DOCUMENT RESUME

ED 308 205 TM 013 415

AUTHOR Yap, Kim O.
TITLE TAC Evaluation Assistance: A Ten-Year Evolution from Accountability to Program Improvement.
PUB DATE Mar 89
NOTE 39p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Educational Assessment; Educational Change; *Educational Trends; Elementary Secondary Education; Evaluation Methods; *Evaluation Research, Evaluation Utilization; *Information Centers; Program Evaluation; *Program Improvement; School Districts; State Agencies; *Technical Assistance
IDENTIFIERS Educational Information; *Technical Assistance Centers; *Title I Evaluation and Reporting System

ABSTRACT

The Title I/Chapter 1 Technical Assistance Centers (TACs) were originally created to provide assistance to state and local education agencies in implementing the Title I evaluation and reporting system. However, over a 10-year period, TAC role has changed significantly. Its service foci have evolved from providing evaluation assistance to helping clients in the use of evaluation data for the improvement of program operations and outcomes. The evolving service foci reflect the influence of the fervor of the school reform movement and the nudging of federal directives to expend increasingly greater amounts of TAC resources on program improvement. Based on experience in 13 western states and the Pacific territories, this paper traces the evolution of service delivery, identifies the forces that have shaped the evolution, and discusses implications for future TAC work. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED308205

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Kim O. Yap

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TAC Evaluation Assistance: A Ten-Year Evolution from Accountability to Program Improvement

Kim O. Yap

Northwest Regional Educational Laboratory

A paper presented at the annual meeting of the American Educational Research Association, San Francisco, March 27-31, 1989.

TM013415

BEST COPY AVAILABLE

TAC Evaluation Assistance: A Ten-Year Evolution from Accountability to Program Improvement

Kim O. Yap

Northwest Regional Educational Laboratory

101 S.W. Main, Suite 500

Portland, Oregon 97204

ERIC

TAC Evaluation Assistance: A Ten-Year Evolution from Accountability to Program Improvement

Abstract

The Title I/Chapter 1 TACs were originally created to provide assistance to state and local education agencies in implementing the Title I evaluation and reporting system. However, over a 10-year period, TAC role has changed significantly. Its service foci have evolved from providing evaluation assistance to helping clients in the use of evaluation data for the improvement of program operations and outcomes. The evolving service foci reflect the influence of the fervor of the school reform movement and the nudging of federal directives to expend increasingly greater amount of TAC resources on program improvement. Based on experience in 13 western states and the Pacific territories, this paper traces the evolution of service delivery, identifies the forces that have shaped the evolution and discusses implications for future TAC work.

TAC Evaluation Assistance: A Ten-Year Evolution from Accountability to Program Improvement

Introduction

Since the inception of the Title I/Chapter 1 TACs in 1976, a primary function of these centers has been to help state departments of education (SEAs) and local education agencies (LEAs) to improve their evaluation practices. Initially, improvement meant that the LEAs could provide evaluation data which met a set a technical standards and which could be aggregated at the national level and reported to Congress. Later, improvement took on a new dimension; evaluations were also to provide useful information for local districts, particularly for the purpose of program improvement.

The Title I evaluation and reporting system (TIERS), as originally constituted, was not intended to serve the needs of two masters: local school districts and the federal government (Wisler & Anderson, 1979). It was designed to generate Chapter 1 evaluation data in a common metric for national aggregation. It was prescriptive, although not restrictive. It set the minimum standards, but in no way limited the LEA evaluation activities to the TIERS. Indeed, LEAs were urged to do more than what the TIERS entailed -- and many LEAs did. TAC's mission included helping SEA/LEA

clients to move beyond the TIERS to increase the usefulness of evaluation in program planning and improvement.

With these gradual but fundamental changes in TAC role, much of what we do has taken on a new look and new purposes. This paper presents a historical account of TAC evaluation assistance from its formative days in 1976 through the reauthorization of Chapter 1 in 1988. We will do this by describing a set of evolving foci which has encompassed much of TAC work during the past decade in 13 western states and the Pacific territories. We will describe this evolution, discuss its implications, and identify some issues we will face in carrying out future TAC work.

Background

The TIERS was born in an era of accountability. Pre-TIERS Title I efforts were not required to and did not furnish information that could be aggregated to obtain a national picture of what was offered to disadvantaged students and its effects. Following years of frustration (Cross, 1979), Congress in 1974 instructed the U.S. Office of Education to implement a set of standards and criteria which would provide for a common metric to assess the effects of educational services provided to the disadvantaged youth. Under contract with the U.S. Office of Education, the RMC Research Corporation developed a set of evaluation

procedures, later known as the TIERS (Tallmadge & Wood, 1982; Tallmadge, 1982). The system attended to concerns of reliability and validity of evaluation information and, most importantly, offered a common metric, the normal curve equivalent (NCE), to measure the effects of Title I activities across the country.

