A stable core group, with some additions or deletions, seemed to have worked steadily at the process over several years. They were thus able to smooth out the shifts in direction that might have resulted from changes in school board composition, legislation, court orders, funding levels, etc.

Comprehensive rather than ad hoc problem analysis on the part of district managers meant that the core central office staff prepared for next steps beyond the immediate task of the moment. Typically our districts did not develop elaborate, multi-year blueprints of their projected information management activities in advance of action. However, from the beginning they did have a sense of where they were going and the steps that were necessary to get there.

District personnel in our study seemed to have the ability to deal with delay and ambiguity; that is, they were aware of, and accommodated to a strategy for developing an instructional information system that was

uneven, interrupted, and in some cases oft-postponed. Creating arrange-
ments to merge together usually-separated district operations into integra-
ted configurations took time, patience, administrative talents, and adher-
ence to an overall idea rather than to specific details.

WHAT DO INSTRUCTIONAL INFORMATION SYSTEMS LOOK LIKE?

The information systems of the districts we studied varied on a number
of dimensions; among them, the purpose of the system, the number of ele-
ments which were coordinated, and the locus of decision making.

Here are thumb-nail sketches of two districts which differ substan-
tially on these three dimensions. District A is small and has a central-
ized, highly coordinated management system, characterized by a district-
wide curriculum scope and sequence, district-wide criterion-referenced
testing, district-wide staff development activities, clear expectations for
principal supervision, classroom supports in the form of learning and media
specialists. Its purpose is to provide ongoing information and resources
to classroom teachers who then differentiate their instruction to meet in-
dividual needs. District B is medium sized, has a decentralized, school-
level decision making focus. The district sends out a variety of testing
and survey data to help school planning teams do their own resource alloca-
tion. The district's purpose is to provide information and support to site
teams who then help solve school level problems. The sketches are adapted
from district administrators' own accounts.

District A: A Centralized Information System to Monitor and Improve
Classroom Instruction

In 1970, the district attempted to meet the need for student progress
reporting by purchasing test items to measure a sequence of behavioral

objectives. The objectives themselves turned out to be sequenced differently than those existing in classroom materials. Teachers hated the objectives and the tests and eventually district administration discarded both.

Rather than purchase another available testing system, district administration formed teacher committees to write their own tests. The feeling was that tests written by district teachers would be better received by other teachers than tests developed by an outside agency. In 1973, the first district-developed tests, known as PAL (Pupil Assessed Learning), were ready and administered to students. They were intended to monitor student learning so as to feed back information to teachers about how individualized instruction was working. They also were seen as a way of reassuring parents who were concerned about their children's progress.

Teacher reaction to PAL was extremely negative. Everyone complained. Teachers today remember their complaints: e.g., "It was a waste of time." "It didn't tell us anything we needed to know." "It was not coordinated with anything we taught." Teachers loudly made their feelings known to parents and to the Board. It was a hard time for central office staff who wanted the testing system to work.

The person in charge of developing the system next appointed teacher committees to try to revise the items. Simultaneously, work began on a district level instructional continuum. A group of volunteer teachers were paid to work on a reading continuum during the summer. Their work continued through the school year.

By 1976, teacher committees had generated a reading continuum (RIC) and the beginning of a math continuum (MIC). By 1976, also, the PAL criterion-referenced testing system had been scrapped. Other teacher committees had written reading, math, and language arts test items to form their own criterion-referenced tests (CRT).

Realizing that a testing program by itself could not improve instructional practices, the central office instituted a staff development program in the early 1970's. Nearly all teachers are trained in a diagnostic-prescriptive approach to teaching. A modified version is provided for substitutes and aides. Other staff development programs discuss individualizing instruction, meeting affective needs, using inquiry skills techniques, and teaching specific curriculum content such as reading, writing, or mathematics.

At the same time as the staff development program was being operationalized, the district office decided to create a new school-based role -- that of a learning specialist whose responsibility was coordinating the newly developed testing system in each of the schools.

The district claims that their testing system holds the curriculum, the texts, the staff development, the teacher supervision together. The R&D person notes:

° All tests are directly linked to a kindergarten through eighth grade instructional continuum for reading, language arts, and mathematics.

° All tests are computer-scored. Results are formatted in easy-to-read specially tailored reports to teachers, parents, and site and district administrators.

° Results are reviewed and appropriate actions to remediate children not performing well are immediately undertaken and communicated to everyone concerned.

