

## DOCUMENT RESUME

03404 - [A2553678]

Problems and Needed Improvements in Evaluating Office of Education Programs. HRD-76-165; B-164031(1). September 8, 1977. 76 pp. + 4 appendices (53 pp.).

Report to the Congress; by Elmer B. Staats, Comptroller General.

Contact: Human Resources Div.

Budget Function: Education, Manpower, and Social Services: Elementary, Secondary, and Vocational Education (501).

Organization Concerned: Office of Education; Department of Health, Education, and Welfare.

Congressional Relevance: House Committee on Education and Labor; Senate Committee on Human Resources; Congress.

Authority: Elementary and Secondary Education Act of 1965 (20 U.S.C. 241a). Education Amendments of 1974 (P.L. 93-380).

General Education Provisions Act (20 U.S.C. 1226c).

Education Amendments of 1972 (P.L. 92-318). 20 U.S.C. 1231a.

A review was conducted to determine the effectiveness of federally supported education evaluations, primarily those concerning elementary and secondary education programs, in order to obtain objective data for allocating resources and in deciding whether or not programs should be continued or modified. Questionnaires were sent to education agencies in all States and the District of Columbia and to a statistical sample of local school districts to obtain State and local agencies' views on Federal education program evaluations. Federal, State, and local education agencies frequently use standardized norm-referenced achievement tests to measure the effect of Federal education programs. Findings/Conclusions: The Office of Education's (OE) evaluation studies can better serve Congress by having them timed to coincide with the legislative cycle and by more frequent briefings of congressional committee staffs. OE needs to make a better effort to set forth specific qualitative and quantitative program objectives in order to provide a clear basis for program evaluation. The usefulness of the State and local evaluation reports needs improvements in the areas of: relevance of reports to policy issues, data completeness and comparability, and report timeliness. If the reporting systems based on aggregated local agency data are to be effective, standardization of data collection efforts is needed. Educators and test experts disagree on the use of standardized norm-referenced tests versus criterion-referenced tests. More research may be needed on criterion-referenced tests and on how to reduce racial, sexual, and cultural biases in standardized tests. Recommendations: The Secretary of HEW should direct OE to: (1) emphasize congressional information needs when planning, implementing, and reporting on evaluation studies; (2) seek agreement with Congress on the specific program objectives to be used for evaluations as well as acceptable evaluation data and measures for each program to be evaluated; and (3) improve the

implementation of evaluation results by giving greater attention and priority to procedures such as the issuance of policy implication memoranda designed to assure implementation of those results. OE should review the types of State and/or local program evaluation information collected on programs authorized by titles I and VII of the Elementary and Secondary Education Act to determine if it is realistic to serve Federal, State, and local levels with aggregated data based on local agency evaluation reports. (SW)

# REPORT TO THE CONGRESS

BY THE COMPTROLLER GENERAL  
OF THE UNITED STATES

---

## Problems And Needed Improve- ments In Evaluating Office Of Education Programs

Office of Education  
Department of Health, Education, and Welfare

The Office of Education should more strongly emphasize serving congressional needs in planning and carrying out evaluation studies, should define its program objectives more clearly, and should improve the implementation of evaluation results.

The Office of Education should assess whether State and/or local evaluation reports, under titles I and VII of the Elementary and Secondary Education Act, can realistically be improved to supply Federal, State, and local officials with the reliable program information they need for decisionmaking.

Serious questions have been raised about standardized "norm-referenced" tests, which are frequently used to measure Federal education programs' effectiveness. This report discusses some of these questions and various suggestions for alleviating problems in conducting large-scale evaluations of compensatory education and desegregation programs.

03404





COMPTROLLER GENERAL OF THE UNITED STATES  
WASHINGTON, D.C. 20548

B-164031(1)

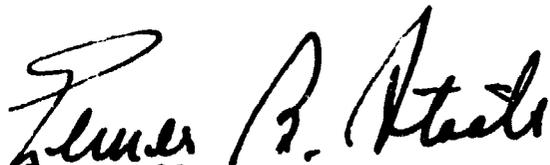
To the President of the Senate and the  
Speaker of the House of Representatives

This report points out that the Office of Education needs to more strongly emphasize the purpose of providing information to the Congress when planning and implementing evaluation studies, more clearly define its program objectives, and improve implementation of evaluation results. The report contains recommendations to improve the usefulness of educational evaluations. Also discussed in the report are some criticisms and defenses of standardized tests and suggestions for alleviating problems in evaluating large-scale compensatory education and desegregation programs.

Our review was made because of the increasing concern shown by the Congress and various Federal agencies with evaluating the effectiveness of major Federal education programs. This concern stems from the need for objective data to be used in allocating resources and in deciding whether or not programs should be continued, modified, or discontinued.

We made our review pursuant to the Budget and Accounting Act, 1921 (31 U.S.C. 53), and the Accounting and Auditing Act of 1950 (31 U.S.C. 67).

We are sending copies of this report to the Director, Office of Management and Budget, and the Secretary of Health, Education, and Welfare.

  
Comptroller General  
of the United States

COMPTROLLER GENERAL'S  
REPORT TO THE CONGRESS

PROBLEMS AND NEEDED  
IMPROVEMENTS IN EVALUATING  
OFFICE OF EDUCATION PROGRAMS  
Office of Education  
Department of Health, Education,  
and Welfare

D I G E S T

In recent years the Congress and various Federal agencies have become increasingly interested in evaluating the effectiveness of major Federal education programs. Such evaluations should help them decide whether programs should be continued, modified, or discontinued. (See p. 1.)

To obtain State and local agencies' views on Federal education program evaluations and related matters, GAO sent questionnaires to education agencies in all States and the District of Columbia and to a statistical sample of local school districts throughout the Nation. (See p. 11.)

This report discusses the questionnaire results, the pros and cons of standardized tests, and suggestions from testing and evaluation experts and others for alleviating problems in conducting large-scale evaluations of compensatory education and desegregation programs. (See pp. 53 and 63.)

OFFICE OF EDUCATION  
EVALUATION STUDIES

Office of Education evaluations are to help those making decisions about education, including the Congress. Recently, the main emphasis has been on evaluating the effectiveness of major Office of Education programs throughout the Nation.

The Office of Education's evaluation studies can better serve the Congress. Timing the studies to coincide with the legislative cycle and more frequently briefing congressional committee staff would help.

Evaluations can reach conclusive and useful findings about a program's effectiveness only if the program's objectives are defined and measurable. For example, a national evaluation of the migrant program did not adequately assess the program's effectiveness because no one had developed acceptable criteria and objectives for measuring the program's success.

Program objectives to be evaluated need to be better defined, and the ways evaluation results are used and implemented need more attention.

The Congress should recognize that the Department of Health, Education, and Welfare (HEW) is not complying, and does not intend to comply, with the legislative requirement to set forth goals and specific objectives for individual programs and projects included in its annual evaluation report to the House Committee on Education and Labor and the Senate Committee on Human Resources.

HEW feels that its authority and ability to comply with this legislative requirement are limited.

#### PROBLEMS WITH STATE AND LOCAL EVALUATION REPORTS

The money spent at the State and local education agency levels to evaluate selected Federal elementary and secondary education programs is large--over \$42 million in fiscal year 1974.

If the reporting systems based on aggregated local agency data are to be effective, various aspects of the evaluation reports need to be improved. These include the credibility of findings and the qualification and quantification of measurement data.

The usefulness of the State and local evaluation reports also needs to be improved with respect to

- relevance of the reports to policy issues,
- completeness and comparability of the data reported, and
- report timeliness.

Valid, complete, and comparable evaluation data is important if evaluation results are to meaningfully contribute to decisions at all levels. If local and State evaluation data continues to be aggregated for use at higher levels, data collection efforts and techniques need to be standardized to provide comparable results.

Because of constraints on the Federal role in education, the Government probably will not try to provide needed valid and comparable data by dictating uniform evaluation methods and procedures to State and local education agency grantees.

In addition, questions exist about whether the models for evaluating title I (aid for disadvantaged children) will be able to provide valid and acceptable data to meet program information needs. If these issues are not resolved, GAO questions whether improvements can be made that will enable the reporting systems (based on aggregated local data) to meet program information needs at Federal, local, and/or State levels.

#### USES OF STANDARDIZED TESTS AND PROGRAM EVALUATION

Federal, State, and local education agencies frequently use standardized norm-referenced achievement tests to measure the effectiveness of Federal education programs. These tests measure an individual's performance against a "norm" group.

However, testing experts and educators disagree about the adequacy of these tests for their intended uses. Serious questions have been raised about the tests, and some organizations have called for a moratorium

on their use and a higher priority on development and use of alternatives.

Although the tests' critics and defenders agree that certain problems exist, views differ greatly about the importance or severity of the problems, and their remedies. Defenders recognize that improvements are needed in such areas as test and test question bias; appropriate test norms; test selection, interpretation, and administration, including test uses for program evaluation; and other issues.

Some questions about the tests have great importance in determining the appropriateness, validity, and proper conduct of educational program evaluations, including those that are federally funded. Those responsible for making educational decisions should be aware of these issues when using such information.

Additional research may be needed on:

- Criterion-referenced and other tests for uses which include program evaluation, as alternatives to standardized norm-referenced achievement tests.
- How to reduce racial, sexual, and cultural biases in standardized tests.

States and especially local education agencies need to be more aware of available information intended to help them select the most appropriate tests for evaluation.

#### EVIDENCE OF PROGRAM EFFECTIVENESS

Local and State evaluation reports on Federal elementary and secondary education program effectiveness are intended to provide information that local, State, and Federal officials can use to make policy and program decisions. However, State and local officials see important differences in the types of evidence of program effectiveness that they and officials at other levels prefer.

Better communication is needed among the three levels on information they need to facilitate policy and program decisions. Questionnaire results raise this question: Should all three levels be served by a reporting system based on the same reports?

Although State officials view Office of Education program officials as being most impressed by standardized norm-referenced test results, and local officials view State and Office of Education officials in the same manner, State and local officials say that they are not most impressed by such results.

Local officials prefer broader, more diverse types of information on program results than just these test scores, and they are most impressed by improvements in curriculum and instructional methods and gains in the affective domain (likes, dislikes, interests, attitudes, motives, etc.). State officials are most impressed by results from criterion-referenced tests.

The widespread use of standardized norm-referenced tests to evaluate State and local programs indicates that State and local officials have more frequently based their evaluations on the kinds of results they believe would be likely to most impress higher level officials than on their own preferences.

Although HEW's comments were not responsive to certain aspects of GAO's recommendations, it agreed, at least in general, with all but one recommendation. (See pp. 21, 37, 52, and 76.)

# C o n t e n t s

		<u>Page</u>
DIGEST		i
CHAPTER		
1	INTRODUCTION	1
	Origin and development of Federal evaluation efforts	1
	Administration of Federal evaluation activities	2
	Requirements for State and local evaluations of OE programs	5
	Limitations of educational evaluations	6
2	SCOPE AND METHOD OF REVIEW	11
3	OPPORTUNITIES TO IMPROVE OE'S EVALUATION STUDIES	14
	Congressional needs should be taken more into account	14
	Program objectives need to be defined	15
	Use of evaluation results needs more attention	17
	Conclusions	20
	Recommendations to the Secretary of HEW	20
	Agency comments and our evaluation	21
	Recommendation to the Congress	23
4	PROBLEMS WITH STATE AND LOCAL EVALUATION REPORTS	24
	State and local evaluation expenditures	24
	State and local evaluation reports need improvement	26
	Other problems	30
	Conclusions	36
	Recommendation to the Secretary of HEW	37
	Agency comments and our evaluation	37
5	USING STANDARDIZED TESTS	42
	Uses and implications of standardized norm-referenced tests	42
	Widespread State and local use of standardized norm-referenced tests to evaluate Federal programs	43
	Efforts to evaluate standardized tests	45
	Criterion-referenced tests	48
	Conclusions	51
	Recommendations to the Secretary of HEW	51
	Agency comments and our evaluation	52

		<u>Page</u>
CHAPTER		
6	STANDARDIZED TESTS AND PROGRAM EVALUATION	53
	Criticism of standardized norm-	
	referenced tests	53
	Defense of standardized tests	61
	Conference on using tests to evaluate	
	programs	63
	Conclusions	67
7	PREFERENCES FOR EVIDENCE OF PROGRAM	
	EFFECTIVENESS DIFFER	69
	State and local education officials	
	believe standardized test results	
	most impress higher level officials	70
	Conclusions	75
	Recommendation to the Secretary of HEW	76

#### APPENDIX

I	Results of GAO's State education agency	
	questionnaire	77
II	Results of GAO's local education agency	
	questionnaire	93
III	Letter dated June 15, 1977, from the	
	Inspector General, HEW	116
IV	Additional suggestions by OE conferees	
	for improving testing and evaluation	125
V	Principal officials of the Department of	
	Health, Education, and Welfare responsible	
	for activities discussed in this report	128

#### ABBREVIATIONS

GAO	General Accounting Office
HEW	Department of Health, Education, and Welfare
NIE	National Institute of Education
OE	Office of Education

## CHAPTER 1

### INTRODUCTION

#### ORIGIN AND DEVELOPMENT OF FEDERAL EVALUATION EFFORTS

In recent years the Congress and various Federal agencies have become increasingly concerned with evaluating the effectiveness of major Federal education programs. Because of tight budget constraints, limited financial and human resources, and the need for objective data on the effectiveness of education programs, the Congress, the Office of Management and Budget, officials of the Department of Health, Education, and Welfare (HEW), and others are requesting such data to help them allocate resources and decide whether or not programs should be continued, modified, or discontinued.

#### Legislative background

Legislative mandates for evaluation of education programs can be traced back at least to title I of the Elementary and Secondary Education Act of 1965 (20 U.S.C. 241a), which required States to assure the adoption of

"\* \* \* effective procedures, including provision for appropriate objective measurements of educational achievement \* \* \* for evaluating at least annually the effectiveness of the programs in meeting the special educational needs of educationally deprived children."

Congressional interest in evaluation was further reflected in the Education Amendments of 1974 (Public Law 93-380). The Office of Education states that 22 new studies and reports are required to be submitted to the Congress by the Commissioner of Education, the Assistant Secretary for Education, or the Secretary of HEW. Of the 22, 7 are evaluation studies to be conducted by the Office of Education's Office of Planning, Budgeting, and Evaluation; another is a thorough evaluation and study of Federal title I and State compensatory education programs by the National Institute of Education (NIE).

The General Education Provisions Act (20 U.S.C. 1226c) requires that HEW provide the House Committee on Education and Labor and the Senate Committee on Human Resources (previously the Senate Committee on Labor and Public Welfare) an annual evaluation report "which evaluates the effectiveness of applicable programs in achieving their legislated

purposes" and recommends improvements. Any evaluation report evaluating specific programs and projects is required to

- set forth goals and specific objectives in qualitative and quantitative terms for all programs and projects and relate those goals and objectives to program purposes;
- report on the progress made during the previous fiscal year in achieving such goals and objectives;
- describe the cost and benefits of each program evaluated; and
- contain plans for implementing corrective action and legislative recommendations, where warranted.

In addition to an increasing number of congressionally mandated national studies, legislation and HEW regulations often require local and/or State evaluations of program effectiveness at least once a year.

The Education Amendments of 1972 (Public Law 92-318), established NIE as part of HEW's Education Division. The Director of NIE reports to the Secretary of HEW through the Assistant Secretary for Education, as does the Office of Education (OE). NIE is charged with improving education by, among other things, building an effective educational research and development system. Since educational research includes not only basic and applied research and surveys, but also evaluation, the law provides a new mechanism for evaluating educational programs. However, principal responsibility for evaluating OE programs remains with OE.

#### ADMINISTRATION OF FEDERAL EVALUATION ACTIVITIES

According to OE, the main goal of Federal educational evaluation studies is to provide information on which policy decisions about Federal education programs and OE resource allocations may be based. To achieve this goal, OE

- conducts national evaluations of the effectiveness of Federal education programs;
- analyzes major educational problems or issues;
- reports annually to the Congress on the effectiveness of OE programs in meeting their legislative intent;

--attempts to identify the program approaches that work and do not work, and determine why; and

--attempts to identify and validate for dissemination locally initiated innovative practices and exemplary programs.

HEW's education evaluations are carried out by several entities: the Office of the Assistant Secretary for Planning and Evaluation; OE's Office of Planning, Budgeting, and Evaluation; a limited number of OE program bureaus; the National Advisory Councils; and NIE. Although the Office of the Assistant Secretary for Planning and Evaluation primarily reviews OE's evaluation activities, it also receives a small portion of education evaluation funds to conduct studies for the Secretary.

OE is among the few Federal agencies that attempt to integrate the program evaluation process with their budget and legislative cycle. OE's evaluation activities are largely centralized in the Office of Planning, Budgeting, and Evaluation. This was done to try to emphasize evaluation more strongly.

The table below lists the funds available to OE for planning and evaluation. According to OE, these sums, although substantial, represent less than three-tenths of 1 percent of OE's total annual program appropriations and must cover approximately 85 legislative programs. OE's Assistant Commissioner for Planning, Budgeting, and Evaluation estimated that from about 1971 on, approximately two-thirds of the OE planning and evaluation appropriation funds have been used for OE evaluation activities. (Chapter 4 provides funding information on State- and local-level evaluations of elementary and secondary education programs.

OE Planning and Evaluation Funds

<u>Fiscal year</u>	<u>OE planning and evaluation appropriations</u>	<u>OE program funds used for evaluation (notes a and b)</u>	<u>Total</u>
----- (000 omitted) -----			
1968	\$ 1,250	-	\$ 1,250
1969	1,250	-	1,250
c/1970	9,512	\$ 4,155	13,667
d/1971	12,475	8,724	21,199
d/1972	11,225	3,950	15,175
d/1973	10,205	9,880	20,085
d/1974	5,200	5,268	10,468
d/1975	6,858	11,043	17,901
1976	6,383	10,512	16,895

a/Includes funds authorized from Follow Through, Emergency School Assistance Act, title I of the Elementary and Secondary Education Act, Basic Opportunity Grants, Project Information Packages, and Career Education programs.

b/Does not include program funds used by State and local education agencies for evaluations under Elementary and Secondary Education Act, titles I, III, VII, and VIII.

c/Does not include \$5 million appropriated for grants to States for planning and evaluation under Elementary and Secondary Education Act, title V, part C--Comprehensive Educational Planning and Evaluation.

d/Includes support for the Educational Policy Research Centers (at Stanford Research Institute and Syracuse University Research Center) for the following fiscal years: \$900,000 (1971); \$900,000 (1972); \$950,000 (1973); and \$450,000 (1974). Monitorship of the centers was transferred to the Office of the Assistant Secretary for Education in fiscal year 1974.

e/Excludes \$1 million earmarked for NIE planning.

Systematic, comprehensive evaluation of Federal education programs at the Federal level dates back only to 1970. At that time the Congress increased OE evaluation funds in response to HEW's request. According to OE, such efforts were largely precluded before then by insufficient appropriated funds for evaluation and too few technically qualified

evaluation staff members. Since fiscal year 1970, OE has attempted to expand and upgrade its evaluation activities and capabilities. The equivalent of about 23 professional full-time staff members are now assigned to evaluation.