To ensure the proper use of the system, the U.S. Office of Education established in 1976 regional centers to provide technical assistance to SEAs and LEAs which intended to adopt the system. These centers, now known as the TACs, took on the task of explaining the merits of the TIERS and ways of ensuring a high fidelity implementation of the system in their client states. By 1979, most states had adopted the TIERS and reported NCE data to the U.S. Department of Education (USDE), formerly the U.S. Office of Education.

The mission of the early TACs was simple: to help LEAs and SEAs to provide aggregatable evaluation data to the USDE and Congress. For the first time in over a decade, Congress was able to peruse national data on Title I activities and their impact. For the first time, accountability permeated the entire Title I program, from local districts to the federal level. Congress was pleased, the TIERS survived and the TACs were re-funded.

At about this time, voices of dissatisfaction with the TIERS began to come from various quarters. Some questioned the methodological rigor of the RMC models (Linn, 1979; Jaeger, 1979; Wiley, 1979). Others raised the more practical concern over the usefulness of the TIERS data at the local level (Barnes & Ginsburg, 1979). Among efforts to respond to these voices of concern, the notion of "sustained effects" emerged. Sustained effects studies were to assess the longer-term (over one year) effects of Title I/Chapter 1. But more importantly, they were to make the TIERS more responsive to local needs. The sustained effects component was to afford local project staff the opportunity to conduct studies and use the data to address questions of local interest. This was something project staff would do by themselves and for themselves. There were no requirements to share the information with the SEAs or the USDE. The only mandate was that it be conducted and that the resulting information be used for purposes of program improvement.

At the time the USDE made its third procurement of TAC contracts in 1983, a distinction was made between evaluation and program improvement. A minimum of 25 percent of TAC resources was to be expended on program improvement. This figure has increased to 50 percent for the current contract cycle (1985-88). The enforcement of sustained effects requirements has become more stringent. Local districts are now required to conduct such studies for all grade levels

and all subject areas. At the same time, program improvement has taken on increasingly greater importance. It might well become a mandated activity in the reauthorized Chapter 1 in 1988.

The Evolution

In the sections to follow, we will trace the evolution of TAC work during the past decade. We start with the initial implementation of the TIERS, progressing to the gradual improvement of data quality and utility and, finally, to the use of assessment and evaluation data in program improvement. This evolution came about through the persistent work of TAC staff, helped by the gentle nudging of USDE directives which channeled TAC and client efforts onto the path of program improvement.

Orientation to TIERS

Much of the early TAC work consisted of explaining, demonstrating and selling the TIERS to client states. Workshops on orientation to TIERS typically included an explanation of the merits and demerits of the three RMC models and their respective implementation steps. By 1979, it became obvious that few, if any, of the states would attempt the adoption of the methodologically more rigorous but computationally more complex models. Most clients

abandoned the comparison and regression models following some preliminary exposure to their methodological complexity. Many were concerned that the implementation of the comparison model might invite legal challenge on behalf of disadvantaged youngsters in the comparison group who, by design, would be deprived of Chapter 1 service. It appeared that most LEAs would implement the norm-referenced model because the procedures were most similar to what they had been doing in the past.

From then on, the TIERS workshops focused mostly on the implementation of the norm-referenced model and related issues. Regression to the mean, for example, became a dominant topic of discussion between TAC staff and clients. Separation of pretest from selection test emerged as a standard of high fidelity implementation of the norm-referenced model. It, unfortunately, also became a hindrance to acceptance of the model in LEAs which would like to use the same measure for selection and pretest.

Quality Control

Having sold the norm-referenced model to its clients during the first contract cycle, TAC turned its attention to quality control issues. It became imperative that the LEA clients achieve a high fidelity implementation of the the norm-referenced model. To shore up data quality, TAC began

to offer workshops on test selection, test administration, score conversion and data management. Test selection emphasized the importance of both the psychometric qualities of the instrument and its content validity. A close match (over 75 percent) between test items and instructional objectives became an overriding criterion for test selection. Test administration workshops sought to increase client awareness of the effects of poor test administration on results, particularly in the case of individually administered tests. TAC staff demonstrated proper administrations and developed various checklists to guide teachers through the proper steps.

The conversion of raw scores to scale scores to percentiles and finally to NCEs continued to be a major source of error. Finding the right norms table and then the normative score from the correct row or column continued to tax the understanding and patience of TAC clients. The followup task of managing the large volumes of data on paper or card files was also a challenge. TAC staff provided a large number of "trouble-shooting" workshops and consultation sessions to make TIERS as error-free as possible.