° The tests are reviewed on an ongoing basis with revisions occurring four times over the last several years.

District B: A Decentralized Information System to Improve Schools

"Our evaluation/planning model is one which takes place where the action is, at the school-site level. It involves those with high stakes in its outcome, namely, the school's staff, students and parents. Significantly, our model is one which begins with a broad data base. Most importantly, it goes on serving throughout the school year, long after the formal process of planning has been concluded."

Throughout the evolution of their evaluation/planning model, this district has held to several fundamental beliefs as to the reasons for evaluation and planning. They believe that an evaluation and planning model would deliver the greatest "payoff" to the district's schools if these are its primary goals:

° to improve the quality of the local educational program;

° to encourage the most effective or efficient use of scarce financial resources;

° to assist with the attainment of the goals of the school, the school district, and the special program(s) mandating the evaluation/planning activities.

Ten years ago, when the Evaluation Specialist position was created, the district had six elementary schools in its Title I program. School personnel continually asked that they be allowed to create their own program plans suited to the uniquenesses of their individual school sites. Serious doubt was often expressed by federal or state administrators with regard to the ability of local school sites to accomplish such a task.

However, in 1971, the legislature passed a bill authorizing the Early Childhood Education (ECE) program and, with it, a framework for evaluation and planning at the school level. The California State Department of Education developed the evaluation/planning model as the planning format. The creation of the ECE program prompted the district itself to develop its own participatory evaluation/planning model to be based on a broad needs assessment process. Their efforts were shaped by the state's process and forms.

The earliest versions of the staff, parent, and student needs assessment survey instruments came to be widely known as the Educational Program Assessment (EPA) process. Since its creation, the ongoing work on the EPA instruments involved large numbers of parents, staff, administrators, and secondary students. This has provided district staff with a unique opportunity to test out their theories regarding the beneficial effects of involvement on school and on community feelings of ownership and support for such processes.

"We believe that broad school and community involvement has contributed to responsive instruments. These, in turn, have led to high response rates, averaging about 70 percent and rising to an unbelievable high of 100 percent response from one school and its community."

While surveys of needs assessment responses provide a large amount of information for evaluation and planning, the primary source of the school district's evidence about student achievement is the Iowa Test of Basic Skills.

The evaluation/planning process takes people through a six-step problem solving sequence at the school-community level. Evaluation and planning activities take place under the patronage of the School Site Council, a body whose composition and influence has grown out of the California School Improvement Program (SIP) legislation. This committee, with the site administrator, represents each of the major constituencies within school-community (i.e., parents, staff, administrators, and secondary students).

For planning purposes, the school's curriculum is typically (although not always) separated into academic components, such as reading, language, mathematics, etc. The School Site Council usually creates "component committees" of parents, staff members, and students (in secondary schools) to do evaluation and planning for each academic area. The final school program plan has sections addressing each component written by the separate committees.

At the outset, each component committee tries to define its area of the curriculum comprehensively, to ensure that all sub-areas are addressed in the evaluation/planning process. The members of the component committees examine their information to identify the strengths and weaknesses of their students. With this step complete, the component committees turn their attention to the existing programs. In those areas where student performance has been found to be unsatisfactory, an effort is made to identify probable program causes, that is, program gaps or weaknesses which, if eliminated, would result in improved student performance.

Then, the School Site Council does a component-by-component review of
all aspects of the proposed program after an anticipated projected cost has
been attached to each proposed expenditure. Through negotiation and com-
promise, the activities of lowest priority, along with their related costs,
will be reduced or eliminated to the point where the amount which the
school proposes to spend will exactly equal the amount which it expects to
receive.

HOW DO INSTRUCTIONAL INFORMATION SYSTEMS VARY?

Purpose of System. In the two districts described above, the purpose
of one instructional information system was to tailor classroom instruction
to individual student needs, of the other to improve school functioning.
Some of our other study districts have more narrowly defined the purpose of
their instructional information system as to raise the average of the dis-
trict's test scores. Other districts defined the purpose of their instruc-
tional information system as monitoring teacher behavior. Their idea was
to assert district responsibility for instruction by mandating a particular
instructional strategy and then tracking the effects of that strategy by
test score analysis.

Extent of Administrative Coordination. Our eight districts differed
from one another in the complexity of their coordination arrangements at
the central office level. Usually autonomous district functions which were
linked in formal or informal arrangements included: staff development,
instructional materials (texts and adjunct materials), testing and evalua-
tion, supervision of principals, curriculum development. Coordination was
carried out by some or all of the following: informal personal links among

a few people; formal, inter-organizational reporting arrangements among units or departments; on an <u>ad hoc</u> or routine basis, central office staff and principals' discussion of district-wide and school site implications of data analysis.