The Office of Planning, Budgeting, and Evaluation has designed and begun over 100 evaluation and planning studies; instituted an annual evaluation plan highlighting yearly priorities; and implemented a process for disseminating, chiefly at the Federal level, the major results of evaluation studies.

Almost all OE evaluation and planning studies are performed under contract. OE's evaluation office issues a request for proposals after determining the study's design and the techniques to be used--for example, sample size, analysis method, and data collection method. Contractors are selected competitively. After a contract is awarded, an OE project monitor from the evaluation office monitors the contractor's performance by exercising approval over the approach to be used, making site visits, and reviewing progress reports. The project monitor also reviews and approves the draft report's technical adequacy, completeness, and responsiveness before the report is finally accepted.

OE develops and implements policy recommendations on the basis of the evaluation findings.

#### REQUIREMENTS FOR STATE AND LOCAL EVALUATIONS OF OE PROGRAMS

Legislation and HEW regulations often require annual evaluations of Federal programs by State and/or local education agencies. This is the case for programs and projects funded under titles I, III, and VII of the Elementary and Secondary Education Act. A brief summary of the provisions in each title follows:

- Title I provides funds through State education agencies to local education agencies serving areas with concentrations of children from low income families. The funds are intended to meet the special educational needs of educationally deprived children.
- Title III has provided funds to local education agencies, principally through State education agencies, for (1) stimulating and assisting in the development and establishment of exemplary programs to serve as models for regular school programs and (2) assisting

the States in establishing and maintaining guidance, counseling, and testing programs. (The Education Amendments of 1974 consolidated the title II and most title III activities into a new title IV. Fiscal year 1976 was both the first year of funding under the new title IV and the last year of funding under title III. However, final title III projects will not run out until the end of fiscal year 1977, and requirements for State and local evaluations are in effect until that time. Title IV requires evaluation by the advisory council in each State.)

--Title VII provides OE discretionary funds for local education agencies and others to help carry out projects designed to meet the special educational needs of children who speak a language other than English and who come from low income families. Title VII is a demonstration program designed to build up the resources needed to start bilingual projects.

Local education agencies are required to evaluate their title I, III, and VII projects annually. State education agencies are also required to administer and annually evaluate their title I and III programs. However, title VII projects (bilingual education) operate under direct grants from OE; therefore, no State evaluations are required for such projects. OE is nevertheless required to consult with State education agencies before approving local education agencies' title VII grant applications and to give States the opportunity to make recommendations on the applications.

#### LIMITATIONS OF EDUCATIONAL EVALUATIONS

The central issue in most educational evaluation studies is whether programs such as title I of the Elementary and Secondary Education Act affect student progress. The President's Commission on School Finance was established to make recommendations to the President regarding the proper Federal role in financing elementary and secondary education. To be able to make its recommendations in the light of educational research results, the Commission asked a major research corporation to assess the available knowledge on what determines educational effectiveness. The resulting 1972 report 1/

---

1/Harvey A. Averch, et. al., "How Effective Is Schooling? A Critical Review And Synthesis Of Research Findings," R-956-PCSF/RC, The Rand Corporation, Mar. 1972, pp. iii-xii, 125, and 158.

states that research has found nothing that consistently and unambiguously makes a difference in student "outcomes." That is, research has not found any educational practice that offers a high probability of widespread success. OE's fiscal year 1975 annual evaluation report makes a similar statement about the attempt to identify the attributes of successful projects. A November 1976 OE-funded report, based on a study of educational innovations, found that the innovations make little difference in student achievement.

The 1972 report's findings were based on an examination of how valid the approach and results of numerous individual studies were. The report states that while some studies show a given educational practice to be effective, other similar studies find the same educational practice to be ineffective. It is unclear, the report adds, why this discrepancy exists. According to the report, four substantive problems appear in virtually every area of educational research that limit evaluation studies:

- Research data is, at best, a crude measure of what is happening. For example, student achievement is typically measured by scores on standardized achievement tests despite the many serious problems involved in interpreting such scores. (See chs. 5 and 6.)
- Educational outcomes are almost exclusively measured according to cognitive achievement, often leading to sparse and inconclusive results that provide little guidance on what practices are effective.
- There is almost no examination of the cost implications of research results, which makes it very difficult to translate research/evaluation results into policy-relevant statements.
- Few studies adequately monitor the relationship between what actually goes on in the classroom and student achievement, so that data may be affected by circumstances unrecognized in analysis.

Because of the problems above, according to the report researchers are confronted by the virtually impossible task of measuring those aspects of education they wish to study. That is, it is impossible for current research to reach definitive conclusions about educational outcomes because it cannot measure most of them well.

Other studies point out that numerous poorly designed and implemented evaluations result in very questionable or invalid data on which to base decisions about policy or programs.

A 1975 study, <sup>1/</sup> which reviewed the major title I evaluation efforts from 1965 through 1972, discusses some limitations in educational evaluations. The study traces the accepted belief in the necessity to evaluate education programs to title I of the 1965 act. This legislation established local reporting on projects in the hope that timely and objective information about the results of title I projects could reform the local administration of education and the methods of educating poor children. It was also hoped that systematic evaluation could make Federal management of education programs more efficient. According to the study, those pursuing educational reform saw evaluation as central to achieving it and assumed that reporting requirements would generate valuable information which would be used rationally in contributing to policy and program decisions.

The study concluded that after 7 years, more than \$52 million of expenditure on evaluations, and creating a number of alternative evaluation models, evaluation had failed to meet the expectations of those urging reform or even to serve the self-interest of Federal program managers. Regarding this conclusion, the study stated:

"There are numerous reasons why efforts to evaluate Title I failed \* \* \* The central cause is that school districts had no incentive to collect or report output data, and federal officials lacked the political muscle to enforce evaluation guidelines or to require cooperation with other federal evaluation efforts \* \* \*."

The study added:

--Those interested in reform efforts failed to take into account the difficulty of evaluating the process of schooling in general and title I in particular.

---

<sup>1/</sup>Milbrey Wallin McLaughlin, "Evaluation and Reform," a Rand Educational Policy Study, Ballinger Publishing Company, Cambridge, Mass., 1975, pp. vii-ix and 117-120.

- Legislatively mandated evaluation, intended to make school administrators accountable, has led to local evaluation that is, in the view of many observers, little more than an annual ritualistic defense of program activities. In addition, the Federal evaluation efforts have not contributed to the formulation of short-run management strategies or long-range planning. Evaluations based on an "impact cost-benefit" model have been used selectively to lend an appearance of rationality to decisions that are essentially political.
- Contrary to the expectations of those interested in reform efforts, neither Federal decisionmakers nor local school personnel showed much ability or interest in using evaluations to formulate title I policy or practice.
- Local perceptions of Federal initiatives and commitments as inherently unstable, combined with a basic local defensiveness about achievement measures, will probably continue to frustrate Federal attempts to secure objective, reliable information on program results.
- The highly political way that title I evaluation has been conducted, including use of its results, has weakened the credibility of evaluation as a policy instrument, in the opinion of many program personnel.
- A realistic and useful evaluation policy should acknowledge the inherent constraints that the policy system and the behavior of bureaucracies place upon evaluation.

In addition, the study stated:

"The history of Title I evaluation also suggests a number of implications about the conduct and use of evaluation in a multi-level government structure \* \* \* In a federal system of government, and especially in education, the balance of power resides at the bottom, with special interest groups \* \* \* Thus a federal evaluation policy that conflicts in fundamental ways with local priorities is unlikely to succeed \* \* \* Federal evaluators, then, are faced with a specifically political dilemma generated by their inability to insist upon accurate information on school effects and program

impact. And the existence of powerful social sanctions against a strong federal data requirement means that these barriers to the implementation of federal evaluation policy will remain."

OE's Assistant Commissioner for Planning, Budgeting, and Evaluation said that he agrees with this historical analysis. He believes that the evaluation approach that OE's evaluation office follows takes these problems into account because it is based not on school district data but on contractor data collected nationally.

## CHAPTER 2

### SCOPE AND METHOD OF REVIEW

Our objectives were to review the usefulness and limitations of federally supported education evaluations--focusing mostly on elementary and secondary education programs--and to solicit suggestions on needed program evaluation improvements, including needed research and development.

We reviewed OE's evaluation activities--principally carried out through the Office of Planning, Budgeting, and Evaluation--and related NIE activities. We also reviewed legislation, policies, procedures, and various Federal, State, and local education agency program evaluation reports relating to the Elementary and Secondary Education Act, titles I, III, and VII.

We interviewed officials from the Office of the Secretary of HEW, the Office of the Assistant Secretary for Education, the National Center for Education Statistics, OE, NIE, and the Office of Management and Budget. In addition, we interviewed the staff members of various congressional committees and officials from 11 education research organizations, including 4 publishers of commercial tests and 6 research/evaluation organizations. We also interviewed officials from two national interest groups concerned with education, and attended conferences of educators and measurement and evaluation experts which were held to improve student assessment or educational program evaluation.

To obtain State and local education agencies' views on Federal education program evaluations and related matters, we sent questionnaires to education agencies in all States and the District of Columbia, and to a statistical sample of local school districts throughout the Nation. The two sets of questionnaires were sent in April 1975 and were returned by June 1975.

The District of Columbia and 49 States responded to our State-level questionnaire. (To simplify questionnaire results in this report, we consider the District of Columbia to be a State.) Appendix I compiles the responses on the State education agency questionnaire. Respondents to section A of the questionnaire were almost always officials responsible for statewide assessment, accountability, and/or testing activities. Respondents to sections B and C were nearly always officials responsible for titles I and III programs, respectively.

Our questionnaire sample for local school districts was largely the same as a national sample used by the Office of Education in 1973. Neither sample included school districts having fewer than 300 pupils; both were stratified according to enrollment as follows: 125,000 pupils or more; 35,000 to 124,999 pupils; 9,000 to 34,999 pupils; 3,000 to 8,999 pupils; and 300 to 2,999 pupils.

Nineteen school districts compose the first group--the largest school districts--and all were included in the sample. An independent random sample of 813 school districts was drawn from the remaining groups. We received responses from 710 (85 percent) of the 832 school districts included in the sample.

As a result of the high response rate, the attitudes and opinions expressed in response to our local school district questionnaire are representative of the entire universe of 11,666 such districts in the Nation having 300 or more pupils. However, we projected the responses to a total of 8,936 local education agencies because this method, based on the weighting and the response rates across the various strata in our sample, allows us to obtain the most accurate percentages on the answers given.

Local education agency questionnaire results appear in appendix II. The numbers shown there represent the number of local school districts in the Nation to which our local questionnaire sample responses have been projected. Most local education agency respondents to section A of the questionnaire were directors for testing. Most respondents for sections B, C, and D were directors for titles I, III, and VII projects, respectively. However, in some cases superintendents responded to the questionnaire.

Our questionnaires focused on program evaluations of titles I, III, and VII of the Elementary and Secondary Education Act for a number of reasons. The largest Federal emphasis in education has been placed on the attempt to deal with various inequalities in educational opportunity. Programs of this kind have attempted to equalize educational opportunity for groups and individuals who are at a disadvantage (titles I and VII) and to improve the quality and relevance of American education through research and demonstration and dissemination of results (title III). Appropriations for the programs under these three titles are substantial. For fiscal year 1975 they were \$2.2 billion, which represented about one-half of all Federal elementary and secondary education program dollars. In addition, Federal legislation and HEW regulations

require national, State, and local evaluations for titles I and III, and national and local evaluations for title VII.

To supplement information obtained from the questionnaires, we interviewed education agency officials from five States, the District of Columbia, and 10 local school districts.

## CHAPTER 3

### OPPORTUNITIES TO IMPROVE

#### OE'S EVALUATION STUDIES

Evidence we gathered indicates that opportunities exist for OE's evaluation studies to better serve the Congress. Specifically, this includes timing the studies better and briefing congressional committee staff more frequently.

OE also needs to better define the program objectives to be evaluated and improve the use of evaluation results.

#### CONGRESSIONAL NEEDS SHOULD BE TAKEN MORE INTO ACCOUNT

According to OE, its evaluations are intended primarily to assist those involved in making educational decisions. This includes the Congress. In recent years the main emphasis has been on evaluating the national impact or effectiveness of major OE programs.

Decisions on education programs are made at various levels--in the Congress, OE, State education agencies, and local school districts. Decisionmakers at different levels sometimes need different information. For evaluation studies to be most useful to them, the views of those who are to use the results should be taken into account in evaluation planning and design. This increases the chance that their information needs will be adequately fulfilled and that resultant decisions will be well defined. It should also increase the chance that evaluation results will be effectively communicated to those intended to use them.

One such user to which OE should give greater attention in providing program evaluation information is the Congress. To obtain information on how useful OE evaluation studies are to the Congress and how much congressional views are taken into account in those studies, we contacted four key congressional committee staff members responsible for education matters, including the Majority and Minority Counsels for the Subcommittee on Elementary, Secondary, and Vocational Education, House Committee on Education and Labor; the Minority Counsel for the Subcommittee on Education, Senate Committee on Human Resources; and the Chief Counsel for the Senate Appropriations Committee. Two persons interviewed believed that OE does not obtain enough congressional input in designing its evaluation studies.

One of the four staff members stated that OE evaluation studies are generally useful and reasonably good. The other three stated that the studies often have been completely ineffective or have had little impact on legislation. Their most frequently cited reasons were that

- OE studies are not timed to coincide with the legislative cycle;
- OE efforts to interpret data and highlight important findings are insufficient; and
- OE briefings for congressional committee staff are not frequent enough.

OE's Assistant Commissioner for Planning, Budgeting, and Evaluation agreed with each of these statements and said that the evaluation office has not been sufficiently sensitive and responsive to congressional needs. He noted that the long leadtime necessary to plan and implement studies contributes to this problem.

OE's Assistant Commissioner for the Office of Legislation said that he considers the poor timing of evaluation studies to be a major factor inhibiting their impact on legislation. He stated that OE's evaluation office has not made the effort necessary to assure that evaluation results are arrived at and communicated soon enough to be considered in developing legislative proposals.

Commenting on our report, HEW stated that it does not concur with one of the three reasons cited above for the limited impact of its studies. HEW believes that its procedure for interpreting and summarizing evaluation study results is not deficient. Brief summaries of each evaluation study are sent to all members of the cognizant House and Senate authorizing and appropriation committees and their staffs, as well as to appropriate HEW Education Division staff and others.

PROGRAM OBJECTIVES NEED  
TO BE DEFINED

Evaluations can reach conclusive and useful findings about a program's effectiveness only if the program's objectives are defined and measurable. However, our review of selected evaluations and reports, as well as discussions with OE and non-OE experts, showed that one major problem in assessing education programs is the lack of sufficiently defined objectives.

For example, a national evaluation of the migrant program under title I of the Elementary and Secondary Education Act had not adequately assessed the program's effectiveness, an OE migrant program official said, because no one had developed acceptable criteria and objectives for measuring the program's success.

In addition, officials from OE's evaluation office stated that the purpose and objectives of title I itself have not been specified clearly. They said that this has made it difficult to evaluate the effectiveness of the title I program.

Other educational researchers have pointed out that it is difficult to establish criteria for Federal programs designed to respond to multiple needs, such as titles I and III of the Elementary and Secondary Education Act. They have stated that when the purposes of a program such as title I or III are ambiguous, various criteria are applied in assessing the program which leads to noncomparable evaluation results.

Likewise, officials representing a major test publisher said that the objectives of compensatory education programs should be clarified.

According to some educational researchers and contractors who compete for OE evaluation contracts, OE studies do not always define clearly the criteria and major questions that should be addressed. For example, a researcher who frequently prepares OE policy studies said that OE evaluation studies do not help in formulating policy because they are not set up to answer the major short-term or long-term questions.

Other educational experts stated that lacking measurable and sufficiently defined objectives often led to evaluation studies that addressed unanswerable questions and produced inconclusive results. They added that the language used in legislation, regulations, policy manuals, plans, and budgets is generally ambiguous and fails to precisely define program objectives and make the evaluation useful.

As mentioned in chapter 1, legislation requires that HEW provide the congressional committees having responsibility for education with an annual report evaluating program effectiveness in achieving legislative purposes. The report is required to set forth goals and specific objectives in

qualitative and quantitative terms for all programs and projects assisted which are evaluated and relate those goals and objectives to the program's purposes.

Our review of the annual evaluation report on OE programs for fiscal year 1975 showed that most of its statements of program goals and objectives merely restated the legislative purposes or general goals, and did not set forth specific objectives. Quantitative objectives, even in the broadest sense, were established for very few programs.

OE's Assistant Commissioner for Planning, Budgeting, and Evaluation agreed with our observations and said that OE has seldom established specific objectives. He stated that this should be corrected, but that OE sometimes faces opposition from the Congress and others when it specifies objectives. We agree that the Congress has major responsibility for specifying program objectives. In our view, however, the legislative requirement and OE's limited responsiveness to it, as well as the need for providing a clear basis for program evaluation, dictate that OE make a better effort to set forth specific qualitative and quantitative program objectives for congressional consideration.

#### USE OF EVALUATION RESULTS NEEDS MORE ATTENTION

In 1972, OE's Office of Planning, Budgeting, and Evaluation instituted a procedure which entails drafting and implementing a "policy implication memorandum" to increase the use of the evaluation findings with which OE concurs in policy and program decisions. However, the procedure has not been used to its potential.

The memorandum procedure was developed to translate the findings of evaluation studies into a list of "action items" for program management. It represents an attempt to make sure that study results receive proper attention from OE and department decisionmakers. For instance, the Commissioner of Education may use the memorandum as a base for policy decisions. He may selectively direct actions to be taken, the office responsible for implementing them, and their due dates.

According to OE's Assistant Commissioner for Planning, Budgeting, and Evaluation, the procedure is one of the most important parts of the whole evaluation process, which encompasses evaluation planning through implementation of results. He believes it is superior to other implementation methods

because the implications of evaluation results which OE accepts for implementation, including related followup requirements, are explicitly set forth in areas such as basic policy, budgeting, staffing, and program regulations. Also, the Assistant Commissioner believes, decisions are more likely to be made on action items under such a procedure.

An OE evaluation official said that policy memorandums were to be written after the completion of each evaluation study in which important findings were produced. From March 1972 to March 1974, OE's evaluation unit completed 32 studies, costing a total of about \$8.2 million. The Secretary's evaluation office, using education evaluation funds of about \$3.5 million, completed 18 studies during a similar period.

Although OE considers the policy implication memorandum procedure a key to assuring the use of study results, it was followed on only two evaluation studies completed during this period. 1/ We reviewed its use in both instances--the studies cost \$120,000 and \$772,000--to ascertain how it affected policy and program changes.

The first policy memorandum was dated December 26, 1972; the other August 19, 1974. Nine months elapsed between the first study's completion date and the preparation of the policy memorandum, and 10 months elapsed for the second study.