It was at this juncture, in the late 70s, that the microcomputer made its propitious appearance on the education scene. Its use actually mushroomed in classroom

instruction before the technology showed its appeal in program management and evaluation. TAC staff capitalized on the micro-revolution and began advocating the use of high-tech in Title I, particularly in test score conversion and data management. For clients not using the test publishers' scoring service, the same results could be gotten through a home-grown package, developed with the assistance of TAC staff. Finally, there was hope that evaluation need not entail drudgery and tedium to produce accurate results.

Testing Issues

Although neither TAC staff nor clients would equate evaluation with testing, a large portion of TAC resources was devoted to testing issues. Among other things, we offered workshops on out-of-level testing, test equating, test planning and testwiseness. Testing out of level was an issue of particular pertinence to TAC clients, Title I/Chapter 1 youngsters being low-achieving students. It was important to decide whether out-of-level testing should be used and, if used, whether it should be conducted on an individual or group basis. TAC clients needed to know the appropriate criteria for making such decisions and procedures for implementing such decisions.

Some TAC clients considered criterion-referenced tests (CRTs) to be more meritorious than norm-referenced tests

(NRTs). There was a need to afford the opportunity to use CRTs and still be able, ultimately, to report results in NCEs. Test equating became a topic of interest. We offered the equi-percentile equating model for its relative mathematical simplicity and conceptual clarity. Understandably, not many clients found themselves eager to embark on a test equating project. Of those that made the attempt, even fewer completed the task feeling satisfied that the benefits were worth the effort.

Far more widespread was the desire to plan or streamline a districtwide testing program to reduce the burden on both Title I/Chapter 1 and the regular program. In many LEAs, the district administered a testing program which considerably overlapped with a school testing program which in turn duplicated the Title I/Chapter 1 testing requirements. In some cases, the same tests were used, but test administrations occurred at different times for no compelling reasons. In other cases, different tests were in use, again for no compelling reasons. The duplication of effort not only entailed a waste of time and precious resource but also resulted in poor quality or uninterpretable data. In most instances, a streamlined testing program produced considerable cost-savings and higher quality data. In some cases, where political or jurisdictional issues prevented the implementation of a

streamlined testing program, duplicated testing practices have persisted.

Many an LEA has claimed that its evaluation data were invalid by virtue of the fact that its Title I/Chapter 1 students did not know how to take tests. Testwiseness began to loom large as a data quality issue and TAC provided a large number of workshops on the topic to vast numbers of clients. The objective of testwiseness was not necessarily to raise test scores as some of the catchy titles of testwiseness publications would imply. The overall goal was to obtain a more valid measure of achievement and to enable LEAs to implement practices conducive to achieving that goal. TAC offered training to Title I/Chapter 1 project staff on how to prepare students, parents and test administrators so that extraneous variables (e.g., lack of test taking skills) did not exert an undue influence on test scores.

Test Use

Having ensured that clients were in possession of high quality data, TAC now turned its attention to promoting the use of such data. We offered workshops on data interpretation, data graphing and curriculum mapping, among others. There was a continuing need to discuss with clients the meaning of various types of test scores. There was even

a greater need to explain the concept of NCE gain and the use of comparative and quality standards (Messick, 1985) to evaluate the strength of such gains. There was, for example, a need to point out that children in lower grades tended to make greater gains and that children who scored lower on the pretest tended to make greater gains. TAC staff also felt compelled to point out that test scores were not infallible and that there would always be some measurement error. We continually reinforced the notion that test results based on a larger sample of students deserved more of our confidence that those based on a smaller sample.

We offered data graphing as an effective means of promoting data use. Data should appear not just in a tabular format but also in bar graphs, line graphs, pie charts and their fancier variants to get the message across. A picture speaks louder than a thousand words. Profiling a Title I/Chapter 1 class, grade, or school to highlight successes and failures became an important way of sharing program outcomes with parents and other school and community groups.

Curriculum mapping involved the use of item level information to identify strengths and weaknesses in instruction relative to the specified instructional objectives. We trained teachers in item analysis and the

clustering of items, along with their p values, relative to specific instruction objectives. The process highlighted strong and weak areas in instruction so the instructional staff could plan activities to overcome any identified weaknesses. We urged the use of collegial learning, peer coaching and external expertise as means of developing corrective actions.