Locus of Decision-making

The foregoing discussion of variations in purpose and in coordinating arrangements alluded to a number of organizational levels at which decisions were made about translating the analyses of test scores into changes in instructional activities. Districts differed from one another in the nature of the decisions which were made at the central office level, at the school level, or the classroom level. The following list indicates the range of activities we found at each level.

CENTRAL OFFICE decision-making consisted of one or more of the following:

- ° district-level construction of grade-level objectives by subject areas

- ° district-level construction of criterion-referenced tests

- ° district-wide selection and use of a norm-referenced test

- ° district-wide evaluation of instruction through classroom visits, surveys, etc., of instructional efforts

- ° district-wide formulation and conduct of staff development programs

- ° district-wide selection of texts to match tests

SCHOOL LEVEL decision-making consisted of one or more of the following:

- ° school-level planning teams, teacher/parent/community

- ° school-level receipt of information about student outcomes from either teacher-created tests, teacher-option CRT's, district-wide mandated CRT's; district-wide norm-referenced tests

° school-level conduct of instructional evaluation via supervision, teacher self-reports, district evaluations, outside evaluations

° school-level planning for school-year activities

° school-level allocation of services, both personal and financial, to support local plans

° school-level decisions about texts

CLASSROOM LEVEL decision-making consisted of one or more of the following:

° individual teacher receipt of information about student learning, e.g., norm-referenced test scores, CRT's, teacher observations, text tests, student assignments, etc.

° individual teacher decisions about grouping, remediation, enrichment, alternative instruction, etc.

° individual teacher participation in professional development activities

CONCLUSIONS

We have come to some tentative conclusions about testing, evaluation, and instruction in "phase three" districts.

1. Building an instructional information management system can be done in districts. It is, however, difficult to do, requires a specific set of advantageous circumstances, and takes time.

2. There are many yet unanswered questions about such systems. For example, what do they cost? What are their negative as well as positive consequences? Are they worth doing? Can the pace of the development process be accelerated?

3. There is no single method or design which districts have used in developing their systems. Whether such systems are uniquely configured

due to the specifics of community priorities, history of the district, personal biases of the "idea champions" or, on the other hand, because the state-of-the-art of instructional information systems is so underdeveloped, we do not yet know. We do know that there is much district interest in the use of information to guide instructional decision making; we suggest that the topic of instructional information systems be added to the agenda of researchers and practitioners interested in practical district-level supports for effective instruction.

REFERENCES

Alkin, M.C., Daillak, R., & White, P. Using evaluations: Does evaluation make a difference? Beverly Hills, CA: Sage Publications, 1979.

Bank, A., & Williams, R.C. Evaluation design project: School district organizational study. Report to NIE, CSE, 1981.

Burry, J., Catterall, J., Choppin, B., Dorr-Bremme, D. Teaching in the nation's schools and districts: How much? What kinds? To what ends? At what costs? CSE Report 194, 1982. Los Angeles: Center for Study of Evaluation, 1982.

David, J. Local uses of Title I evaluations. Stanford, CA: Stanford Research Institute Report EPRC 21, July 1978.

Kennedy, M.M., Apling, R., & Neumann, W.F. The role of evaluation and test information in public schools. Cambridge, MA: The Huron Institute, 1981.

Lortie, B. School teacher. Chicago: University of Chicago Press, 1975.

Lyon, C.D., Doscher, L., McGranahan, P., & Williams, R.C. Evaluation and school districts. Los Angeles: Center for the Study of Evaluation, University of California, 1978.

Meyer, J.W., & Rowan, B. Institutionalized organizations: Formal structure as myth and ceremony. American Journal of Sociology, 1977, 83, 340-363.

Pincus, J., & Williams, R.C. Planned change in urban school districts. Phi Delta Kappan, June 1979, 729-733.

Weick, K.E. Educational organizations as loosely coupled systems. Administrative Science Quarterly, 1976, 21, 1-19.

Zucker, L.G. Institutional structure and organizational processes: The role of evaluation in schools. In A. Bank, R. Williams, & J. Burry (Eds.). Evaluation in school districts: An organizational perspective. CSE Monograph No. 10. Center for the Study of Evaluation, University of California, Los Angeles, 1981.