Several sources doubted the impact of the first study and memorandum. A program official affected by the study, relating to title I of the Elementary and Secondary Education Act, said that there was no evidence showing that memorandum recommendations were being taken seriously or were influential in causing program change. In addition, OE program officials said that they were pursuing some of the recommendations contained in the memorandum prior to availability of the results of the evaluation study. Also, OE's Assistant Commissioner for the Office of Legislation stated that neither the policy memorandum nor the study has had much influence on the process of developing new legislation for title I. H. explained that one reason for their lack of impact was the Administration's emphasis on educational revenue-sharing at the time of the study's publication. This emphasis reduced interest in amending existing legislation that might have been substantially

---

1/In November 1976, OE's Assistant Commissioner for Planning, Budgeting, and Evaluation stated that a third memorandum had been written.

eliminated and replaced if the educational revenue-sharing legislation had been enacted.

The evaluation office official who wrote the memorandum said in November 1976 that the evaluation office was not following up on all memorandum recommendations despite the fact that some were still open issues. He agreed that better followup of such open issues should be a part of the policy implication memorandum system.

An OE official involved in the evaluation study said that although very few of its recommendations were acted upon, it compiled evidence to support certain conclusions for the first time.

A program official stated that if program officials had been consulted about the subject matter, content, and design of the study, they would have been in a better position to use the study's results. Evaluation office officials disagreed with this view and stated that they extensively involve program officials in designing evaluation studies.

The second policy memorandum contained only one major recommendation which required further action. The recommendation was implemented, and substantially changed program emphasis. The OE project monitor for this study said that a policy implication memorandum was written for it because the procedure was given high priority at the time.

Regarding other evaluation studies for which no policy memorandums were written, OE project monitors gave these explanations:

- The priority placed on writing the memorandums was not high enough.
- Evaluation studies had overlapping cycles; therefore, before one was completed another could start, detracting from full appreciation of the earlier one.
- Delays in receiving study reports could have affected writing the memorandums.

An OE official commented on this situation, stating that because policy memorandums are not being written, meaningful study conclusions fail to reach policy planners and program administrators who have a voice in the legislative process. He felt that the procedure is needed to call attention to the significant recommendations in each study.

In our opinion, delays of nearly 1 year before the two policy memorandums were written and approved and the general lack of such memorandums clearly point to the need for more OE emphasis on assuring increased use of the evaluation findings with which OE concurs. This includes giving a higher priority to policy implication memorandums or some other procedure for achieving this purpose. OE's Assistant Commissioner for Planning, Budgeting, and Evaluation agreed that OE has not successfully carried out the policy implication memorandum system. Although he believes the memorandums are of central importance, other priority matters have preempted staff time.

### CONCLUSIONS

Opportunities exist for OE's evaluation studies to better serve the Congress. These include better timing of the studies and more frequently briefing congressional committee staff.

There is a need to better define the program objectives to be evaluated. The use and implementation of evaluation results also need more attention.

### RECOMMENDATIONS TO THE SECRETARY OF HEW

We recommend that the Secretary of HEW direct OE to:

- More strongly emphasize the purpose of providing information to the Congress when planning, implementing, and reporting on evaluation studies. In particular, more attention should be given to timing the studies so that they more clearly coincide with the legislative cycle and briefing congressional committee staff more frequently.
- Take steps to comply with the General Education Provisions Act (20 U.S.C. 1226c) requirement that, in the annual evaluation report to the House Committee on Education and Labor and the Senate Committee on Human Resources, HEW set forth goals and specific objectives in qualitative and quantitative terms for all programs which are evaluated. OE should indicate in the evaluation report that it is setting forth specific objectives tentatively in response to the congressional requirement and as a basis for future discussion and agreement with the committees on program evaluation matters. These matters should include the acceptable evaluation data needed by congressional decisionmakers and the measures to be used. If HEW still does not intend to comply with

this requirement, it should propose legislative changes to the Congress to avoid continued agency noncompliance. In the meantime OE should initiate dialogues with the appropriate House and Senate committees to seek understanding and agreement on the specific program objectives to be used for evaluations as well as acceptable evaluation data and measures for each program to be evaluated.

--Improve the implementation of evaluation results by giving greater attention and priority to procedures such as the issuance of policy implication memorandums designed to assure implementation of those results.

### AGENCY COMMENTS AND OUR EVALUATION

HEW commented on matters discussed in this report in a letter dated June 15, 1976. (See app. III.)

HEW agreed with our recommendation that OE should more strongly emphasize meeting congressional information needs by timing the studies to better coincide with the legislative cycle and briefing congressional staff more frequently. HEW stated that it has initiated a series of reviews of its studies, focusing on predicted production dates for findings and recommendations in relation to critical dates for legislative input. HEW also stated that the need for congressional committee staff briefings has not been given proper attention but that it has recently decided to institute such briefings on all major evaluation studies and will initiate this procedure in the coming weeks.

Regarding our recommendation that OE should improve the implementation of its evaluation results, HEW agreed and stated that the policy implications memorandum procedure, which is an invention of OE's evaluation office, has not been used nearly as extensively as it should have been. HEW said that efforts are currently underway to expand the use of these memorandums and that agency officials are now conducting periodic reviews of the schedule for producing the memorandums and emphasizing their high priority.

Our draft report proposed that OE better define the program objectives to be evaluated as required by the General Education Provisions Act (20 U.S.C. 1226c). This includes translating the legislative purposes of individual programs evaluated into specific qualitative and quantitative program objectives, and clearly stating these objectives, and the progress made toward achieving them, in the annual evaluation report.

HEW disagreed with this proposal because it believes there are limits on OE's "authority and ability" to increase the clarity and specificity of its program objectives. HEW comments, however, ignored the fact that the General Education Provisions Act requires the agency to set forth such objectives for individual programs included in its annual evaluation report to the House Committee on Education and Labor and the Senate Committee on Human Resources.

HEW said that in most cases legislation fails to state a program's objectives with sufficient clarity for evaluation. It appears that the Congress has also recognized that program legislation does not provide sufficiently clear and specific objectives for evaluation; therefore, in the General Education Provisions Act it has required HEW to set these forth in a report to the appropriate congressional committees. HEW comments did not respond to its noncompliance with this requirement, although OE's Assistant Commissioner for Planning, Budgeting, and Evaluation agreed with our finding on this matter. (See p. 17.)

HEW stated that it "proceeds at considerable peril in trying to further specify legislation" and that

"\* \* \* in many cases it has been the Congress' specific intention to avoid specification of program objectives and leave such judgments and decisions up to State and local officials."

However, in conducting national program evaluations the Office of Education is often implicitly establishing and using specific program objectives; for example, standardized tests, frequently used in these evaluations, are based on specific instructional objectives. We believe there is an important distinction between specific program objectives explicitly set forth for Federal evaluations and those which would be established to dictate to State and local education agencies specifically how to design and run their programs which use Federal funds. If OE cannot explicitly set forth specific program objectives it would use for Federal evaluation purposes, then we believe it is inconsistent to conduct national program evaluations which contain such objectives implicitly.

We agree with HEW that there is political opposition, but we believe such opposition is really directed toward Federal infringement on State and local education agency prerogatives. Such opposition effectively restrains the Federal agency from

trying to dictate State and local program goals, specific objectives, approaches, etcetera. In our view, HEW should comply with the requirement, but in doing so it should clearly indicate in the evaluation report that the specific objectives are tentative, are being set forth in response to the legislative requirement, and are intended only for congressional scrutiny and as a basis for mutual discussion and agreement on program evaluation matters, including the acceptable evaluation data needed by congressional decisionmakers.

In its general comments on our report HEW stated its belief that there is growing professional opinion that OE's studies have, over the past 10 years, been responsible for many major changes in existing legislation. Also, in HEW's view the assumption that there are certain decisionmakers, and that effective evaluations provides timely data to them, is increasingly being questioned. Instead HEW believes that effective evaluations affect the broad political climate within which particular decisions are made.

Obviously, HEW believes that OE studies are affecting legislative decisions. However, as discussed in our report, three of the four key congressional committee staff members interviewed, who are responsible for education matters, said that the studies often have been completely ineffective or have had little impact on legislation. Therefore, we continue to believe that the primary purpose of these evaluations should be to provide useful information to decisionmakers.

#### RECOMMENDATION TO THE CONGRESS

The Congress should recognize that HEW is not in compliance and does not intend to comply with the General Education Provisions Act requirement (20 U.S.C. 1226c) as noted above. HEW feels that its authority and ability to comply with this legislative requirement are limited. The Chairmen of the House Committee on Education and Labor and the Senate Committee on Human Resources should discuss this matter further with agency officials to seek a common understanding with them on the process or approach to be used for (1) clarifying program objectives for evaluation and (2) reaching agreement on acceptable evaluation measures and data for each program to be evaluated.

## CHAPTER 4

### PROBLEMS WITH STATE AND LOCAL EVALUATION REPORTS

A large amount is spent to evaluate State education agency title I and III, and local education agency title I, III, and VII elementary and secondary education programs. Agency officials at these levels, responding to our questionnaires, indicated a need to improve program evaluation reports, including these important areas for determining program effectiveness: the credibility of findings and the qualification and quantification of measurement data.

Other areas needing attention and improvement to make the State and local evaluation reports more useful include

- the relevance of the reports to policy issues,
- the completeness and comparability of the data reported, and
- report timeliness.

The significance of these problems and other factors raise a question about whether the present reporting systems based on aggregated local data can be improved so that they meet program information needs at Federal, local, and/or State levels.

### STATE AND LOCAL EVALUATION EXPENDITURES

In addition to the funds authorized for program evaluation at the Federal level (see ch. 1), the Elementary and Secondary Education Act requires annual State and local education agency evaluations of title I and III programs, and local evaluations for title VII. The following tables show--on the basis of State and local agency responses to our questionnaire--estimates of the evaluation funds expended by State and local education agencies for the programs during fiscal year 1974.

Estimated State-level Expenditures  
for Evaluating Selected  
Elementary and Secondary Education Programs  
Fiscal Year 1974 (note a)

	<u>Title I</u>	<u>Title III</u>
Total State-level program expenditures reported for evaluation	\$ 2,066,020	\$1,723,805
Average State program grant	24,520,132	2,127,455
Average expenditures on evaluation per State	43,958	39,177
Average percent of grant spent for evaluation (note b)	1.2%	5.4%

a/All amounts are based on unverified questionnaire responses from 47 States for title I and 44 States for title III.

b/The percentages shown are based on the averages of the percent of grant funds reportedly spent for evaluation by the States. However--overall, two-tenths of one percent of title I and 1.8 percent of title III funds were reportedly spent for evaluation. The differences between these percentages and those shown are because larger percentages of the smaller grants were generally used for evaluation.

Estimated Local-level Expenditures  
for Evaluating Selected Elementary  
and Secondary Education Programs  
Fiscal Year 1974 (note a)

	<u>Title I</u>	<u>Title III</u>	<u>Title VII</u>
Local project expenditures reported for evaluation	\$31,790,960	\$5,089,344	\$1,574,320
Average project grant per local district grantee	161,417	50,240	164,705
Average evaluation expenditures	3,860	2,101	4,537
Average percent of grant spent for evaluation (note b)	6.4%	5.0%	3.1%

a/All amounts are based on unverified questionnaire responses from local school district respondents representing 8,236 title I, 2,422 title III, and 347 title VII projects.

b/The percentages shown are based on the averages of the percent of grant funds reportedly spent for evaluation by the local projects. However--overall, 2.4, 4.2, and 2.8 percent of title I, III, and VII funds, respectively, were reportedly spent for evaluation. The differences between these percentages and those shown above are because larger percentages of the smaller grants were generally used for evaluation.

OE officials stated that OE does not collect State and local education agency data on evaluation expenditures.

STATE AND LOCAL EVALUATION  
REPORTS NEED IMPROVEMENT

Our questionnaires asked State and local officials connected with administering title I, III, and VII programs under the Elementary and Secondary Education Act to rate the adequacy of these aspects of the State and local evaluation reports: credibility of findings, presentation of required management information needs, qualification of findings, qualification and quantification of measurement data, and focus and scope. In our opinion, the adequacy of these aspects is likely to greatly influence how much the reports satisfy the policy, management, and program information needs of State and local officials.

In most cases, more respondents rated the various aspects of the evaluation reports in the "adequate or better" categories than in the "less than adequate" categories. However, in each aspect, many respondents to the questionnaires indicated that State and local evaluation reports were inadequate; most of these ratings were in the "marginal" category. In addition substantial numbers of State and local officials rated State and local evaluation reports to be less than adequate in these important areas for determining program effectiveness: the credibility of findings and the qualification and quantification of measurement data. In our opinion, such large numbers of less-than-adequate ratings indicate a serious need for improvement in the reports.

The following table summarizes the respondents' less-than-adequate ratings of State evaluation reports.

State Evaluation Reports:  
Summary of Respondents' Less-Than-  
Adequate Ratings (note a)

	State program officials of title:		Local project officials of title:		
	(percent)				
	<u>I</u>	<u>III</u>	<u>I</u>	<u>III</u>	<u>VII</u>
Credibility of findings	59	40	45	36	61
Presentation of required management information needs	59	40	48	38	59
Qualification of findings	47	47	50	39	69
Qualification and quantifica- tion of measurement data	57	55	49	42	51
Focus and scope	42	43	38	33	60

a/Percentages for State program officials are in all cases based on questionnaire responses from 48 or 49 States for title I and between 45 and 47 States for title III. Percentages for local project officials are based on sample responses and in all cases represent more than 8,300 title I projects, more than 2,350 title III projects, and about 300 title VII projects. See app. I, questions 9-13 and 20-24, and app. II, questions 9-13, 21-25, and 33-37.

The following table summarizes the questionnaire respondents' less-than-adequate ratings of local evaluation reports:

Local Evaluation Reports:  
Summary of Respondents' Less-Than-  
Adequate Ratings (note a)

	<u>State</u> <u>program</u> <u>officials</u> <u>of title:</u>		<u>Local</u> <u>project</u> <u>officials</u> <u>of title:</u>		
	—————(percent)—————				
	<u>I</u>	<u>III</u>	<u>I</u>	<u>III</u>	<u>VII</u>
Credibility of findings	67	60	38	31	47
Presentation of required management information needs	71	56	39	31	42
Qualification of findings	69	60	41	34	39
Qualification and quantifica- tion of measurement data	57	58	44	40	41
Focus and scope	50	40	30	25	43

a/Percentages for State program officials are in all cases based on questionnaire responses from 48 or 49 States for title I and 48 States for title III. Percentages for local project officials are based on questionnaire sample responses and in all cases represent more than 8,500 title I projects, more than 2,400 title III projects, and about 350 title VII projects. See app. I, questions 9-13 and 20-24, and app. II, questions 9-13, 21-25, and 33-37.

Credibility of findings

The questionnaire defined this aspect as the degree of confidence expressed in the findings through statements about statistical certainty, soundness of method, evidence of replication, consensual agreements, similar experiences, supporting expert judgment and opinions, and reasonableness of assumptions.

As the table on page 27 shows, between 36 and 61 percent of State and local respondents from the various programs rated the credibility of findings in their program's State evaluation reports to be less than adequate.

As the table above shows, the percentage of State and local respondents from the various programs that rated local evaluation reports to be less than adequate in this aspect ranged from 31 to 67 percent.

### Presentation of required management information needs

The questionnaire defined this category as the extent to which the reports are informative to those who evaluate and update current policies and transfer policy decisions into plans, budgets, curriculum or program implementation, operational oversights, resource allocations, forecasts, status assessments and reports, educational accountability, costs, benefits, and efficiency assessments.

The percentage of State and local respondents finding this aspect of their program's State evaluation reports to be less than adequate ranged from 38 to 59 percent. Similarly, for local evaluation reports, the range was from 31 to 71 percent.

### Qualification of findings

The questionnaire defined this aspect as the extent to which the reports properly qualify the findings and assumptions and identify those conditions and situations to which the findings are not applicable.

The percentage of State and local respondents rating this aspect of their program's State evaluation reports to be less than adequate ranged from 39 to 69 percent; for local evaluation reports, the percentage ranged from 34 to 69 percent.

### Qualification and quantification of measurement data

The questionnaire defined this category as the extent to which the evaluation assessments can be qualified and quantified into measurable attributes and parameters that address the problem in measurable, operational, or concrete terms.

The percentage of State and local respondents rating this aspect of their program's State evaluation reports to be less than adequate ranged from 42 to 57 percent; for local evaluation reports, the range was from 40 to 58 percent.

### Focus and scope

The questionnaire defined focus and scope as the adequacy with which the reports covered essential and related material and the appropriateness of the emphasis and treatment given to the relevant topics and details and high and lower priority information.

The percentages of State and local respondents rating this aspect of their program's State evaluation reports to be less than adequate ranged from 33 to 60 percent; for local evaluation reports, the range was from 25 to 50 percent.

#### Other questionnaire results

Only 61 percent of State title I officials and 62 percent of State title III officials said that local evaluation reports "generally" or "very often" adequately show evidence of qualifiable or measurable student benefits; and only 64 percent of State title I officials and 50 percent of State title III officials said that State evaluation reports are generally or very often adequate in this respect.

Although over 75 percent of the State respondents said they use local and State evaluation results for policy, programmatic, or management decisions, the only data contained in State and local reports which was frequently found adequate by State officials was

- the number of children in the program and
- the per-pupil expenditures.

Most local school district respondents stated that their reports are generally or more often than generally adequate in providing essential information on the

- number of children in the program,
- per-pupil expenditures for each program,
- evidence of quantifiable and measurable achievement, and
- evidence of quantifiable or measurable pupil benefits.

#### OTHER PROBLEMS

To be useful in making decisions at the Federal level, State and local evaluation reports should be timely, complete, comparable, and relevant to policy issues. Among the Elementary and Secondary Education Act State and local evaluation reports submitted to OE that we reviewed, most did not meet any of these criteria.

Three factors need to be considered:

- The significance of these and other problems discussed in this chapter.
- Constraints on the Federal role in education as discussed in chapter 1. (See pp. 8 to 10.)
- Questions about whether ongoing efforts to resolve these problems with evaluation models for title I will be effective.

If the questions about the models are not resolved, the feasibility of producing improvements that will enable the present reporting systems based on aggregated local data to provide the information needed is questionable.

#### Usefulness and relevance to policy issues

Evaluation reports have the best chance of affecting policy decisions if they are designed to directly address policy issues. The reports should, for example, indicate programs' or projects' successes and failures.

In relation to this, OE officials generally said that State and local evaluation reports were rarely used to support operational or policy changes. State compensatory education program officials made similar statements, saying that State and local evaluations are of limited usefulness to those making decisions. In addition, OE's Assistant Commissioner for Planning, Budgeting, and Evaluation stated that there is no question that the State and local evaluation reports are not useful.

Similarly, two OE-contracted research studies question the usefulness of State and local evaluation reports:

- A study analyzed the policy-relevance rating of title I State evaluation reports for the 5 fiscal years 1969-73. Study results revealed serious problems concerning the validity of data reported by most States, precluding any meaningful interpretation of data. The study noted that the majority of the reports examined were seriously deficient in reporting policy-relevant information.