Improvement-Oriented Assessment

In its steady march toward program improvement, TAC began, in early 1980s, the development of a set of activities which may best be described as improvement-oriented assessment. These activities focused on the assessment of processes linked to effective schooling and provided helpful hints for improving such processes. We offered these activities to clients in the form of workshops bearing such labels as time-on-task, reading diagnosis and instructional planning, writing assessment, higher-order thinking skills, classroom management, parent involvement and cost analysis.

Time-on-task, which later proved to be one of the most popular workshops among TAC clients, drew on research principally conducted by Berliner (1976, 1979) and Stallings (1980). The workshop demonstrated ways of obtaining a picture of time use in the classroom, detailing the

proportion of class time spent on various instructional and non-instructional activities. It discussed ways of interpreting the results and offered helpful hints for remedying any identified misuse of class time.

The workshop on reading diagnosis and instructional planning was developed with the assistance of a reading specialist. The workshop emphasized the use of multiple measures, including the use of informal reading inventories, in diagnosing reading difficulties. We urged the use of flexibility in interpreting the results, particularly in determining reading levels. The rigid use of criteria on word recognition and passage comprehension advocated by various experts may actually lead to mis-diagnosis.

The writing assessment and thinking skills workshops offered ways of assessing these more complex cognitive skills and provided helpful hints on how to improve the teaching of these advanced skills. Writing assessment demonstrated ways of scoring a writing sample holistically to judge its overall quality or analytically to focus on the quality of discrete components (e.g., organization, ideas, wording). In addition, writing samples may be rated on how well they achieve the primary purpose of persuading, informing, or entertaining.

In a classroom setting, thinking skills fall into the categories of recall, analysis, comparison, inference and evaluation. The thinking skills workshop provided a definition of each category and demonstrated ways of assessing these skills through oral questioning during instruction as well as the use of multiple choice items and performance assessment (e.g., composition). We favored the use of an infusion model over a separate curriculum in teaching these skills and emphasized the distinction between learning cognitive skills in school and using such skills in real-life settings (see Resnick, 1987; Sternberg, 1987).

The parent involvement workshop emphasized the importance of parents' role in the instructional process both in school and at home. Parental influence was strongly felt at the advocacy and the decisionmaking levels. We had powerful parent advocates for Title i/Chapter 1 from the LEA to the federal levels. We had parents sitting on various committees that made decisions on program planning, operations and evaluation. Now we need parents who would help provide instruction to their own and other youngsters.

The classroom management workshop addressed what has long been a bane for most teachers: How to maintain an orderly classroom that is conducive to high achievement. The workshop discussed the use of routines and rules and emphasized the importance of <u>teaching</u> them to the students.

It offered advice on the use of "wait-time" -- putting questions to the entire class and waiting up to 3 seconds for an answer -- to hold every student accountable for the answer. It urged the use of quick and mildly unpleasant punishment to save class time and avoid the risk of turning students off permanently.

With limited or declining resources, cost-effectiveness has become a criterion for judging the worth of a program. Given comparable outcomes, a program that costs less is more supportable than one that costs more. The cost-analysis workshop demonstrated ways of determining per-pupil costs and deriving various cost-effect indices. An interesting topic concerned the alluring prospect of using computer-assisted instruction (CAI) or computer-managed instruction (CMI) to serve a greater number of students at a lower cost. Although the use of high-tech in Chapter 1 is still in its infancy, CAI and CMI are undoubtedly here to stay.

Moving closer to program processes, some SEAs and LEAs have begun to embark on process evaluation activities. These efforts generally focus on instructional components, use of personnel and project settings. One LEA, for example, looked at variables affecting the implementation of a mastery learning component. The study revealed that several factors were conducive to high achievement with the use of mastery learning. They included close coordination

with the regular program, staff experience and training in the use of the component and a prolonged period of use of the component. Another LEA examined the use of additional aides to increase time-on-task in teaching mathematics in a learning center setting and found such use to be effective. With TAC assistance, an SEA conducted a secondary analysis of Chapter 1 data covering a period of three years. The analysis was to assess the relative merits of pull-out versus in-class settings, among other things. The findings suggested that the pull-out setting was more expensive but it also showed higher NCE gains. In fact, from a cost-effectiveness standpoint, the pull-out setting was as supportable as, if not more so than, the in-class setting.

Program Improvement

Although program improvement was not an explicit goal of the TIERS, TAC staff has always worked under the assumption that SEA and LEA clients would use evaluation data to improve program operations and outcomes. When USDE first required sustained effects evaluation in 1979, program improvement was very much an integral part of the package. Local project staff were to use the information obtained from such evaluations for program improvement. It was something that local staff could do with a great deal of flexibility so that the resulting information would be useful for program modification and improvement.

A more conspicuous signal that the USDE was serious about program improvement came in 1982, when it provided a set of Secretary's Initiative grants to selected SEAs and LEAs for the development and implementation of program improvement activities. This was followed in 1984 by the initiation of an effort to accord national recognition to "unusually successful" Chapter 1 projects across the nation. This effort, called the National Identification Program, has received an enthusiastic response from the SEAs and LEAs for the past several years. It has completed two successful cycles, each culminating in an awards ceremony at the International Reading Association annual meeting and the preparation of the Chapter 1 Sourcebook as a source of program improvement information for Chapter 1 projects.

The National Identification Program recognizes exemplary projects on the basis on positive outcomes and a set of 13 effective program attributes (Griswold, et al., 1986). Based on recent effective schooling research, the USDE has identified these 13 variables to be particularly pertinent to Chapter 1 projects:

Organizational attributes:

- o Positive school/classroom climate
- o Clear project goals and objectives

- Coordination with the regular school program and other special programs
- Parent and community involvement
- Professional development and training
- Evaluation results used for project improvement
- Strong leadership

Instructional attributes:

- Appropriate instructional materials, methods and approaches
- Maximum use of academic learning time
- High expectations for student learning and behavior
- Regular feedback and reinforcement
- Closely monitored student progress
- Excellence recognized and rewarded

The move toward program improvement became even more entrenched with the award of new TAC contracts in 1983. TAC staff received a directive from the USDE that a minimum of 25 percent of TAC resources was to be devoted to program improvement activities. With the arrival of the current contract cycle in 1986, the resource commitment has increased to 50 percent.

The USDE directives accelerated the already quickening tempo toward program improvement. Following a year or two of exploration and experimentation, we now have three distinguishable approaches to program improvement:

First, the improvement-oriented assessment has continued to be in demand and is becoming an effective way of introducing significant changes in the Chapter 1 instructional process. For example, following the time-on-task workshop, many LEAs have examined their own pattern of time use and have made changes in the way their instructional staff use class time. The thinking skills workshop has led to increased awareness of the importance of including all categories of cognitive skills (not just recall) in teaching. The curriculum mapping workshop has enabled local project staff to use test information to identify strong and weak points in the instructional process and to initiate activities to remedy identified weaknesses. In many cases, the practice has been incorporated as part of routine program operations. The conduct of process evaluation will undoubtedly continue to provide information on program processes in support of program improvement activities. We hope that these improvement-oriented assessment activities will build up to a critical mass and significantly enhance the quality of instruction in Chapter 1.

Second, following a spate of SEA and LEA initiatives, a variety of quality monitoring systems has been developed with input from the district and school level staff. The monitoring team typically consists of a cross-section of the Chapter 1 staff. Quality monitoring is prompted by poor evaluation results or an intrinsic desire to improve the operations and/or outcomes of a project. Its key components include:

- Preparation of a site visit plan
- Onsite interviews with project staff
- Classroom observations
- Written summaries of findings
- Written recommendations for change
- Preparation of a plan for implementing corrective actions
- Followup site visits

Onsite interviews and observations are aided by a package of instruments including interview guides, observation schedules and a variety of checklists and rating scales. Many of these instruments address the 13 effective program attributes identified by the USDE. Members of the monitoring team receive training in the use of these instruments prior to conducting the site visit.

Third, taking advantage of enthusiasm generated by the recent school reform movement, TAC is promoting a systemic process for program improvement. The process is rooted in a broad research base encompassing school effects, teacher effects, instructional leadership, curriculum, program coupling and implementation (Berman & McLaughlin, 1974; Weick, 1976; Berman, 1980; Fullan, et al., 1980; Edmonds, 1982; Purkey & Smith, 1982; Cooley, 1983; Mckenzie, 1983; Purkey & Smith, 1983; Pratzner, 1984).

Operationally, the process consists of the following steps:

- Form a leadership team
- Review relevant research
- Develop project profiles
- Set improvement goals
- Develop prescriptions
- Plan for implementation
- Monitor implementation
- Evaluate and renew improvement effort

The implementation of these steps is research guided, data based, and built upon the assumption that shared decisionmaking, collegial learning and a sense of ownership among all participants are essential ingredients of a workable and effective program improvement effort.

The leadership team typically consists of the building administrator, key project staff and an LEA/SEA staff. The team reviews the relevant research and gathers data to develop a project profile to identify strengths and weaknesses in student performance, including academic achievement, attitudes and behavior. The team then shares the findings with the entire project staff and prepares goal statements to specify areas where improvement is deemed necessary. In prescription development, the leadership team examines the effective schools research base and seeks input from the entire staff to develop instructional methods and techniques for use by all staff during the improvement cycle. In consultation with all staff, the team then prepares specific plans for implementation. The implementation of prescriptions is closely monitored and evaluated to ensure that new practices are being implemented and are leading to the desired outcomes.