--A 1974 study of a nationwide sample of title VII evaluations concluded that no strong relationship could be established between the content or quality of evaluation reports and funding levels awarded to projects. It also concluded that none of the evaluations presented data which would indicate project failure, noting that such information is essential if a report is to be useful to decisionmakers. The study also noted that over half the evaluations it reviewed were of limited or no use for making judgments about project effectiveness.

Our review of selected evaluation reports and discussions with OE program officials generally confirmed that the State and local education agency reports, because of problems cited in this chapter, have little effect on Federal-level decisions.

#### Complete and comparable data in reports

Evaluation of Federal programs at the local level should produce reports containing valid, complete, and comparable results if data from each report is to be aggregated to provide Federal policymakers and program administrators with a good perspective on how well the program as a whole has worked and which approaches have produced the best results.

#### Response to congressional requirement

The Congress has recognized the need for OE to make State and local education agency evaluation reports more usable. The Education Amendments of 1974 (Public Law 93-380) directed the Commissioner of Education to carry out certain evaluation activities for programs authorized by title I of the Elementary and Secondary Education Act. Section 151 of the act directs the Commissioner to

- provide for independent evaluations which describe and measure impact of programs and projects,
- develop and publish standards for evaluation of program or project effectiveness,
- provide models for evaluation to State education agencies,

- provide such technical and other assistance as may be necessary,
- specify objective criteria which shall be utilized in the evaluation of all programs, and
- outline techniques and methodology for producing data which is comparable on a statewide and nationwide basis.

OE has begun activities to address each of these requirements, and evaluation models are being developed and refined. OE plans to require State and local use of the models (or use of alternatives that the Commissioner certifies will generate compatible evaluation data). However, OE Planning and Evaluation Office officials have stated that they do not believe that evaluation needs at the local, State, and Federal levels can all be met by the same approach. In addition, these officials, including OE's Assistant Commissioner for Planning, Budgeting, and Evaluation told us that although the models will make aggregation of locally collected data theoretically possible, accomplishing successful implementation for 14,000 school districts is doubtful, or at least questionable. They noted that the problems that need to be overcome are methodological, fiscal, and political. For example, some State and local officials do not want comparisons of educational results.

An OE evaluation official said further that the models are based on assumptions about such things as the common metric (scale) used, the soundness of the tests employed, and whether those tests are similar enough to provide data that is truly comparable; these assumptions represent compromises, and whether they will satisfy everybody is unclear. He stated that because of its technical nature the information based on the models may be provided to those who are responsible for decisions without an explanation of the assumptions involved. We believe that such information, when provided to the Congress and other users, should set forth these assumptions as clearly as possible.

OE's evaluation office is offering these types of technical assistance on title I: written handbooks on evaluation topics, such as the models; technical assistance workshops to train State administrators and evaluators in using the models and to prepare them to train local school district personnel; and consulting services to States to help them use the models. Technical assistance centers have been established throughout the country under OE contracts to

provide these services. According to OE, such services might include writing computer programs, conducting workshops to train local personnel, and helping with data analysis. To alleviate many of the States' staffing and technical problems in using the models, OE encourages States to call upon the technical assistance personnel to solve problems at the State level, and, depending on the desires of each State, at the local school district level, too.

#### Need to improve data

As stated earlier, OE program officials and State education agency officials feel that State evaluation reports are generally not useful for making management decisions; both expressed the need for uniform evaluation methods which would lead to comparable data reporting.

When local title I officials were asked whether or not local districts could compare the data from their local reports with data on the same program contained in State and Federal reports, 49 percent said they could do this only occasionally or seldom with State reports, and 64 percent replied similarly regarding Federal reports. Corresponding results were obtained for titles III and VII.

The results of a 1972 OE-contracted study done by an educational research firm to identify successful State programs and local projects in compensatory education illustrates the lack of reliable, comparable data. According to OE's Assistant Commissioner for Planning, Budgeting, and Evaluation, the study was able to identify only "a discouragingly low number of successful projects," because many projects did not have an evaluation design good enough to produce reliable data on cognitive results and many projects were poorly designed, poorly managed, or badly implemented. Furthermore, the study concluded that the lack of representative data from each State that could be combined in a meaningful way made it extremely difficult to address the effectiveness of the Elementary and Secondary Education Act title I program.

In an attempt to provide a means for Federal, State, and local education agencies to develop comparable evaluative data, OE, through the National Center for Educational Statistics (which was part of OE at that time), funded the "Anchor Test Study." The study was intended to develop tables and procedures for equating test scores among the eight most widely used standardized reading tests for fourth, fifth, and sixth grade children. OE developed the basic plan and

the detailed specifications for the study, which was intended to become an integral part of Federal elementary and secondary education program evaluation. OE expected the study to be useful at all levels of educational administration; that is, evaluative results would become useful to State and Federal Governments because program evaluation test data could be combined in a meaningful way. In addition, local school systems could have flexibility in selecting achievement tests to be used in their evaluations.

The results of the Anchor Test Study became available in September 1974. According to our questionnaire, however, as of April-June 1975 all of the 282 projected local education agency respondents indicated they had little or no information about the study. Although the great majority of State respondents knew about the study, only 18 percent had used it.

Test publishers and educational evaluators informed us that the study was technically excellent and potentially useful, but OE had not adequately planned for its use and dissemination. They also noted that the study's usefulness is diminishing as new versions of those tests included in the study are developed and published. State and local education officials interviewed said that the study has had little impact on their evaluation efforts primarily because OE did not pursue its implementation; little effort has been made to direct, encourage, or promote the use of the study so that evaluation results could be made more comparable.

### Timeliness

OE often does not receive annual State evaluation reports in a timely manner. For instance, 1 month before the end of fiscal year 1975, less than half of the fiscal year 1974 State elementary and secondary education, title I, program evaluation reports had been received. Notwithstanding OE's attempts to obtain delinquent reports, several States are up to 2 years behind in submitting theirs.

Delinquent reporting helps to prevent meaningful aggregation of State evaluation reports into a national picture of program effectiveness that can affect Federal decisions on funding and program operation. According to an OE program official, some of the more significant reasons they have been given by States for late filing include delays due to

- inclusion of summer program results,
- uncooperative local education agencies,

- differences between OE, State, and/or local agency personnel over what information the evaluation reports should contain,
- low State priority for the programs or their evaluation, and
- schedule slippages in printing, computer, and staff processing of the reports.

## CONCLUSIONS

The amount of funds spent to evaluate State education agency titles I and III, and local education agency titles I, III, and VII elementary and secondary education programs is substantial. If the reporting systems based on aggregated local agency data are to be effective, there is a need to improve the adequacy of various aspects of the evaluation reports, including two areas important in determining program effectiveness:

- The credibility of findings.
- The qualification and quantification of measurement data.

The usefulness of the State and local evaluation reports also needs improvements in

- report relevance to policy issues,
- data completeness and comparability, and
- report timeliness.

Valid, complete, and comparable evaluation data is important for results to meaningfully contribute to decisions at all levels. If local and State evaluation data continues to be aggregated for use at higher levels, standardization of data collection efforts and techniques is needed to provide comparable results.

Because of the constraints on the Federal role in education, as discussed in chapter 1 (see pp. 8 to 10), in our opinion it is unlikely that the Federal Government would seek to provide the valid and comparable data needed by dictating uniform evaluation methods and procedures to State and local education agency grantees. In addition, questions exist

about whether the title I evaluation models will be able to provide valid and acceptable data to meet program information needs. If these issues are not resolved, we question whether improvements can be made that will enable the reporting systems based on aggregated local data to meet program information needs at Federal, local, and/or State levels.

#### RECOMMENDATION TO THE SECRETARY OF HEW

We recommend that the Secretary direct OE to assess whether State and/or local evaluation reports for titles I and VII of the Elementary and Secondary Education Act can be improved so that they supply officials at Federal, local, and/or State levels with the reliable program information they need for decisionmaking. This includes assessing the adequacy of the title I evaluation models and related data.

- If OE determines that it is unrealistic to expend resources on improving these programs' State and local evaluation reporting systems based on aggregated data, then it should take the needed steps to adopt more feasible and effective approaches. This should include eliminating unwarranted reporting requirements and if necessary proposing to the Congress any legislative changes needed to accomplish this.
  
- If, however, these reporting systems are continued, OE should more strongly emphasize improving the adequacy of State and local evaluation reports for title I and local evaluation reports for title VII. This should be done by giving greater management attention and priority to making reports more complete and comparable, relevant to policy issues, timely, credible, and adequate in the qualification and quantification of their measurement data.

#### AGENCY COMMENTS AND OUR EVALUATION

HEW stated that it concurs with the general thrust of our draft report recommendations for chapters 4 and 7, however, it also stated that most of the actions we proposed are already underway, many by legislative mandate, and in some cases they are near completion.

Our report extensively discusses these actions underway. (See pp. 32 to 34.) HEW's statement implies that the actions underway are likely to be successful. Statements by OE evaluation office officials directly contradicted this implication and questioned the practical feasibility of

aggregating local level data. (See p. 33.) We concluded in chapter 4 that there are issues which raise questions about whether the evaluation models will be able to provide valid and acceptable data to meet program information needs. Therefore, we believe that OE needs to assess the effect of the problems connected with its "actions underway" in relation to the feasibility and cost-effectiveness of improving the reporting systems to provide the reliable information needed for decisionmaking at Federal, State, and local levels.

HEW also said that our understanding was incomplete regarding the much needed and legislatively mandated actions to improve State and local evaluations and reporting. In clarifying the meaning of this statement in relation to title I, HEW stated that we should have provided more information relating to the workshops, technical assistance centers, and information dissemination activities relating to the evaluation models mandated by the Education Amendments of 1974. We disagree. Our report discusses each of these issues. (See pp. 32 to 34.) Appendix III provides HEW's additional information on these matters.

Specifically regarding title I, HEW stated that once the models and revised reporting system are in place across the Nation, their use should produce data which can be aggregated across school districts and States. At that time, HEW believes that OE will be in a better position to assess whether or not the data are sufficiently free of systematic errors to support satisfactory aggregations to the State and national levels. If they are not, HEW stated it can then determine whether technical problems could be overcome or whether different kinds of studies are needed to satisfy Federal, State, and local reporting requirements.

We believe that to foster maximum efficiency and economy OE should make the needed assessment as soon as it has a good enough understanding, if it does not already, of such factors as the likely validity, reliability, and comparability of the reporting system data as well as the other methodological, fiscal, and political problems involved.

HEW also stated that OE has interviewed personnel in Federal policymaking roles and received information from an advisory title I group (which included parent representatives) on the kinds of information that should be included in the annual State and local evaluation reports. In addition, evaluation models and their reporting forms were developed and reviewed by each State agency and three of its local agencies to identify possible problems. As a result of

all these efforts, a limited core of essential information was identified as "desirable at the Federal level" and this will become the Federal evaluation requirement when regulations for this portion of the legislation are published.

We commend efforts to identify the Federal information requirements. However, we remain concerned about the identification of State and local agency information requirements and whether these needs can all be met by the same reporting system. We are also concerned about the possible unnecessary duplication of using both this reporting system and OE national evaluations on these programs to meet these Federal information requirements.

With respect to title VII, HEW stated that local evaluation reports can be improved and cited the following steps intended to accomplish this:

--HEW has recently published regulations strengthening requirements for bilingual project evaluation.

--The National Institute of Education and OE have a joint project underway to upgrade the technical expertise of local evaluators.

Although HEW indicated that aggregating local title VII data is more difficult than for title I, aside from concurring with the general thrust of our recommendation, it did not respond to our proposal for OE to assess whether it is realistic to try to improve local title VII reports so that they supply Federal-level decisionmakers with the reliable information they need.

### General comments

In addition, HEW provided several general comments. These observations and our responses follow.

#### HEW comment

HEW stated that the report needs to give more careful consideration to evaluation costs and that the quality of data the report "appears to expect" would require significant additional resources which would be high in relation to the possible payoffs through program improvements.

### Our response

These comments ignore the cost-effectiveness considerations of our recommendations. In addition, HEW has misinterpreted what our report expects. Although HEW comments did not recognize or respond to this proposal, our draft report proposed that OE assess whether it is realistic to expend additional resources on improving the State and local reporting systems for titles I and VII. If OE determines that it is not realistic, then we proposed that it initiate action to eliminate unwarranted reporting requirements, including proposing to the Congress any needed legislative changes. We would certainly expect cost-effectiveness considerations to be a part of OE's assessment.

Our draft report also proposed (see p. 76) that in connection with this assessment, OE determine (1) whether it is realistic to try to serve Federal, State, and local levels with information based on local agency evaluations and (2) how the information needs at Federal, State, and local levels can best be met. HEW comments also did not recognize or respond to this proposal.

These proposals certainly do not require "significant additional resources" on reporting systems. In fact, they question the value of present and proposed expenditures and suggest that OE face this issue. We recommend this assessment because it is not clear and has not been demonstrated that the reporting systems based on aggregated local level data are now providing or can be made to provide valid, useful, and cost-effective data to Federal, local, and/or State decisionmakers. In our view, significant additional resources should not be expended until there is some solid evidence that they would be cost-effective.

### HEW comment

HEW stated that the report does not give adequate recognition to whether the tradeoffs in improved program quality are likely to justify additional spending.

### Our response

Once again, HEW has misinterpreted our draft report proposals. As discussed above, we proposed that OE make the needed assessment to determine the realism of trying to improve these reporting systems. We believe that such an assessment, if properly conducted, would necessarily include consideration of the tradeoffs involved. In our view it is the agency's responsibility to make such assessments. This

is especially true in this situation where, as discussed in this chapter, not only have OE-funded studies shown significant problems but OE evaluation office officials have expressed serious questions about the feasibility of the approach currently being followed to solve these problems.

HEW comment

HEW took exception to our statement that the "amount of funds spent to evaluate State and/or local education agency titles I, III, and VII elementary and secondary education programs is substantial," saying that the percentages of funds involved at the State level are not substantial.

Our response

The amount that we intended to refer to is the total reportedly spent not only on State evaluations for titles I and III, but also on local evaluations for titles I, III, and VII. This totals in excess of \$42 million--this is a substantial amount.

## CHAPTER 5

### USING STANDARDIZED TESTS

State and local education agencies generally use standardized "norm-referenced" tests to measure the effectiveness of Federal elementary and secondary education programs. OE's national evaluations also frequently use these tests.

Most State and local respondents to our questionnaires believe there is a substantial or very great need for increased efforts to develop alternatives to standardized norm-referenced tests, such as "criterion-referenced tests. Many of these officials also believe that increased efforts are needed to reduce racial, sexual, and cultural biases in tests. The questionnaire results indicated a lack of awareness among State and especially local agency officials on information available to help them select appropriate standardized tests.

### USES AND IMPLICATIONS OF STANDARDIZED NORM-REFERENCED TESTS

Standardized norm-referenced tests were devised to measure the status of an individual in relation to other individuals--the norm group. The score an individual receives has meaning in relation to the performance of the norm group, not the educational objectives involved; therefore, such tests are described as norm-referenced.

A standardized norm-referenced test differs from other tests given by schools in that it (1) is almost always constructed by specialists in educational testing, (2) has explicit instructions for standard or uniform administration, and (3) has norms for interpreting test results. These norms have been derived from giving the test to a sample of persons intended to represent the whole group for whom the test is designed.

There are many kinds of standardized tests given in schools, business, and the military services: intelligence, academic aptitude, achievement, personality, attitude, interest inventory, and vocational aptitude tests. In discussing tests, this report deals almost exclusively with achievement tests--those which measure current knowledge or competencies. Most of the standardized tests that children take in school are of this kind. Today the typical school-child takes one to three such tests every year.

Five or six companies account for about three-fourths of the total test sales in the country. These companies have all been in the testing business for a long time; most helped pioneer the testing field in the 1920's. The companies all sell a wide variety as well as a large volume of tests (most list more than 100 tests in their catalogues), and provide extensive services to test customers. These factors contribute to the widespread use and acceptance of the tests.

Standardized norm-referenced achievement tests are used to evaluate both individuals and programs. For student evaluations, the tests are used to rank or compare students for such purposes as counseling them, assigning them to a class within a grade or a group within a class, assigning students to a special program (for example, for the retarded or gifted), indicating the kind of courses a student may take in junior high school or high school, or gaining admittance to college or graduate school. In relation to program evaluations, standardized tests are used widely at the Federal, State, and local levels, to determine the effectiveness of OE programs such as titles I, III, and VII of the Elementary and Secondary Education Act.

#### WIDESPREAD STATE AND LOCAL USE OF STANDARDIZED NORM-REFERENCED TESTS TO EVALUATE FEDERAL PROGRAMS

Most OE-funded evaluations of Federal education programs at the national, State, and local levels have at least one common purpose: they are to measure and report on the effectiveness of federally funded programs and projects. Many State and local evaluations are congressionally mandated, and national evaluations, according to OE evaluation officials, are conducted either because of a congressional request or to provide responsible agency officials and congressional members with nationally comparative data about a particular program or approach to education. These national evaluations frequently use standardized tests.

The General Education Provisions Act (20 U.S.C. 1231a) requires OE to collect information intended to objectively measure the effectiveness of education programs and permits local education agencies to use systematic measurement approaches, approved by the Commissioner of Education, that will assure adequate evaluation of each program.

Our questionnaire asked State education agencies which of several techniques they employed for their 1973-74

evaluations of projects funded through titles I and III of the Elementary and Secondary Education Act. The following table shows the results for respondents in the 48 States that answered this question:

State Education Agency Techniques  
for Evaluating Titles I and III

<u>Techniques</u>	<u>Title I</u>		<u>Title III</u>	
	<u>Number</u>	<u>Percent (note a)</u>	<u>Number</u>	<u>Percent (note a)</u>
Aggregation and analysis of data from local education agency reports	45	94	41	85
Educational audits and their results	9	19	27	56
Statewide testing of students	11	23	5	10
Other	5	10	12	25

a/Does not total 100 percent because some States used more than one technique.

The distribution of responses shows that some States used a combination of techniques, and most States aggregated local education agency data. Only a small number of States employed statewide testing to evaluate their programs. However, the great majority of those that did test statewide in their program evaluations used standardized norm-referenced tests.

Because most States perform their evaluations by aggregating and analyzing data from local education agency reports, it is very important that local education agencies use test measures which reflect meaningful and comparative data. Our questionnaire asked local education agencies to indicate which types of tests they used in their 1973-74 evaluation of Federal programs funded through the Elementary and Secondary Education Act, titles I, III, and VII. The following table shows their responses.

Types of Tests Used by  
Local Education Agencies (note a)

<u>Type of Test</u>	<u>Title I</u>		<u>Title III</u>		<u>Title VII (note b)</u>	
	<u>Number</u>	<u>Percent</u>	<u>Number</u>	<u>Percent</u>	<u>Number</u>	<u>Percent</u>
Standardized norm- referenced tests	8,103	94.4	1,380	79.1	357	90.7
Criterion- referenced tests (note c)	2,110	24.6	424	24.3	148	37.5
Other tests	1,168	13.6	515	29.5	91	23.0

a/Local education agency figures are projected on the basis of a statistical sample of local agencies. (See app. II.) The percent columns do not total 100 percent because some local agencies used more than one type of test.

b/Title VII provides for direct OE grants to local agencies without State level administration.

c/Tests which are designed and scored in relation to specific learning objectives or behaviors and include an explicit statement of performance standards.

The local education agency responses show predominant use of standardized norm-referenced tests for program evaluation, but also relatively large use of criterion-referenced tests

Frequent use of criterion-referenced tests was also reflected in State education agencies' responses to a question on the use of statewide assessment programs to achieve educational accountability--an effort separate and distinct from testing used to evaluate Federal programs. Of 47 States responding to this question, 36 said that criterion-referenced tests have been or will be used extensively in that context.

EFFORTS TO EVALUATE  
STANDARDIZED TESTS

Since its establishment, NIE's major research effort for improving educational measurement has been an ongoing evaluation of commonly used standardized tests. In 1972 NIE

assumed from OE the sponsorship of the Center for the Study of Evaluation at the University of California at Los Angeles. The Center extensively assesses published standardized tests and as a primary objective issues reliable guides for use in selecting tests.

As of September 1975, the Center had published guides which evaluate preschool and kindergarten, elementary, and secondary school achievement tests. Each guide contains a compendium of tests, keyed to educational objectives and evaluated by measurement experts and educators for such characteristics as meaningfulness, examinee appropriateness, administrative useability, and quality of standardization.

A Center official said that he believes that commercially available standardized norm-referenced achievement tests are generally inappropriate for measuring program effectiveness. Various Center reports also state this opinion for reasons discussed in chapter 6, such as the tests' low degree of correspondence with actual instructional objectives, their failure to indicate the extent that the full range of instructional objectives has been mastered, problems in test administration, and the test's lack of information on specific skill and knowledge development. The Center official noted that the tests are especially inappropriate for broad-based intervention programs, such as title I of the Elementary and Secondary Education Act.

Since the major purpose of the guides is to provide State and local educators and other test consumers with information that will assist them in selecting the most appropriate measurement devices for evaluations, our questionnaire included a question relating to the Center's work. We asked State and local education agencies to indicate the degree of familiarity they had with the Center's research on evaluating the utility of many popular commercially available standardized norm-referenced tests. Their responses follow:

<u>Responses</u>	<u>State education agencies</u>		<u>Local education agencies</u>	
	<u>Number</u>	<u>Percent</u>	<u>Number (note a)</u>	<u>Percent</u>
Have little or no information	6	12	5,059	85
Aware of the Center's work in test evaluation	14	28	601	10
Read some of the Center's publications	17	34	192	3
Used the Center's material in the selection of commercially available standardized tests	<u>13</u>	<u>26</u>	<u>135</u>	<u>2</u>
Total	<u>50</u>	<u>100</u>	<u>5,987</u>	<u>100</u>

a/Local education agency figures are projected on the basis of a statistical sample of local agencies. (See app. II.)

There was general agreement among the educators and test publishers interviewed that the Center's work was a definite improvement over other existing evaluations of standardized tests. Some test publishers, however, criticized the Center's work because it was based on incomplete data, excluding, for example, the highly technical specifications from which test questions are developed. They also criticized it for establishing educational goals on which test ratings were based without consulting publishers, and for relying on the work of graduate students.

In response, Center officials indicated that they had sent letters to test publishers advising them of the study and its purpose, stating what information was desired, and asking for any other material the test publishers wished to send. They acknowledged that their curricular goal system is not perfect; however, they believe it is a tremendous step forward, providing a clear statement of expected student behaviors which test users can employ to match tests to curriculums. Officials also stated that the system is justified because test makers often measure skills that are different from what they claim to measure. Center officials also noted

that all participating graduate students demonstrate adequate competence in research and measurement before being hired; they followed set procedures, were routinely monitored, and discussed questionable points with supervisors.

State education agency officials and measurement experts interviewed stated that the Center's work should be continually updated and should include an assessment of available criterion-referenced tests.

Commenting on our report, HEW stated that NIE is sponsoring an effort by the Center to prepare a new test evaluation book reviewing all commercially published criterion-referenced achievement tests for grades kindergarten through 12. The target date for publishing this book was June 1977; however, as of August 1, 1977, it had not been published.

#### CRITERION-REFERENCED TESTS

Our questionnaire asked State and local education agencies to indicate areas, if any, in which there is a need to increase the educational community's efforts devoted to measurement and assessment techniques. Over 75 percent of the State education agency respondents and about 51 percent of school district respondents indicated "development of alternatives to the classic standardized norm-referenced tests (e.g., criterion-referenced tests)" as an area needing substantial to very great increases. More than half of the State respondents and a third of the local respondents also believe there is a great need to reduce racial, sexual, and cultural biases in tests. The following table shows the percent and number of State and local agency respondents indicating a substantial or very great need to increase the educational community's efforts devoted to measurement and assessment techniques.

	State agency responses		Local agency responses	
	Percent (note a)	Number	Percent (note a)	Number (note b)
Development of methods for test design and construction	52	25	34	2,507
Reduction of cultural, racial, and sexual biases in tests	51	25	35	2,581
Development of alternatives to the classic standardized norm-referenced tests (e.g., criterion-referenced tests)	77	37	51	4,061
Development of more and improved standardized norm-referenced tests	28	9	27	2,177
Development and utilization of methods to better evaluate standardized norm-referenced tests in use	40	20	43	3,409

a/Does not add up to 100 percent because more than one item could be checked. The percentages reflect only the number of State and local agencies that responded to each suggested item. (See apps. I and II, question 3.)

b/The number of local agencies is projected on the basis of a statistical sample. (See app. II.)

NIE is currently funding research intended to meet these needs in the areas of test biases and criterion-referenced measurement.

Criterion-referenced tests are designed to remedy some weaknesses in standardized norm-referenced tests (see ch. 6) by (1) being more accurately interpretable, (2) detecting the effects of good instruction, and (3) allowing more accurate diagnoses of the individual learner's capabilities.

A well-devised criterion-referenced test relates scores to specific learning objectives or behaviors and includes an explicit statement of performance standards. The objectives must be described without ambiguity to permit an accurate description of what an examinee does and does not know or can and cannot do. Criterion-referenced tests pinpoint the student's deficiencies, while a norm-referenced test identifies only general student weaknesses.

Because criterion-referenced tests are not required to yield large variances in examinee's scores, they can retain questions based on the primary curricular emphasis even if, after instruction, most learners answer them correctly. Consequently, criterion-referenced tests are considered more capable of discerning instructional effects than norm-referenced tests.

An increasing number of educators have begun to question the use of standardized norm-referenced tests and to propose criterion-referenced tests as an alternative. Some people assume that the latter are fully developed and ready to use. According to experts, however, the technical status of criterion-referenced measurement is far less advanced than many educators and others believe it to be.

An expert in criterion-referenced testing from the University of Michigan has pointed out that producing test questions that can be defended as valid and fair for both the majority of examinees and various minority groups is a major problem that affects the development of both standardized norm-referenced and criterion-referenced tests.

Another expert in criterion-referenced testing from the California Test Bureau of McGraw-Hill has stated that carefully constructed criterion-referenced tests can provide both diagnostic and evaluative information that is appropriate not only for disadvantaged students but also for all students, assuming some consensus can be obtained on instructional objectives to be tested. Therefore, such tests could discover real educational problems and indicate appropriate remedial help for students, rather than simply showing that the students are "below grade level" on a general test of reading or mathematics. He added, however, that constructing such criterion-referenced tests is not a simple matter. More knowledge is needed about the structure of subject matter than now exists. Nevertheless, successful statewide assessments and evaluations have been carried out using only criterion-referenced tests.

This testing expert believes that in the future it will be possible to evaluate basic skills on a large scale using appropriate criterion-referenced tests when sufficient knowledge has been acquired to specify the important skills and subskills required to assure competence in these areas. For this to occur, an understanding of the particular problems facing disadvantaged and minority students as well as the basic logical and cognitive structure of disciplines will be needed. In his opinion, progress is being made on these problems, but as yet no widespread consensus of what is important has emerged. Until it does, no national criterion-referenced evaluations seem likely.

Commenting on our report, HEW stated that, based on the test evaluation being made by the Center for Study of Evaluation at the University of California at Los Angeles commercially published criterion-referenced tests, like many norm-referenced tests, were generally unsatisfactory for program evaluation.

## CONCLUSIONS

Federal, State, and local education agencies frequently use standardized norm-referenced tests to measure the effectiveness of Federal education programs. The next chapter provides further discussion of some of the tests' problems, including test biases. Additional research may be needed on:

- Criterion-referenced tests and other alternatives to standardized norm-referenced achievement tests, for uses which include program evaluation.
- How to reduce racial, sexual, and cultural biases in standardized tests.

There is also a need to increase State and especially local education agency awareness of available NIE-funded information that is intended to help select the most appropriate tests for use in evaluation.

## RECOMMENDATIONS TO THE SECRETARY OF HEW

We recommend that the Secretary direct NIE to:

- Consider the need for funding additional research in the future on (1) criterion-referenced tests and other alternatives to standardized norm-referenced achievement tests for uses that include program evaluation and (2) the nature and extent of racial,

sexual, and cultural biases in standardized tests and how such biases may be reduced.

- Improve dissemination of available NIE-funded information, which is intended to help select the most appropriate standardized tests, thereby increasing State and local education agency officials' awareness and use of this information.

#### AGENCY COMMENTS AND OUR EVALUATION

HEW agreed with the above recommendations. Regarding our recommendation that NIE consider the need for funding additional testing research, HEW described efforts currently underway and stated that more emphasis will be given to these programs in fiscal years 1977-79 if NIE's appropriation levels permit. However, HEW also stated that it is not clear whether we believe NIE deserves additional appropriations for such an effort since NIE cannot divert substantial funds from its present budget. It is not our intention to call for either additional appropriations or redirection of NIE's present budget. Our recommendation calls on NIE to consider funding additional research, as needed to address the problems discussed, that could begin when research efforts currently underway are completed.

Regarding our recommendation that NIE improve dissemination of its materials designed to help select standardized tests, HEW stated that NIE intends to make school personnel familiar with these materials through several dissemination approaches. These include:

- Listing the test evaluation consumer guides in a catalog of NIE products sent to school superintendents and district curriculum directors and which will now be sent to school district evaluation directors.
- Using the new "Lab and Center R&D Exchange," which HEW believes will possibly reach 50 percent of the country's school systems.
- Using the dissemination network formed by NIE's seven research and development utilization contractors.

## CHAPTER 6

### STANDARDIZED TESTS AND PROGRAM EVALUATION

Serious questions have been raised about standardized norm-referenced achievement tests in spite of their wide-spread use. Based on the views of test and evaluation experts and others, this chapter provides information concerning: (1) criticism and defense of these tests and (2) suggestions for alleviating problems in conducting large-scale evaluations of compensatory education and desegregation programs.

#### CRITICISM OF STANDARDIZED NORM-REFERENCED TESTS

The content and use of standardized norm-referenced achievement tests have been widely criticized by testing experts, educators, and others. The National Education Association, the National Association for the Advancement of Colored People, and others have called for moratoriums on using standardized tests. 1/

The National Association of Elementary School Principals convened the appointed representatives of 25 national educational associations, government agencies, and education groups in November 1975 to explore the implications of the widespread use of standardized achievement tests. They recommended that the educators give higher priority to developing and using new assessment processes that are more fair and effective than those currently in use and that educators more adequately consider the diverse talents and cultural backgrounds of children.

#### Criticism of norm-referenced tests

Critics of the standardized norm-referenced achievement tests believe test bias and score interpretation, as well as other problems, are some of the test's deficiencies.

---

1/Robert L. Williams, et al., "Critical Issues in Achievement Testing of Children from Diverse Ethnic Backgrounds," prepared for the Office of Education's Conference on "Achievement Testing of Disadvantaged and Minority Students For Educational Program Evaluation," May 27-30, 1976, pp. 8-12.

### Test bias

A frequent criticism is that the tests discriminate unfairly against racial and cultural minorities because (1) their norms continue to be based on populations not representative of a pluralistic, multicultural society and (2) test questions ask for knowledge most familiar to the white middle class or reflect the cultural biases of test developers and question writers, who mostly represent the white middle class. Therefore, the tests are considered biased against the poor, black, Hispanic, and other minority Americans. Although some test publishers have attempted to minimize cultural and racial biases, these problems have not been solved.

### Score interpretation

Some critics cite certain weaknesses in the interpretation of test scores, as follows:

- Raw scores (scores based on the total number of correct answers on a test) are typically interpreted in terms of national norms, which are estimates of nationwide performance. The national norms are derived from giving the test to what is intended to be a representative sample of students. But since the samples are different and taken at different times, the norms for different tests vary. As a result, the normed score for a student depends partly on which test the student takes. 1/
  
- The grade equivalent score represents the estimated average score that pupils in that month of that grade would achieve on the test nationwide. For example, a 3.8 in reading is the average score for a child in the eighth month of the third grade. A frequent misconception is that the score means the child has mastered the standard curriculum up to that point in schooling. Even if the 3.8 grade-equivalent were always an accurate estimate of the average achievement for the child in the eighth month of the third

---

1/George Weber, "Uses and Abuses of Standardized Testing in the Schools," Council for Basic Education, May 1974, pp. 13-14.

grade, that average child has not necessarily "mastered" the reading curriculum to that point, although parents, and even teachers usually do not understand this. Another problem is that on some tests a few answers one way or the other can make as much as a whole year's difference in the grade-equivalent score, and the tests are simply not that accurate. Despite these and other shortcomings, grade-equivalent scores are usually used in interpreting the results of standardized achievement tests. 1/ 2/

--Test scores are meaningful only in terms of national average achievement, and do not indicate whether this is good, bad, or indifferent in terms of "reasonable" standards defined independently of such average scores. For example, if a given third grade class does as well on a given reading test as the national third grade average, this does not reveal how well the children can read in absolute terms. According to this viewpoint, since reading achievement in the primary grades is generally below what could reasonably be accomplished, reading scores suggest a better achievement than is in fact the case. 3/ Moreover, critics say that norm-referenced tests result in half of the students being above the norm and half below, as a statistical fact of life. Then how, they ask, does one "raise scores to the general norm?" 4/

---

1/Ibid., pp. 14 and 16.

2/Ralph Tyler, "Discussion of Hoepfner's Paper on Achievement Test Selection for Program Evaluation," prepared for the Office of Education's conference on "Achievement Testing of Disadvantaged and Minority Students For Educational Program Evaluation," May 27-30, 1976, p. 7.

3/Weber, p. 20.

4/Miriam Clasby, et al., "Laws, Tests, and Schooling: Changing Contexts for Educational Decision-Making," RR-11, Educational Policy Research Center, Syracuse University Research Corporation, Syracuse, N.Y. Oct. 1973, p. 174.

Other problem affecting test use

Regarding the use of standardized norm-referenced tests in program evaluation, critics state that the tests are inadequate for evaluating program effectiveness and cite the following deficiencies.

Test results not specific--Normative test scores might be very useful for program evaluation if one knew what they meant. But the test results do not indicate specifically what students have learned. 1/ The test scores reflect the number of correct answers students give, but do not indicate how well students achieve intended educational objectives or whether they answer particular questions correctly. 2/ 3/ Therefore, the test results are too general to provide specific guidance for improving the quality of schooling. 4/

Tests do not coincide with instructional objectives--The tests' content often has a low degree of correspondence with actual instructional objectives at any given time or place. This is a serious deficiency because one cannot determine the effectiveness of an educational program unless the tests used actually measure the objectives that the teacher, teachers, or school district is attempting to accomplish over

---

1/Stephen Klien, "Evaluating Tests in Terms of the Information They Provide," Evaluation Comment, Vol. 2, June 1970, Center for the Study of Evaluation, University of California at Los Angeles, p. 2.

2/National Assessment of Educational Progress, "General Information Yearbook," Dec. 1974, Rept. No. 03/04-GIY, pp. 1, 3, and 4.

3/Carmen J. Finley, "Not Just Another Standardized Test," Compact Vol. 6, No. 1, Feb. 1972, Education Commission of the States, pp. 10 and 11.

4/W. James Popham, "Appropriate Assessment Devices for Educational Evaluation," presented at the National Forum on Educational Accountability in Denver sponsored by the Office of Education and the Cooperative Accountability Project, May 8-9, 1975, pp. 3 and 4.

a defined period of time. 1/ As a major test publisher 2/ and others have pointed out, it is only to the extent that a program's instructional objectives coincide with those of the test that the instrument is valid for measuring how well the learning program has succeeded. If the test fails to measure certain objectives included in the learning program and/or measures other objectives that are not part of that program, to that extent the test is not a valid measure of success in the program.

In addition, test publishers describe their standardized norm-referenced tests in very general terms, calling them, for example, tests of reading comprehension. The generality of these descriptions increases the possibility of unrecognized differences between what the schools teach and what the tests specifically measure. According to this view, such differences result in misleading data and false conclusions about program effectiveness. 3/

Tests are designed to differentiate students, not diagnose specific problems--Since the tests are intended to compare examinees, they must yield a reasonably large degree of "response variance"--different scores for different examinees. Test questions that are answered correctly by half the examinees maximize a test's response variance. If a test question is answered correctly by a large or increasing proportion of examinees, it tends to be removed from the test or modified. Thus, as norm-referenced tests are periodically revised, questions on which examinees perform well are systematically eliminated. Yet, test critics maintain that such questions often deal with the very concepts teachers thought important enough to emphasize in their instruction. If a concept is taught well, questions measuring it will likely be removed in the next test revision. The

---

1/Rodney Skager, "The System for Objectives-Based Evaluation-Reading," Evaluation Comment Vol. 3, No. 1, Sept. 1971, Center for the Study of Evaluation, University of California at Los Angeles, p. 6.

2/J. Wayne Wrightstone, et. al., "Accountability in Education and Associated Measurement Problems," Test Service Notebook 33, Issued by the Test Department, Harcourt Brace Jovanovich, Inc., New York, p. 4.

3/Popham, p. 3.

result is that (1) the tests are particularly insensitive to detecting the effects of instruction 1/ 2/ and (2) sometimes the tests do not contain test questions dealing with the central concepts in the field. These are serious deficiencies for a test used for program evaluation. 1/ 3/

Teaching the test--According to some critics, the use of standardized achievement tests for program evaluation and "accountability" has led and will lead to corruption and dishonesty among educational professionals and to the further erosion of public trust in the schools and the people who run them. Faced with public pressure that is often in itself "irrational and destructive," it is all too easy for principals and teachers to respond to subtle pressure to "prepare students for the assessment" by teaching students responses to specific test questions rather than by developing the underlying skills which these questions reflect. Standardized tests are readily available at all levels of any school system, and are brief enough to be highly susceptible to coaching. Test security and control may be feasible for programs like the college boards or the American College Testing Program, in which representatives of the testing agency handle the assessment and the examinees come to a central location. Similar controls are not feasible in large-scale evaluations of school programs. 4/

Inappropriate norms--In the typical large-scale program evaluation, the focus is on performance of groups by students--categorized by classes, buildings, or school systems--not by individuals. The reference norms one needs for such purposes are distributions of averages for appropriate reference schools, not norms for individuals; and these types of

---

1/Ibid., pp. 4 and 5.