The implementation of an improvement cycle, which typically takes a year or two, may be accomplished through a set of carefully sequenced workshop sessions, interspersed with tasks to be completed by clients following each workshop. Alternatively, TAC staff may take clients through the process without a fixed schedule of task completion to allow clients more flexibility in adjusting to various reality constraints. Among improvement cycles being

implemented, many are addressing the 13 effective program attributes. Popular areas include parent involvement, coordination and instruction.

Implications

The progression from accountability to program improvement is certainly congruent with congressional intent as reflected in the pending legislation for Chapter 1 reauthorization. The new legislation is likely to put heavy emphasis on program improvement and perhaps even mandate the use of TAC services in such activities. To continue to serve as an effective resource for evaluation and improvement, our service delivery system will need to accommodate several unmistakable trends.

1. More and more, project schools rather than the SEAs will become the primary beneficiary of TAC service. Although the SEAs will remain the primary contact point in our service delivery system, schools will likely be the primary recipient of TAC service. TAC work will address school level concerns such as instructional strategies, inservice training for instructional staff, coordination w[illegible]gular pr[illegible]ram and school level evaluation results. With a great number of Chapter 1 schools in each of our client states, we face the urgent need to agree to a set of priorities which will maximize the impact of our services.

2. With the TIERS taking hold in most states, the time has come for us to devote resources to the measurement of program processes as well as outcomes. TAC has worked past the macro stage of summative evaluation (Scriven, 1979) and entered the micro stage of formative evaluation and program improvement. To aid the program improvement process, we need to come up with procedures for assessing such variables as quality of instruction, appropriate level of instruction, incentive for learning and use of class time.

3. In the early days, TAC staff served as outside experts. Contacts with clients were formal, short-term and discrete. Over the years, the working relationships have become much more collaborative, enabling TAC staff to contextualize evaluation information and provide specific suggestions instead of general recommendations. We are attending more to the potential use of evaluation and less to the methodological sophistication of an evaluation's design features. There is a lesser need to build credibility on our technical expertise. There is a greater need to maximize our helpfulness in the potential use of evaluation information by clients.

4. Within the context of use, no amount of methodological sophistication will provide the implicit knowledge which only professional consultation and judgment can muster.

For example, we need to know something about the local management style. We need to know whether a school administrator is an initiator, manager, or responder (Hord, et al., 1984). We should strive to be more practical than technical. Sophisticated information does not get used if it is not also practical (David, 1981).

5. In an effort to clearly define TAC role and to preserve the integrity of TAC work, the USDE has placed a number of restrictions on our scope of work. In particular, the restriction against advising clients on curricular and instructional matters will continue to hinder some aspects of our work. With increased emphasis on program improvement, it appears necessary that this restriction be removed from the scope of work so that TAC staff have maximum flexibility in providing program improvement assistance to clients.

Issues

A decade of working with SEA and LEA personnel has enabled TAC to resolve some critical issues facing Chapter 1 programs. By and large, Congress is getting what it requested and the TIERS has remained a popular system for evaluating the impact of Chapter 1. Our work, on the other hand, has also surfaced some thorny issues. Some examples follow.

1. The TIERS was designed primarily to produce aggregatable data for reporting to Congress. It sidestepped such complex issues as what constitutes appropriate criteria of success for Chapter 1. Most researchers would advocate the use of multiple measures. The single measure of NCE gain is perhaps convenient for reporting purposes. One could, however, think of other criteria which are just as pertinent. For example, what about the quality of regular school work of Chapter 1 participants? What about the way students feel about class work and school in general? What about mainstreaming? Surely, a primary goal of any compensatory education program is to "graduate" students who are able to function effectively in the regular school setting.

2. Even if we restrict ourselves to NCE gains, it is certainly debatable whether fall-spring or annual (e.g.,spring-spring) gains are the better or more accurate measures of achievement. The annual cycle has the obvious advantage of reducing testing burden. It offers a stronger safeguard against extraneous factors (e.g., test administration) influencing the size of achievement gains by virtue of the fact that posttest data for any given year will be used as pretest data for the following year. Furthermore, unlike fall-spring gains, annual gains are cumulative over the years. On the other hand, if we follow

the premise that data based on a larger number of students deserve more of our confidence than data based on a smaller number of students, surely we should have more faith in fall-spring results than spring-spring results. Spring-spring testing entails a much higher degree of attrition resulting from a wide array of anticipated and unanticipated circumstances. Attrition occurs when we have transient student populations; when a testing program allows or requires the use of a different test from one school year to the next; and when the school desegregation schedule requires the busing of students at certain grades. Recently, an LEA reported that these factors reduced the proportion of students with both pretest and posttest scores to less than 15 percent of the Chapter 1 enrollment.