2/Richard M. Jaeger, "A Discussion of Classical Test Development Solutions," prepared for the Office of Education's conference on "Achievement Testing of Disadvantaged and Minority Students For Educational Program Evaluation," May 27-30, 1976, p. 6.

3/W. James Popham, Statement presented at U.S. House of Representative hearings on the Elementary and Secondary Education Act, March 28, 1973, pp. 2323 and 2324.

4/See Skager, p. 7.

norms are seldom available unless they are collected as part of the evaluation study itself. 1/

### Measurement of growth

A major test publishing company has described 2/ various technical problems with the tests that are related to their use in measuring academic growth, as contrasted against their traditional use for measuring present status. This is significant, because according to the company most educational program evaluations have involved using nationally standardized norm-referenced achievement tests--especially for measuring "growth." Besides the problems related to using tests to measure present status--such as selecting a test that measures what the user intends, assuring that teachers follow directions, and the like--the test publisher noted that when the tests are used to measure growth they are attended by special problems, such as the following:

- Defining "normal growth." There are serious questions about the legitimacy of defining normal growth in terms of grade-equivalent scores. However, expected or normal gain is almost universally defined in terms of grade equivalents for standardized achievement tests used at the elementary level.
- Interpolating or estimating norms so that they may be applied to tests taken at times during the year for which norms have not been empirically determined. These estimates are almost certainly in error by some small amount in most cases and by a substantial amount in some cases.
- Converting scores from different levels and alternative forms of a standardized test series so that the scores are equivalent. If they are not equivalent, this can lead to invalid measurement of gains and possibly erroneous conclusions as to the merit of the program evaluated.

---

1/William E. Coffman, "Classical Test Development Solutions," prepared for the Office of Education's conference on "Achievement Testing of Disadvantaged and Minority Students For Educational Program Evaluation," May 27-30, 1976, p. 24.

2/Wrightstone, et. al., pp. 5-12.

The above criticisms of standardized norm-referenced tests indicate that the tests may be inappropriate for wide-scale use in evaluating educational programs.

Why are standardized tests used?

Considering all these criticisms, why are the tests used for Federal program evaluations? A former OE evaluation official explained that standardized achievement tests are used in educational program evaluation "for many good and not-so-good reasons" such as the following:

--Since many standardized achievement tests or subtests were developed primarily for basic skill performance measurement, they become prime candidates for evaluating programs that seek to improve basic skills.

--Such tests are readily available in large quantities, at short notice, and at relatively low cost. If off-the-shelf tests were not available, the cost of developing and standardizing such measures for a specific evaluation might be prohibitive and therefore might cause abandoning evaluation plans.

Other, more technical reasons given by this former OE official for the widespread use of standardized achievement tests in education program evaluation include their general technical excellence, their standardized administration procedures, the representativeness of their questions to the possible universe of questions on basic skill performance, their normative reference, ease in scoring, alternative and equated test forms, high reliabilities, and apparent validity. Another factor, he stated, is that most achievement tests are part of a battery of tests designed so that student growth can be measured as the student progresses through school by administering different test levels and forms. 1/

---

1/Michael J. Wargo, "An Evaluator's Perspective," prepared for the Office of Education's conference on "Achievement Testing of Disadvantaged and Minority Students For Educational Program Evaluation," May 27-30, 1976, pp. 19-21.

## DEFENSE OF STANDARDIZED TESTS

Defenders of the tests say that there is no convenient alternative for those outside the schools to evaluate students' collective achievements. Therefore, in their view, despite the tests' shortcomings and abuses, they provide the best information available. According to OE's Assistant Commissioner for Planning, Budgeting, and Evaluation, it is true that standardized tests do not allow for program and project differences, but the Congress wants to know if the overall program is effective, and he believes the tests provide this information acceptably.

Defenders of standardized tests tend to emphasize test misuse as the major problem, and they express the need for training teachers and administrators in proper test selection, administration, and interpretation. Many testing experts admit that the tests have deficiencies--such as test and test question bias--but state that wholesale rejection of the tests and their norm-referenced interpretations is unwarranted. Instead, they favor refining the tests and learning to avoid pitfalls, such as lack of congruence among (1) test content, (2) course content or curricular emphasis, and (3) the purpose and design of the evaluation. Some state that progress has been made in the last 10 to 15 years in such areas as

- constructing efficient tests that reliably measure important educational skills,
- developing nationally representative norms, and
- providing test users with relevant information about the test areas. 1/ 2/

---

1/John C. Bianchini, "Achievement Tests and Differentiated Norms," prepared for the Office of Education's conference on "Achievement Testing of Disadvantaged and Minority Students For Educational Program Evaluation," May 27-30, 1976, pp. 36-37.

2/Ralph Hoepfner, "Achievement Test Selection For Program Evaluation," prepared for the same conference, p. 2.

One expert from Systems Development Corporation, for example, stated that the quality, forthrightness, and focus of standardized achievement tests have improved remarkably, particularly within the last decade, and that many well-aimed attacks on standardized achievement tests made about 10 years ago are no longer valid. 1/ This expert, who has considerable experience in evaluating test adequacy, stated that his review of available criterion-referenced tests--the most likely alternative--found them to be of uniformly bad quality.

An expert from RMC Research Corporation said that the standardized tests are not the problem. In his view, OE's large-scale national evaluations as well as most local evaluations have been poorly designed and poorly done, and the tests have often been misused by evaluators. He said that he recently spent about 2 weeks in every State working on title I, dividing his time equally between the State education agencies and some local education agencies in each State. Based on this experience, he believes that the great majority of local title I projects is not providing students with educational treatments that differ in any significant way from regular classes. In such circumstances, the general lack of evidence of marked improvement in basic skills should not be too surprising.

#### Suggested improvements

Included among suggestions for improvement offered by test defenders are the following:

- Test publishers, in developing standardized tests of basic skills, should break away--at least in the elementary grades--from the current practice of designing tests for measuring achievement at multiple grade levels. Test publishers should develop series of tests, each designed for a specific grade, with sufficient numbers of questions at various difficulty levels to yield reliable measurement for essentially all students at that grade.
  
- Test publishers should provide more detailed information about the content of test questions, the instructional objectives on which questions are based, and the skill characteristics needed to answer them to

---

1/Ibid., p. 2.

provide test users with a general framework for assessing the logical congruence between the test content and the content of the curriculum. In one expert's view, such a detailed classification of individual questions is as important to the test user as is the extensive statistical data currently provided about the mental measurement characteristics of the questions.

- Test publishers should expand the services they provide their clients to include developing special norms when they would produce more appropriate use of test results. Test publishers need to be more active in assuring that their tests and subsequent test results are used fairly and effectively.
- The state of the art should be extended to provide test users with practical procedures to assist them in selecting tests and relating test results to instructional programs and program evaluation.
- Program evaluators should recognize that the process for selecting appropriate standardized tests for evaluation must go beyond a naive inspection of the test and normative data. The process ought to include a careful inquiry into such elements as relevant student and school characteristics, test content and its relationship to curriculum content, and the adequacy of normative data in relation to the evaluation design. 1/ Evaluators should select tests which maximize coverage of the objectives desired. 2/

#### CONFERENCE ON USING TESTS TO EVALUATE PROGRAMS

In May 1976 OE sponsored a special four-day conference on "Achievement Testing of Disadvantaged and Minority Students for Educational Program Evaluation." OE invited about 50 experts in testing, program evaluation, and related fields including university and other researchers, and representatives of leading test publishers. Federal and local education agencies, and education and other interest groups were also represented. The conference focus was on large-scale

---

1/Bianchini, pp. 36-38.

2/Hoepfner, p. 33.

program evaluations of elementary and secondary school compensatory education and desegregation programs--programs on which OE concentrates most of its effort. The purpose of the conference was to identify, define, and analyze the many problems associated with using standardized achievement tests in the context of large-scale evaluations of these programs and to develop interim and long-term solutions to those problems.

Conference participants mentioned many of the same problems and issues discussed previously in this chapter. Five small working groups were formed at the close of the conference to write recommended solutions. All five groups recommended (1) either limiting or ceasing large-scale Federal education program evaluations like those contracted for by OE's Office of Planning, Budgeting, and Evaluation and/or (2) placing greater emphasis on local evaluations.

The value of conducting large-scale evaluations was questioned because of problems such as the current state of the art in evaluation; inherent bias problems in data collection instruments, methods, and analysis procedures; and community differences in the education programs being evaluated. Suggestions for increased emphasis on local level evaluations included providing funding for adequate local evaluations, training local evaluators, and technical assistance in designing and implementing technically sound evaluations.

Related conclusions and recommendations included the following:

1. The Federal Government's basic policy for evaluating its education programs for culturally different students should require that each local education agency carry out evaluation studies designed to assess how well its local project objectives have been achieved. Also, approved local agency budgets should include sufficient funds to provide for adequate evaluation study design, data collection, analysis, and reporting. The studies should involve at all stages the participation of members of the minority culture or cultures involved. Beyond this, Federal responsibility should be limited to (a) conducting and publicizing the results of audits that determine whether funds were used as intended and whether evaluation data relevant to program objectives were collected, analyzed and reported; (b) providing general guidelines and training

for evaluation, and encouraging the development of guidelines and consulting resources by State agencies and Federal regional offices; and (c) developing summary reports based on the aggregation of information from local evaluations.

2. To resolve Federal program evaluation problems:

- The funds and activities devoted to large-scale evaluation should be immediately rechanneled into development of program evaluation methods and tools that are likely to be more productive, while affording safeguards for recipient populations.
- Studies should be initiated to explore whether it is feasible to draw overall program impact conclusions based on aggregations. Valid, locally relevant project evaluations should also be started.
- Immediate congressional needs for program impact information should be satisfied through careful and extensive analysis in phases of data already collected and data currently being gathered under contract.
- Model local project evaluations should be funded as demonstrations, with support from experts funded by OE's Office of Planning, Budgeting, and Evaluation. This support should be provided to local school systems on a cooperative basis. Local evaluation personnel should be trained. The most effective local evaluation strategies demonstrated should be adopted in an ever-widening pattern, to build a basis for effective national program evaluation by aggregating valid local project evaluations.
- Studies should be funded on tools and procedures needed to make local evaluations that will reflect valid conclusions on the worth of specific projects and will adequately identify the processes and inputs of those projects. This includes holding a workshop to identify needed tools, methods, and priorities.
- Studies should be funded on OE and State education agency regulations, guidelines, and administrative decisions that affect the quality of local education agency evaluation activities and reports. These

studies should produce model administrative and regulatory strategies, including incentives, for upgrading the quality and validity of local education agency evaluations of federally supported projects.

3. One reason for greater emphasis on local evaluations is that instructional treatments are not uniform under nationwide programs. The Emergency School Aid Act program, for example, does not provide uniform treatments for students. Neither do title I or title VII programs. But there are various identifiable, describable, instructional treatments funded by these programs and a small number of them are effective. However, information about these effective treatments is lost in large-scale evaluations that cover a large number of ineffective treatments.
4. Guidelines for local evaluation studies should include the following:
  - Recommendations that each evaluation report include a description of what actually happened to pupils involved in the program. Without such information it is impossible to reach a meaningful interpretation of any measures of change.
  - Encouragement to local projects to collect evidence of progress toward improved skills in reading and mathematics and obtain data regarding other outcomes of the particular methods employed, particularly on developing self-concept and interpersonal relations.
  - Encouragements to local education agencies to include in their evaluation procedures systematic attention to the selection or development of techniques designed to minimize cultural bias in tests and other data-collecting procedures used in evaluation.
5. Money saved by curtailing large-scale evaluations should support a national panel responsible for exploring and developing more responsive and effective alternative program evaluation models.

6. Research into alternative evaluation models should also be conducted to determine (a) the evaluation information the Congress, OE, other Federal agencies, and local agencies receiving Federal funds need and (b) whether adequate evaluation designs can be effectively implemented to meet these needs, and their cost.
7. Because of conflicts in the interpretation of off-the-shelf commercial standardized tests, only custom-designed tests should be used for federally sponsored large-scale program evaluation, when such large-scale evaluations are essential. The associated cost and effort to define program objectives and define measures of their effect should be part of the Federal agency's responsibility along with survey and analysis costs.

Other small working group conclusions and recommendations are shown in appendix IV.

In response to the conferee's recommendations, OE's Assistant Commissioner for Planning, Budgeting, and Evaluation said he agrees that there is a need for increased evaluation activity and capability at the State and local levels. He disagreed, however, that OE's nationally planned evaluations should be deemphasized.

## CONCLUSIONS

As noted in chapter 5, Federal, State, and local education agencies frequently use standardized norm-referenced achievement tests to measure the effect of Federal education programs. However, there is a great deal of disagreement among testing experts and educators on the adequacy of these tests for their intended uses. Serious questions have been raised about the tests, and some organizations have called for a moratorium on their use and a higher priority on development and use of alternatives.

Although the tests' critics and defenders agree that certain problems exist, views differ greatly about the importance or severity of the problems and their remedies. Those who defend the tests and their continued use recognize that improvements are needed on such issues as test and test question bias; appropriate test norms; test selection, interpretation, and administration, including uses for program evaluation; and other issues.

Some questions raised about the adequacy of the tests have great importance in determining the appropriateness, validity, and proper conduct of educational program evaluations, including those that are federally funded. Decision-makers should be aware of these issues when using such information.

## CHAPTER 7

### PREFERENCES FOR EVIDENCE OF

### PROGRAM EFFECTIVENESS DIFFER

Responses to our questionnaire showed that State and local education agency officials responsible for administering Federal title I, III, and VII elementary and secondary education programs perceive important differences in the type of evidence of program effectiveness that Federal, State, and local officials prefer.

- Local education agency officials believe that State and OE officials are predominantly interested in standardized norm-referenced test scores for demonstrating program results. Local officials themselves prefer broader, more diverse types of information than only these test scores.
- State education agency officials prefer criterion-referenced tests. They favor less emphasis on standardized norm-referenced test results as evidence of program effectiveness than they believe OE officials want, but prefer them more than local agency officials do.

OE's Assistant Commissioner for Planning, Budgeting, and Evaluation believes that hard objective data on students' cognitive improvements is the centrally important information needed. He noted that this means gain scores on standardized norm-referenced achievement tests because these are most available.

As noted in chapter 5, State and local evaluations for titles I, III, and VII have been most often based on standardized norm-referenced tests. Therefore, evaluations usually have not reflected the kinds of results that State and local officials themselves prefer, but rather those that they believe would be likely to most impress higher level officials.

STATE AND LOCAL EDUCATION OFFICIALS BELIEVE  
STANDARDIZED TEST RESULTS MOST  
IMPRESS HIGHER LEVEL OFFICIALS

Legislative and other requirements for annual evaluations at the State and local levels of titles I, III, and VII of the Elementary and Secondary Education Act have several purposes, according to OE's Assistant Commissioner for Planning, Budgeting, and Evaluation. These include

- reporting on local project effectiveness or providing other data, on which local, State, and Federal officials can base programmatic, financial, and policy decisions; and
- providing data to State and/or Federal officials to select successful and exemplary projects for dissemination.

Educators must increasingly evaluate their instructional endeavors and present evidence which will permit others to judge their effectiveness. Through our questionnaire we have attempted to ascertain (1) what kinds of evidence of program effectiveness local officials administering federally supported elementary and secondary education programs think State and OE program officers expect, (2) what kinds of evidence of effectiveness State officials administering the programs think OE program officers expect, and (3) what types of evidence of effectiveness the State and local officials think is most useful to them.

Our questionnaire asked

- State program officials to rate, on a seven point scale, the types of information or findings which most impress them and, they believe, OE officials, as demonstrating program results and
- local project officials to rate on the same seven point scale the types of information or findings which most impress them and, they believe, OE or State officials as demonstrating program results.

As noted in chapter 5 (see pp. 43 to 44), only a few States employed statewide testing to evaluate their title I and III programs, and the great majority of these States used standardized norm-referenced tests. Most State evaluation reports for titles I and III were aggregations of local education agency evaluation data. As shown on page 45,

based on our questionnaire sample responses, local title I, III, and VII projects indicated overwhelming use of standardized norm-referenced tests--far greater than their use of criterion-referenced tests and all other tests.

State officials' views

The demonstration of program results should be an important factor in deciding to continue OE program funding. As the table below shows, State title I and III officials agreed on the types of program results they think are likely to impress OE officials most in this regard.

State officials overwhelmingly perceive OE officials to be most impressed by results obtained from standardized norm-referenced tests. Eighty-two percent of the State title I and 65 percent of the State title III officials responding ranked this category first. (See app. I for details.) State title I and III officials ranked results on improvements in educational management and accountability, as well as findings obtained from criterion-referenced tests, as either the second or third most impressive data for OE officials.

Titles I and III State Officials' Perceptions of  
What Impresses OE Officials and What Impresses  
State Officials (note a)

	Impresses OE		Impresses self	
	Title I	Title III	Title I	Title III
1. Improvement in educational management or accountability	3.1	3.9	4.0	4.5
2. Improvement in school services or facilities	6.0	6.1	5.9	6.1
3. Student improvements through gain scores or grades on teacher ratings	4.1	4.6	4.3	5.1
4. Student improvement through gain scores on standardized norm-referenced tests	1.3	1.6	2.5	3.3
5. Student improvement through gain scores on criterion-referenced tests	3.2	2.7	2.4	2.4
6. Student improvement through gains in the affective domain (e.g., likes, dislikes, interests, attitudes, motives, etc.)	5.3	4.4	3.7	3.1
7. Improvements in curriculum and instructional methods	4.5	4.5	3.8	3.4

a/The numbers shown are the respondents' average rankings of the alternatives given. Each respondent was asked to give the most impressive type of result a ranking of "1" and the next most impressive "2," etc.

State officials themselves appear to be impressed by slightly different types of data and information on program results. The table above shows that officials from titles I and III would be most impressed by gain scores on criterion-referenced tests. Although State title I officials ranked gain scores from norm-referenced tests almost equal to those from criterion-referenced tests, the title III respondents consider gains in the affective domain (for example, likes, dislikes, interests, attitudes, motives) as the second most important result, followed by gain scores from norm-referenced tests. State title I officials ranked gains in the affective domain third. Title I and III officials agreed on the remaining categories.

In general, State officials believe objective test results of program performance are most likely to impress both OE and State program officers. However, State respondents see themselves as more likely than OE program officials to be impressed by criterion-referenced test results and other factors cited above. State officials see themselves as more open than OE officials to various types of information demonstrating program results. At the same time, they agree with what they see as OE officials' view that impact evaluations based on test scores are the most impressive findings.

Emphasis on test scores for evidence of program effectiveness continues at the State and national levels. Reasons may include the following:

- Legislators and Federal and State officials are demanding evidence that the infusion of Federal and State dollars for special programs works.
- Test scores have traditionally been the only measure of effectiveness.
- Few alternatives to test scores exist and those that are available are unproven and not as widely accepted.

#### Local officials' views

The local title I and III respondents' perspectives are very nearly alike on what evidence of program results impresses State and OE officials. Title VII respondents' perspective is somewhat different; however, this may be at least partly because local title VII projects are responsible directly to OE program officers and not to State program officers.

The table below shows that local title I, III, and VII project officials feel that OE or State program officers are most impressed by norm-referenced test results.

Local Title I, III, and VII Officials' Perceptions  
of What Impresses State or OE Officials and  
What Impresses Local Officials (note a)

	Impresses OE or State			Impresses self		
	Title I	Title III	Title VII	Title I	Title III	Title VII
1. Improvement in educational management or accountability	3.9	3.8	3.8	4.7	4.5	4.4
2. Improvement in school services or facilities	5.2	5.1	4.8	4.9	4.9	4.8
3. Student improvements through gain scores or grades on teacher ratings	4.3	4.3	5.1	4.0	4.3	4.4
4. Student improvement through gain scores on standardized norm-referenced tests	2.2	2.6	2.1	3.7	3.7	3.4
5. Student improvement through gain scores on criterion-referenced tests	3.1	3.5	3.9	3.4	3.7	4.0
6. Student improvement through gains in the affective domain (e.g., likes, dislikes, interests, attitudes, motives, etc.)	4.6	4.2	4.4	3.3	3.3	3.4
7. Improvements in curriculum and instructional methods	4.2	3.9	3.7	3.2	3.0	3.2

a/The numbers shown are the respondents' average rankings of the given alternatives. Each respondent was asked to give the most impressive type of result a ranking of "1" and the next most impressive "2," etc.

Local title I and III project officials ranked criterion-referenced test results as the second most impressive data for OE or State officials, followed closely by improvements in educational management and accountability, and improvements

in curriculum and instructional methods. Local title VII project officials, however, considered improvements in curriculum and instructional methods the second most impressive program result to OE, followed by improvements in educational management and accountability. Program results from criterion-referenced tests were ranked fourth.

Program results that impress local project officials are different from what they believe impresses OE or State program officers. Officials from all three local project types ranked improvement in curriculum and instructional methods as the most impressive program result. Local title I, III, and VII project officials consider gains in the affective domain the second most impressive program result, but with title VII, gain scores on norm-referenced tests also received the same ranking.

Concerning these results, (1) local title I project officials appear to consider results from criterion-referenced tests more impressive than norm-referenced test results, (2) local title III officials consider them equally impressive, and (3) local title VII officials clearly prefer norm- to criterion-referenced test results. In all cases, however, test results are not the most impressive program result to local project officials.

Since local project officials perceive results from improvements in curriculum and instructional methods and improvements in the affective domain as more meaningful to them than to State or OE officials, the extent to which such results are excluded from evaluations will probably reduce the adequacy of the evaluations and perhaps make them less useful to local officials. Generally, the degree to which local perceptions of what will most impress State or OE officials causes evaluations to emphasize test results, and educational management and accountability will also probably affect the adequacy and perhaps the usefulness of evaluations at the local level.

As with the State officials' perception of OE, local project officials generally believe OE or State program officers are most impressed by student outcome measures of program effectiveness, such as norm- and criterion-referenced tests. Local project officials themselves are considerably less impressed by these measures. They are more impressed by a variety of measures, but do not believe OE or State program officers fully share this interest.

Correspondingly, 51 percent of the local agency respondents and 77 percent of State agency respondents to our questionnaire stated that the educational community needs to greatly increase efforts to develop alternatives to the classic standardized norm-referenced tests.

Certain differences in questionnaire ratings among Federal, State, and local projects may be because of differing project objectives and priorities. However, local officials clearly do not regard test scores as the sole or most impressive criterion for determining program effectiveness. This may indicate a growing disenchantment with using standardized norm-referenced tests as program evaluation tools. Local officials are apparently interested in knowing how the projects as a whole are functioning and may see test measures as only one of several factors to be used in assessing individual and project performance.

### CONCLUSIONS

Local and State evaluation reports on Federal elementary and secondary education program effectiveness are intended to provide information on which local, State, and Federal officials can base policy and program decisions. However, our questionnaire results show that State and local officials see important differences in the types of evidence of program effectiveness that they themselves and officials at other levels--Federal, State, and local--prefer. Therefore, better communication is needed among the three levels about the information they need to facilitate policy and program decisions at each level. The questionnaire results also raise this question: Should all three levels be served by a reporting system based on the same reports?

Although State officials view OE program officials as being most impressed by standardized norm-referenced test results, and local officials view State and OE officials in the same manner, State and local officials say that they are not most impressed by such results. Local officials prefer broader, more diverse information on program results than just these test scores and they are most impressed by improvements in curriculum and instructional methods and gains in the affective domain (likes, dislikes, interests, attitudes, motives, etc.). State officials are most impressed by results from criterion-referenced tests.

OE's Assistant Commissioner for Planning, Budgeting, and Evaluation believes that hard, objective data on students' cognitive improvements is the most important information

needed. HE noted that this means gain scores on standardized norm-referenced achievement tests because these are most available.

The widespread use of standardized norm-referenced tests to evaluate State and local title I, III, and VII programs indicates that State and local officials have more frequently based their evaluations on the kinds of results they believe would be likely to most impress higher level officials than on their own preferences.

#### RECOMMENDATION TO THE SECRETARY OF HEW

In connection with the assessment recommended in chapter 4, we recommend that the Secretary direct OE to review the types of State and/or local program evaluation information collected (or planned to be collected) on programs authorized by titles I and VII of the Elementary and Secondary Education Act. The review should include an assessment of the information's usefulness at each level and should determine:

- Whether it is realistic to attempt to serve Federal, State, and local levels with aggregated data based on local agency evaluation reports.
- How the information needs at the local, State, and Federal levels can best be met.
- Whether unnecessary duplication exists or will exist in meeting Federal information requirements through the State and local reporting systems as well as through OE national evaluations on these programs, and if so, how it should be eliminated.

During this review process, to define the evaluation information needed at State and local levels, OE should seek the views and cooperation of the State and local officials who are intended to use the results. 1/

---

1/HEW combined its comments on our recommendations for this chapter and chapter 4. For discussion of these comments, see "Agency comments and our evaluation" section on p. 37.

RESULTS OF GAO'SSTATE EDUCATION AGENCY QUESTIONNAIRE (note a)

	Number responding from 50 (note b)	<u>Responses</u>	
		<u>Number</u>	<u>Percent</u> (note c)
<u>Section A: General</u>			
1. How familiar are you with the research being conducted by the Center for the Study of Evaluation at the University of California at Los Angeles to evaluate the utility of many popular commercially available standardized norm-referenced tests? (note d) (Check the one response which best expresses your familiarity with the Center's research.)	50		
___ Have little or no information		6	12.0
___ Aware of the Center's work in test evaluation		14	28.0
___ Read some of the Center's publications on the evaluation of norm-referenced tests		17	34.0
___ Used the Center's material to assist in the selection of commercially available standardized norm-referenced tests		13	<u>26.0</u>
			<u>100.0</u>
2. How familiar are you with the Anchor Test Study conducted by the Educational Testing Service for the U.S. Office of Education (OE) to provide the ability to translate a child's score on any one of the eight most widely used standardized reading tests into a score on any of the other tests.	50		
___ Have little or no information		3	6.0
___ Aware of the Anchor Test Study		13	26.0
___ Read the Anchor Test Study		25	50.0
___ Used the Anchor Test Study		9	<u>18.0</u>
			<u>100.0</u>

3. From your experience, how great is the need, if any, to increase the effort by the educational community devoted to measurement and assessment techniques in any of the following areas? (Indicate your answer by checking the appropriate column for each of the seven line items listed below.) (notes b and e)

	Little or no increase		Some increase		Moderate increase		Substantial increase		Very great increase		
	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	
(1) Development of methods for test design and construction.	48	3	6.3	9	18.8	11	22.9	17	35.4	8	16.7
(2) Reduction of cultural, racial, and sexist biases in tests	49	3	6.1	10	20.4	11	22.4	15	30.6	10	20.4
(3) Development of alternatives to the classic standardized norm-referenced tests (e.g., criterion-referenced tests) (note f)	48	2	4.2	3	6.3	6	12.5	16	33.3	21	43.8
(4) Development of more and improved standardized norm-referenced tests	50	17	34.0	20	40.0	4	8.0	7	14.0	2	4.0
(5) Development and utilization of methods to better evaluate standardized norm-referenced tests in use	50	9	18.0	14	28.0	7	14.0	9	18.0	11	22.0
(6) Development and implementation of methods for the dissemination and utilization of assessment techniques for educational practitioners	46	1	2.2	0	0	4	8.7	14	30.4	27	58.7
(7) Other (please specify)	10	1	10.0	0	0	1	10.0	1	10.0	7	70.0

Section B: State title I programs

4. The demonstration of program results should be an important factor in the continuance of program funding from the Office of Education. Consider each of the following types of findings used to assess programs. FROM YOUR EXPERIENCE, which types of results do you think are more likely to impress OE program officers? Indicate your answer by rank ordering each of the following alternatives from the most to least impressive. Select the type of results which you think is most likely to impress an OE official. Rank this list. Do the same for all the remaining categories, ranking them 2nd, 3rd, 4th, 5th, 6th, and 7th. (notes b and e)

	Number responding from 50	1st		2nd		3rd		4th		5th		6th		7th	
		Num-ber	Per-cent												
(1) Improvement in educational management or accountability	49	10	20.4	14	28.6	6	16.3	6	12.2	1	2.0	7	14.3	3	6.1
(2) Improvement in school services or facilities	49	0	0	2	4.1	2	4.1	4	8.2	5	10.2	9	18.4	27	55.1
(3) Student improvement through gain scores on grades or teacher ratings	49	2	4.1	4	8.2	11	22.4	13	26.5	10	20.4	5	10.2	4	8.2
(4) Student improvement through gain scores on standardized norm-referenced tests	49	40	81.6	5	10.2	3	6.1	0	0	1	2.0	0	0	0	0
(5) Student improvement through gain scores on criterion-referenced tests	49	2	4.1	18	36.7	13	26.5	6	12.2	7	14.3	2	4.1	1	2.0
(6) Student improvement through gains in the affective domain (e.g., likes, dislikes, interests, attitudes, motives, etc.)	49	0	0	3	6.1	2	4.1	9	18.4	10	20.4	15	30.6	10	20.4
(7) Improvements in curriculum and instructional methods	49	1	2.0	7	14.3	4	8.2	12	24.5	12	24.5	8	16.3	5	10.2

5. Regardless of the types of results which are likely to impress OE program officials, what types of results are likely to impress you, an educator? Indicate your answer by rank ordering each of the following alternatives with the same procedure used in the previous question (notes b and e)

	Number re- spond- ing from 50	1st		2nd		3rd		4th		5th		6th		7th	
		Num- ber	Per- cent												
(1) Improvement in educational management or accountability	48	8	16.7	4	8.3	6	12.5	7	14.6	12	25.0	7	14.6	4	8.3
(2) Improvement in school services or facilities	49	1	2.0	1	2.0	2	4.1	4	8.2	5	10.2	12	24.5	24	49.0
(3) Student improvement through gain scores on grades or teacher ratings	48	5	10.4	2	4.2	10	20.8	9	18.8	8	16.7	9	18.8	5	10.4
(4) Student improvement through gain scores on standardized norm-referenced tests	48	20	41.7	10	20.8	9	18.8	3	6.3	0	0	2	4.2	4	8.3
(5) Student improvement through gain scores on criterion-referenced tests	49	11	22.4	23	46.9	7	14.3	3	6.1	1	2.0	4	8.2	0	0
(6) Student improvement through gains in the affective domain (i.e. likes, dislikes, interests, attitudes, motives, etc.)	49	6	12.2	7	14.3	8	16.3	12	24.5	10	20.4	2	4.1	4	8.2
(7) Improvements in curriculum and instructional methods	49	8	16.3	5	10.2	8	16.3	10	20.4	8	16.3	8	16.3	2	4.1

## APPENDIX I

## APPENDIX I

	Number responding from 50 (note b)	Responses	
		Number	Percent (note c)
6. Which of the following techniques did your State employ for its 1973-74 evaluation of title I? (Check all that apply.)	48		
___ Aggregation and analysis of data from local education agency reports		45	93.8
___ Educational audits and their results		9	18.8
___ Statewide testing of title I students		11	22.9
___ Other (please specify)		5	10.4
Note: If your State did not test title I students statewide, skip questions 7 and 8.			
7. Which of the following types of tests did your State administer for its title I evaluation?	26		
___ Standardized norm-referenced tests (note d)		21	80.8
___ Criterion-referenced tests (note f)		3	11.5
___ Other tests (please specify)		4	15.4
If you do not use standardized norm-referenced tests, skip question 8. If you do, continue.			
8. How did you report the results for the standardized norm-referenced testing?	28		
___ Raw scores		2	7.1
___ Grade equivalents		24	85.7
___ Percentiles		6	21.4
___ Quartiles		1	3.6
___ Stanines		4	14.3
___ Other (please specify)		2	7.1

APPENDIX I

APPENDIX I

**NOTE:** In the following five questions you will be asked to rate several aspects of the Elementary and Secondary Education Act, title I, local and State evaluation reports that affect the degree to which these documents satisfy your policy, management and programmatic information needs. You are asked to provide overall judgments on the adequacy of the quality, informational content, and utility of the evaluation reports. Do this by considering each of these attributes: very deficient, deficient, marginal, adequate, and more than adequate. Check the box which most appropriately reflects how you feel about the respective local and State evaluations with regard to the particular aspects in question.

9. Rate the FOCUS AND SCOPE of the local and State evaluation reports (notes b and e)

(Focus and Scope: the adequacy with which the report covers the essential and related material and the appropriateness of the emphasis and treatment given to the relevant topics, details, and high and lower priority information). (Check one box in each row.)

	Number re- spond- ing from 50	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent
Local reports	48	0	0	5	10.4	19	39.6	23	47.9	1	2.1
State reports	48	0	0	4	8.3	16	33.3	25	52.1	3	6.3

10. Rate the local and State evaluation reports on THE PRESENTATION OF REQUIRED MANAGEMENT INFORMATION NEEDS (notes b and e)

(Presentation of Required Management Information Needs: the extent to which the report presents the information needed to evaluate and update current policies by those who transfer policy decisions into plans, budgets, program implementation, operational oversight, resource allocations, forecasts, status assessments and reports, educational accountability, costs, benefits and efficiency assessments). (Check one box in each row.)

	Number re- spond- ing from 50	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent
Local reports	49	2	4.1	7	14.3	26	53.1	14	28.6	0	0
State reports	49	1	2.0	6	12.2	22	44.9	18	36.7	2	4.1

11. Rate the local and State evaluation reports on the adequacy with which they properly QUALIFY FINDINGS (notes b and e)

(Qualification of Findings: the extent to which the report properly qualifies the findings and assumptions and identifies those conditions and situations where the findings are not applicable.) (Check one box in each row.)

	Number re- spond- ing from 50	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent
Local reports	49	3	6.1	11	22.4	20	40.8	14	28.6	1	2.0
State reports	49	2	4.1	8	16.3	13	26.5	24	49.0	2	4.1

12. Rate the local and State evaluation reports on the CREDIBILITY OF FINDINGS (notes b and e)

(Credibility of Findings: the degree of confidence expressed in the findings through statements of statistical certainty, soundness of method, evidence of replication, consensual agreements, similar experiences, supporting expert judgment and opinions, and reasonableness of assumptions). (Check one box in each row.)

	Number re- spond- ing from 50	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent
Local reports	49	4	8.2	12	24.5	17	34.7	15	30.6	1	2.0
State reports	49	2	4.1	10	20.4	17	34.7	18	36.7	2	4.1

13. Rate the local and State evaluation reports on the adequacy of the QUALIFICATION AND QUANTIFICATION OF MEASUREMENT DATA (notes b and e)

(Qualification and Quantification of Measurement Data: the extent to which the evaluation assessments can be qualified and quantified into measurable attributes and parameters that address the problem in measurable, operational, or concrete terms). (Check one box in each row.)

Number reporting from 50	Very deficient		Deficient		Marginal		Adequate		More than adequate		
	Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-	
	ber	cent	ber	cent	ber	cent	ber	cent	ber	cent	
Local reports	49	2	4.1	7	14.3	19	38.8	19	38.8	2	4.1
State reports	49	2	4.1	4	8.2	22	44.5	17	34.7	4	8.2

14. How often do your Elementary and Secondary Education Act, title I, local and State program evaluations adequately report information on the needs assessment, number of children, per pupil expenditure, project achievement, and pupil benefit parameters? (Check one box in each row for both local and State evaluations.) (notes b and e)

Local reports	Number reporting from 50		Very often		As often as not		Generally		Occasionally		Seldom	
	Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-
	ber	cent	ber	cent	ber	cent	ber	cent	ber	cent	ber	cent
(1) Information on the manner in which the needs of children are assessed	46	15	32.6	18	39.1	1	2.2	9	19.6	3	6.5	
(2) Information on the number of children in the program	46	30	65.2	12	26.1	3	6.5	1	2.2	0	0	
(3) Per pupil expenditures of each program	46	15	32.6	16	34.8	5	10.9	4	8.7	6	13.0	
(4) Evidence of qualifiable or measurable achievements	46	13	28.3	23	50.0	6	13.0	4	8.7	0	0	
(5) Evidence of qualifiable or measurable pupil benefits	45	9	19.6	19	41.3	11	23.9	7	15.2	0	0	

State reports	Number responding from 50	Very often		Generally		As often as not		Occasionally		Seldom	
		Number	Per cent	Number	Per cent	Number	Per cent	Number	Per cent	Number	Per cent
Information on the manner in which the needs of children are assessed	44	11	25.0	19	43.2	3	6.8	7	15.9	4	9.1
(2) Information on the number of children in the program	45	29	64.4	11	24.4	3	6.7	2	4.4	0	0
(3) Per pupil expenditures of each program	44	17	38.6	18	40.9	4	9.1	1	2.3	4	9.1
(4) Evidence of qualifiable or measurable achievements	45	16	35.6	18	40.0	8	17.8	2	4.4	1	2.2
(5) Evidence of qualifiable or measurable pupil benefits	45	11	24.4	18	40.0	8	17.8	8	17.8	0	0

Section C: State title III programs

15. The demonstration of program results should be an important factor in the continuance of program funding from the Office of Education. Consider each of the following types of findings used to assess programs. FROM YOUR EXPERIENCE, which types of results do you think are more likely to impress OE program officers? Indicate your answer by rank ordering each of the following alternatives from the most to least impressive. Select the type of results which you think is most likely to impress an OE official. Rank this list. Do the same for all the remaining categories, ranking them 2nd, 3rd, 4th, 5th, 6th and 7th. (notes b and e)