3. The use of test norms as standards for judging program impact is not as clear-cut as it might seem -- even if we accept the equi-percentile assumption. Fall-spring norms are essentially cross-sectional norms. Since the derivation of these norms typically does not involve the matching of individual students from fall to spring, they are not longitudinal norms. This is also true of spring-spring norms. The manner in which the TIERS uses test norms implies that fall-spring and spring-spring norms are truly longitudinal norms based on the same individual students. It is tempting to ask how much of the summer dip in Chapter

1 achievement is a result of using cross-sectional norms as longitudinal norms in measuring achievement growth.

4. In providing program improvement assistance, we need to examine the overall goal/priority structure of an educational system, from the SEA to the LEA to the Chapter 1 project. A congruent goal/priority structure is built on a set of criteria of success that is agreed upon by all key players. Yet, incongruent goal structures often exist, making the setting of goals for program improvement at the project level very difficult. For example, the LEA goal of raising the performance level of a majority, say 60 percent, of its students to the 50 percentile on a standardized test may be reasonable and attainable. Yet, applying the same goal to a Chapter 1 project would be a bit too ambitious, if not unrealistic. When a data-based goal setting process is pre-empted or superseded by an administrative fait accompli, the entire program improvement effort might be short-circuited and doomed for failure. An incongruent priority structure can lead to a proliferation of innovations some of which might duplicate or compete with each other, greatly reducing the impact of the innovations. For example, an LEA or SEA may make the improvement of self-esteem its top priority when the Chapter 1 program's greatest need lies in basic skills instruction. In the worst scenario, we see an "innovation overload" at the receiving end of the reform process -- at the school and classroom level. Faced with

the impossible task of juggling a plethora of "projects," school level staff often find it most sensible to orchestrate a perfunctory implementation or a downright trashing of the innovations.

5. We need to weigh the relative importance of profession judgment against empirical data. That TIERS data play a critical role in program improvement activities is perhaps the greatest pleasant surprise arising out of a decade of TAC work. This is a far cry from earlier years when lament over the lack of data use in Chapter 1 was commonplace (David, 1981). In most instances, it is the impact data that trigger a desire or need to initiate an improvement effort. These data, however objective and valid, must be tempered with the professional judgment of project staff. The popular axiom notwithstanding, data do not usually speak for themselves. They need to be interpreted and put in the context of program operations. One of the reasons why evaluation data do not get used is that they remain in a decontextualized form untempered with the crucial and often implicit knowledge of program operations available only to project staff. The critical role that implicit knowledge plays in attaining success, by an individual or a program, is receiving increasingly more attention in research on intellectual skills (Sternberg, 1987). Empirical data, often short-term and precise, need to be fitted in a reality context which is usually fluid and dynamic involving

countless situational specific variables. These variables co-exist and interact with one another sequentially and simultaneously. Empirical data provide only a "slice-of-life" perspective. Implicit knowledge, on the other hand, is often long-term and able to provide insights on the simultaneous interactions of variables. A successful implementation of a program improvement effort calls for the interplay of empirical data and professional judgment. A balanced use of these two ingredients makes a good recipe for program improvement.

6. Many researchers of effective schooling (e.g., Cooley, 1983; Purkey & Smith, 1983) advocate the use of schools as the unit of change. A school is said to have a culture, climate or zeitgeist. The school building, being only loosely coupled with the LEA and SEA administration, has its own leadership style and its own pattern of resource use. An improvement effort that is not schoolwide lacks the critical mass in resource and incentive to move it along. Yet, research also shows that most improvement efforts focus on one or two subject areas, one or two grade levels, one or two aspects of school operations (Rowan, et al., 1983). By and large, improvement occurs on an incremental basis. The dilemma facing Chapter 1 projects stems from the fact that they typically represent less than 10 percent of the overall school operations. How strong a commitment can we expect from a school administrator when the entire effort is

directed to less than 10 percent of the school enrollment? Is Chapter 1 program improvement doomed to fail because it lacks a schoolwide momentum? Are there ways of making it a schoolwide effort by involving non-Chapter 1 personnel? Would such an effort raise compliance issues?

7. Finally, there is the issue of resource, time and incentive. Most existing improvement activities do not entail a substantial, if any, increase in funding. Out of a strong sense of commitment and dedication, school or project staff volunteer their time and energy to carry out the various improvement tasks. There is very little incentive other than the professional satisfaction of seeing things improve. Yet, our experience suggests that time has been the primary obstacle to implementing a program improvement effort. More than financial resource and incentive, project staff find themselves in need of more time to plan and implement new practices -- time that they cannot find in their already busy work schedule. The struggle to find time can undoubtedly be eased if additional resources and/or personnel are made available to help implement improvement activities. Renewed priorities and commitment at the LEA and SEA administrative levels can do wonders in finding time and resources that hitherto have been unavailable.

References

Barnes, R. E., & Ginsburg, A. L. (1979). Relevance of the RMC models for Title 1 policy concerns. Educational Evaluation and Policy Analysis, 1 (2), 7-14.

Berliner, D. C. (1976). A status report on the study of teacher effectiveness. Journal of Research in Science Teaching, 13, 369-382.

Berliner, D. C. (1979). Tempus Educare. In P. Peterson & H. Walberg (Eds.), Research in Teaching. Berkeley: McCutchan.

Berman, P. (1980). Toward an implementation paradigm of educational change. Santa Monica, CA: Rand Corporation.

Berman, P., & McLaughlin, M. W. (1974). Federal programs supporting educational change. Volume I: A model of educational change. Santa Monica, CA: Rand Corporation.

Cooley, W. W. (1983). Improving the performance of an educational system. Educational Researcher, 12 (6), 4-12.

Cross, C. T. (1979). Title 1 evaluation -- A case study in congressional frustration. Educational Evaluation and Policy Analysis, 1 (2), 15-21.

David, J. L. (1981). Local uses of Title 1 evaluations. Educational Evaluation and Policy Analysis, 3 (1), 27-39.

Edmonds, R. R. (1982). Programs of school improvement: An overview. Educational Leadership, 40 (3), 4-11.

Fullan, M., Miles, M. B., & Taylor, G. (1980). Organization development in schools: The state of the art. Review of

Educational Research, 50, 121-183.

Griswold, P A., Cotton, K. J., & Hansen, J. B. (1986). Effective compensatory education sourcebook. Washington, D. C.: U.S. Department of Education.

Hord, S., Stiegelbauer, S., & Hall, G. (1984). Principals don't do it alone: Researchers discover second change facilitator active in school improvement efforts. The Newsletter of the Research & Development Center for Teacher Education, Vol. II, No. 3.

Jaeger, R. M. (1979). The effect of test selection on Title 1 project impact. Educational Evaluation and Policy Analysis, 1 (2), 33-40.

Linn, R. L. (1979). Validity of inferences based on the proposed Title 1 evaluation models. Educational Evaluation and Policy Analysis, 1 (2), 23-32.

Mckenzie, D. E. (1983). Research for school improvement: An appraisal of some recent trends. Educational Researcher, 12 (4), 5-17.

Messick, S. (1985). Progress toward standards as standards for progress: A potential role for NAEP. Educational Measurement Issues and Practice, 4 (4), 16-19.

Pratzner, F. C. (1984). Quality of school life: Foundation for improvement. Educational Researcher, 13 (3), 20-25.

Purkey, S. C., & Smith, M. S. (1982). Too soon to cheer? Synthesis of research on effective schools. Educational Leadership, 40 (3), 64-69.

Purkey, S. C., & Smith, M. S. (1983). Effective schools --

A review. Elementary School Journal, 83 (4), 427-452.

Resnick, L. B. (1987). Learning in school and out. Educational Researcher, 16 (9), 13-20.

Rowan, B., Bossert, S. T., & Dwyer, D. C. (1983). Research on effective schools: A cautionary note. Educational Researcher, 12 (4), 24-31.

Scriven, M. (1979). Viewpoints on education evaluation. Educational Evaluation and Policy Analysis, 1 (2), 66-72.

Stallings, J. A. (1980). Allocated academic learning time revisited, or beyond time on task. Educational Researcher, 9 (11), 11-16.

Sternberg, R. J. (1987). Teaching critical thinking: Bridging the gap between classroom instruction and everyday life. A presentation at the Oregon Educational Research Association annual meeting in Welches, Oregon.

Tallmadge, G. K. (1982). An empirical assessment of norm-referenced evaluation methodology. Journal of Educational Measurement, 19 (2), 97-112.

Tallmadge, G. K., & Wood, C. T. (1982). User's guide ESEA Title I evaluation and reporting system. Mountain View, CA: RMC Research Corporation.

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

Wiley, D. E. (1979). Evaluation by aggregation: Social and methodological biases. Educational Evaluation and Policy Analysis, 1 (2), 41-45.

Wisler, C. E., & Anderson, J. K. (1979). Designing a Title 1

evaluation system to meet legislative requirements. Educational Evaluation and Policy Analysis, 1 (2), 47-55.