	Number responding from 50	1st		2nd		3rd		4th		5th		6th		7th	
		Num-ber	Per-cent												
(1) Improvement in educational management or accountability	49	8	16.3	5	10.2	6	12.2	5	10.2	15	30.6	8	16.3	2	4.1
(2) Improvement in school services or facilities	48	1	2.1	2	4.2	1	2.1	4	8.3	2	4.2	8	16.7	30	62.5
(3) Student improvement through gain scores on grades or teacher ratings	47	0	0	1	2.1	12	25.5	14	29.0	6	12.8	4	8.5	10	21.3
(4) Student improvement through gain scores on standardized norm-referenced tests	48	31	64.6	10	20.8	5	10.4	1	2.1	1	2.1	0	0	0	0
(5) Student improvement through gain scores on criterion-referenced tests	48	8	16.7	19	39.6	9	18.8	7	14.6	2	4.2	3	6.3	0	0
(6) Student improvement through gains in the affective domain (e.g., likes, dislikes, interests, attitudes, motives, etc.)	49	0	0	5	10.2	10	20.4	11	22.4	11	22.4	8	16.3	4	8.2
(7) Improvements in curriculum and instructional methods	46	2	4.3	6	13.0	7	15.2	4	8.7	10	21.7	12	25.1	5	10.9

16. Regardless of the types of results which are likely to impress OE program officials, what types of results are likely to impress you, an educator? Indicate your answer by rank ordering each of the following alternatives with the same procedure used in the previous question. (notes b and e)

	Number responding from 50	1st		2nd		3rd		4th		5th		6th		7th	
		Num-ber	Per-cent												
(1) Improvement in educational management or accountability	48	3	6.3	2	4.2	9	18.8	9	18.8	9	18.8	12	25.0	4	8.3
(2) Improvement in school services or facilities	47	0	0	2	4.3	1	2.1	2	4.3	6	12.8	10	21.3	26	55.3
(3) Student improvement through gain scores on grades or teacher ratings	48	0	0	5	10.4	3	6.3	11	22.9	6	12.5	9	18.8	14	29.2
(4) Student improvement through gain scores on standardized norm-referenced tests	47	12	25.5	7	14.9	8	17.0	6	12.8	6	12.8	5	10.6	3	6.4
(5) Student improvement through gain scores on criterion-referenced tests	48	20	41.7	11	22.9	5	10.4	6	12.5	3	6.3	3	6.3	0	0
(6) Student improvement through gains in the affective domain (e.g., likes, dislikes, interests, attitudes, motives, etc.)	48	3	6.3	15	31.3	16	33.3	6	12.5	6	12.5	2	4.2	0	0
(7) Improvements in curriculum and instructional methods	48	0	0	9	18.8	6	12.5	7	14.6	10	20.8	7	14.6	0	0

## APPENDIX I

## APPENDIX I

	Number responding from 50 (note b)	Responses	
		Number	Percent (note c)
17. Which of the following techniques did your State employ for its 1973-74 evaluation of title III? (Check all that apply.)	48		
___ Aggregation and analysis of data from local education agency reports		41	85.4
___ Educational audits and their results		27	56.3
___ Statewide testing of title III students		5	10.4
___ Other (please specify)		12	25.0
Note: If your State did not test title III students statewide, skip questions 18 and 19.			
18. Which of the following types of tests did your State administer for its title III evaluation?	14		
___ Standardized norm-referenced tests (note d)		9	64.3
___ Criterion-referenced tests (note f)		8	57.1
___ Other tests (please specify)		7	50.0
If you do not use standardized norm-referenced tests, skip question 19. If you do, continue.			
19. How did you report results for the standardized norm-referenced tests?	10		
___ Raw scores		2	20.0
___ Grade equivalents		7	70.0
___ Percentiles		4	40.0
___ Quartiles		2	20.0
___ Stanines		5	50.0
___ Other (please specify)		2	20.0

APPENDIX I

APPENDIX I

**NOTE:** In the following five questions you will be asked to rate several aspects of the Elementary and Secondary Education Act, title III, local and State evaluation reports that affect the degree to which these documents satisfy your policy, management and programmatic information needs. You are asked to provide overall judgments on the adequacy of the quality, informational content, and utility of the evaluation reports. Do this by considering each of these attributes: very deficient, deficient, marginal, adequate, and more than adequate. Check the box which most appropriately reflects how you feel about the respective local and State evaluations with regard to the particular aspects in question.

20. Rate the FOCUS AND SCOPE of the local and State evaluation reports (notes b and e)

(Focus and Scope: the adequacy with which the report covers the essential and related material and the appropriateness of the emphasis and treatment given to the relevant topics, details, and high and lower priority information). (Check one box in each row.)

	Number re- spond- ing from 50	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-
		ber	cent	ber	cent	ber	cent	ber	cent	ber	cent
Local reports	48	1	2.1	2	4.2	16	33.3	27	56.3	2	4.2
State reports	46	1	2.2	5	10.9	14	30.4	25	54.3	1	2.2

21. Rate the local and State evaluation reports on THE PRESENTATION OF REQUIRED MANAGEMENT INFORMATION NEEDS (notes b and e)

(Presentation of Required Management Information Needs: the extent to which the report presents the information needed to evaluate and update current policies by those who transfer policy decisions into plans, budgets, program implementation, operational oversight, resource allocations, forecasts, status assessments and reports, educational accountability, costs, benefits, and efficiency assessments). (Check one box in each row.)

	Number re- spond- ing from 50	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-
		ber	cent	ber	cent	ber	cent	ber	cent	ber	cent
Local reports	48	2	4.2	4	8.3	21	43.8	19	39.6	2	4.2
State reports	45	3	6.7	1	2.2	14	31.1	26	57.8	1	2.2

22. Rate the local and State evaluation reports on the adequacy with which they properly QUALIFY FINDINGS (notes b and e)

(Qualification of Findings: the extent to which the report properly qualifies the findings and assumptions and identifies those conditions and situations where the findings are not applicable). (Check one box in each row.)

	Number re- spond- ing from <u>50</u>	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent
Local reports	48	1	2.1	7	14.6	21	43.8	17	35.4	2	4.2
State reports	47	3	6.4	3	6.4	16	34.0	23	48.9	2	4.3

23. Rate the local and State evaluation reports on the CREDIBILITY OF FINDINGS (notes b and e)

(Credibility of Findings: the degree of confidence expressed in the findings through statements of statistical certainty, soundness of method, evidence of replication, consensual agreements, similar experiences, supporting expert judgment and opinions, and reasonableness of assumptions). (Check one box in each row.)

	Number re- spond- ing from <u>50</u>	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent
Local reports	48	1	2.1	4	8.3	24	50.0	17	35.4	2	4.2
State reports	47	1	2.1	3	6.4	15	31.9	24	51.1	4	8.5

APPENDIX I

24. Rate the local and State evaluation reports on the adequacy of the QUALIFICATION AND QUANTIFICATION OF MEASUREMENT DATA (notes b and e)

(Qualification and Quantification of Measurement Data: the extent to which the evaluation assessments can be qualified and quantified into measurable attributes and parameters that address the problem in measurable, operational, or concrete terms). (Check one box in each row.)

	Number responding from 50	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent
Local reports	48	0	0	7	14.6	21	43.8	19	39.6	1	2.1
State reports	47	1	2.1	3	6.4	22	46.8	19	40.4	2	4.3

25. How often do your Elementary and Secondary Education Act, title III, local and State program evaluations adequately report information on the needs assessment, number of children, per pupil expenditure, project achievement, and pupil benefit parameters? (Check one box in each row for both local and State evaluations.) (notes b and e)

	Number responding from 50	Very often		As often as not		Occasionally		Seldom	
		Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent
(1) Information on the manner in which the needs of children are assessed	47	9	19.1	22	46.8	8	17.0	6	12.8
(2) Information on the number of children in the program	47	33	70.2	9	19.1	4	8.5	0	0
(3) Per pupil expenditures of each program	47	19	40.4	14	29.8	4	8.5	5	10.6
(4) Evidence of qualifiable or measurable achievements	47	14	29.8	19	33.3	11	23.4	3	6.4
(5) Evidence of qualifiable or measurable pupil benefits	47	10	21.3	19	40.4	11	23.4	6	12.8

APPENDIX I

APPENDIX I

APPENDIX I

State reports	Number re- spond- ing from 50	Very often		Generally		As often as not		Occa- sionally		Seldom	
		Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent
(1) Information on the manner in which the needs of children are assessed	40	4	10.0	18	45.0	7	17.5	6	15.0	5	12.5
(2) Information on the number of children in the program	40	21	52.5	13	32.5	4	10.0	1	2.5	1	2.5
(3) Per pupil expenditures of each program	40	12	30.0	13	32.5	5	12.5	8	20.0	2	5.0
(4) Evidence of qualifiable or measurable achievements	39	8	20.5	15	38.5	11	28.2	2	5.1	3	7.7
(5) Evidence of qualifiable or measurable pupil benefits	40	6	15.0	14	35.0	10	25.0	5	12.5	5	12.5

a/We requested that the questionnaire be completed by State officials familiar with the State's evaluation efforts conducted for Elementary and Secondary Education Act, titles I (regular, migrant, neglected, and delinquent, etc.) and III, and the State's own assessment efforts. The questionnaire was divided into three sections: section A, National Assessment of Educational Progress and statewide assessment; section B, Elementary and Secondary Education Act, title I evaluation efforts; and section C, title III evaluation efforts. We suggested that the questionnaire be taken apart and distributed to persons most familiar with each part, such as the State director of research, planning, and evaluation for section A, the State director of title I for section B, and the State director of title III for section C. Some of the questions from the questionnaire have been omitted from this report. Questions pertaining to the National Assessment of Educational Progress and statewide assessments are shown in GAO's report to the Congress (HRD-76-113), dated July 20, 1976.

b In April 1975 we sent the questionnaire to the education agencies in all States and the District of Columbia. By June 1975 the District of Columbia and all but one State responded. For purposes of compiling responses to the questionnaire, the District of Columbia is considered to be a State.

c This column shows the percentage of respondents to the question that chose each specific answer.

d Standardized norm-referenced tests are tests which purport to assess the individual student's ability or achievement in broad subject areas as compared to the rest of the host population (e.g., the Metropolitan Achievement Test (MAT) or the Wechsler Intelligence Scale for Children (WISC)).

e The percent columns show the percentage of respondents to each line item that chose each category. Where the percentages on each line do not add to 100, it is due to rounding.

f Criterion-referenced tests are tests specifically constructed to measure students' attainment of specific educational objectives or proficiency with specified curriculum material. These tests, which may be standardized, usually provide a specific and operational description of the level and type of task performance or behavioral measures used as a criterion to indicate attainment of the educational objectives. For example, the student must be able to compute the correct product of all single digit numerals greater than zero with no more than five errors.

RESULTS OF GAO'S

LOCAL EDUCATION AGENCY QUESTIONNAIRE (note a)

	Number of projected responses from 8,936 (note b)	<u>Responses</u>	
		<u>Number</u> (note b)	<u>Percent</u> (note c)
<u>Section A: To be completed by questionnaire respondents from all local education agencies sampled</u>			
1. How familiar are you with the research being conducted at the Center for the Study of Evaluation at the University of California at Los Angeles to evaluate the utility of many popular commercially available standardized norm-referenced tests? (note d) (Check the one response which best expresses your familiarity with the Center's research.)	5,987		
___ Have little or no information		5,059	84.5
___ Aware of the Center's work in test evaluation		601	10.0
___ Read some of the Center's publications on the evaluation of norm-referenced tests		192	3.2
___ Used the Center's material to assist in the selection of commercially available standardized norm-referenced tests		135	<u>2.3</u>
			<u>100.0</u>
2. How familiar are you with the Anchor Test Study conducted by the Educational Testing Service for the U.S. Office of Education (OE) to provide the ability to translate a child's score on any one of the eight most widely used standardized reading tests into a score on any of the other tests?	282		
___ Have little or no information		282	100.0
___ Aware of the Anchor Test Study		-	-
___ Read the Anchor Test Study		-	-
___ Used the Anchor Test Study		-	-
			<u>100.0</u>

3. From your experience, how great is the need, if any, to increase the effort by the educational community devoted to measurement and assessment techniques in any of the following areas? (Indicate your answer by checking the appropriate column for each of the seven line items listed below.) (notes b and e)

	Projected number responding from 6,936	Little or no increase		Some increase		Moderate increase		Substantial increase		Very great increase	
		Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent
(1) Development of methods for test design and construction	7,410	1,801	24.3	1,645	22.2	1,457	19.7	1,495	20.2	1,012	13.7
(2) Reduction of cultural, racial, and sexist biases in tests	7,398	2,241	30.3	1,321	17.9	1,256	17.0	1,465	19.8	1,116	15.1
(3) Development of alternatives to the classic standardized norm-referenced tests (e.g., criterion-referenced tests) (note g)	8,028	1,477	18.4	1,296	16.1	1,195	14.9	2,139	26.6	1,922	23.9
(4) Development of more and improved standardized norm-referenced tests	8,019	2,097	26.2	1,600	19.9	2,145	26.7	1,703	16.3	874	10.9
(5) Development and utilization of methods to better evaluate standardized norm-referenced tests in use	8,001	1,149	14.4	1,590	19.9	1,853	23.2	1,959	24.5	1,450	18.1
(6) Development of methods for the dissemination and utilization of assessment techniques	3,026	886	11.0	1,180	14.7	1,332	16.6	2,243	27.9	2,385	29.7
(7) Other (please specify)	328	44	13.4	0	0	0	0	76	23.2	208	63.4

APPENDIX II

Section B: Complete this section only if you have a funded title I program.

4. The demonstration of program results should be an important factor in the continuance of program funding from the Office of Education. Consider each of the following types of findings used to assess programs. FROM YOUR EXPERIENCE, which types of results do you think are more likely to impress OE or State education agency program officers. Indicate your answer by rank ordering each of the following alternatives from the most to least impressive. Select the type of results which you think is most likely to impress OE or State education agency officials. Rank this list. Do the same for all the remaining categories, ranking them 2nd, 3rd, 4th, 5th, 6th, and 7th. (Notes b and e)

Projected number responding from SJSU	1st		2nd		3rd		4th		5th		6th		7th	
	Num	Per cent												
8,481	1,487	17.5	1,089	12.8	1,231	14.5	1,083	12.8	1,492	17.6	983	11.6	1,116	13.2
8,481	290	3.4	481	5.7	781	9.2	1,135	13.4	1,426	16.8	1,972	23.2	2,397	28.3
8,543	479	5.6	1,057	12.4	1,591	18.6	1,802	21.1	970	11.3	1,322	15.5	1,324	15.5
8,564	4,595	53.6	1,437	16.8	844	9.9	653	7.6	473	5.5	298	3.5	266	3.1
8,477	1,301	15.3	2,758	32.5	1,675	19.8	671	7.9	1,021	12.0	657	7.8	394	4.6
8,521	382	4.5	701	8.2	1,642	19.3	1,438	16.9	1,054	12.4	1,363	16.0	1,941	22.8
8,529	784	9.2	1,029	12.1	1,003	11.8	1,300	15.2	1,745	20.5	1,408	16.5	761	8.9

APPENDIX II

5. Regardless of the types of results which are likely to impress OE or State education agency program officials, what types of results are likely to impress you, an educator? Indicate your answer by ranking each of the following alternatives with the same procedure used in the previous question. (notes b and e)

	Projected number responding from 8,936	1st		2nd		3rd		4th		5th		6th		7th	
		Num-ber	Per-cent												
(1) Improvement in educational management or accountability	8,489	685	8.1	1,014	11.9	853	10.0	1,120	13.2	1,132	13.3	1,420	16.7	2,266	26.7
(2) Improvement in school services or facilities	8,508	604	7.1	471	5.5	1,126	13.2	897	10.5	1,481	17.4	2,036	23.9	1,892	22.2
(3) Student improvement through gain scores on grades or teacher ratings	8,491	880	10.4	1,215	14.3	1,431	16.9	1,601	18.9	1,414	16.6	1,165	13.7	786	9.3
(4) Student improvement through gain scores on standardized norm-referenced tests	8,526	1,661	19.5	1,557	18.3	1,009	11.8	1,111	13.0	1,192	14.0	889	10.4	1,108	13.0
(5) Student improvement through gain scores on criterion-referenced tests	8,505	2,073	24.4	1,438	16.9	1,202	14.1	1,327	15.6	746	8.8	946	11.1	774	9.1
(6) Student improvement through gains in the affective domain (i.e., likes, dislikes, interests, attitudes, motives, etc.)	8,515	1,725	20.3	1,833	21.5	1,381	16.2	1,428	16.8	704	8.3	652	7.7	792	9.3
(7) Improvements in curriculum and instructional methods	8,492	1,745	20.6	1,934	22.8	1,487	17.5	913	10.7	1,381	16.3	792	9.3	240	2.8

APPENDIX II

APPENDIX II

	Number of projected responses from 8,936 (note b)	Responses	
		Number (note b)	Percent (note c)
6. Are your local evaluations of federally funded programs usually performed by external or internal evaluators? (Check one.)	8,205		
___ Internal (e.g., local education agency staff)		5,468	66.6
___ External (e.g., consultants)		494	6.0
___ Both		<u>2,243</u>	<u>27.3</u>
			<u>100.0</u> f/
7. Which of the following types of tests did your local education agency administer for its title I evaluation? If your local education agency did not test title I students, skip questions 7 and 8.	8,583		
___ Standardized norm-referenced tests (note d)		8,103	94.4
___ Criterion-referenced tests (note g)		2,110	24.6
___ Other tests (please specify)		1,168	13.6
If you do not employ standardized norm-referenced tests, skip question 8. If you do, continue.			
8. How did you report the results for the standardized norm-referenced testing?	8,029		
___ Raw scores		1,792	22.3
___ Grade equivalents		6,658	82.9
___ Percentiles		2,978	37.1
___ Quartiles		528	6.6
___ Stanines		946	11.8
___ Other (please specify)		163	2.0

APPENDIX II

APPENDIX II

**NOTE:** In the following five questions you will be asked to rate several aspects of the Elementary and Secondary Education Act, title I, local and State evaluation reports that affect the degree to which these documents satisfy your policy, management, and programmatic information needs. You are asked to provide overall judgments on the adequacy of the quality, informational content, and utility of the evaluation reports. Do this by considering each of these attributes: very deficient, deficient, marginal, adequate, and more than adequate. Check the box which most appropriately reflects how you feel about the respective local and State evaluations with regard to the particular aspects in question.

9. Rate the FOCUS AND SCOPE of the local and State evaluation report. (notes b and e)

(Focus and Scope: the adequacy with which the report covers the essential and related material and the appropriateness of the emphasis and treatment given to the relevant topics, details, and high and lower priority information). (Check one box in each row)

	Projected number re- spond- ing from 8,936	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent
Local reports	8,510	45	0.5	426	5.0	2,097	24.6	4,521	53.1	1,422	16.7
State reports	8,350	196	2.3	733	8.8	2,248	26.9	4,098	49.1	1,076	12.9

10. Rate the local and State evaluation reports on the PRESENTATION OF REQUIRED MANAGEMENT INFORMATION NEEDS. (notes b and e)

(Presentation of Required Management Information Needs: the extent to which the report presents the information needed to evaluate and update policies by those who transfer policy decisions into plans, budgets, program implementation, curriculum, operational oversight, resource allocations, forecasts, status assessments and reports, educational accountability, costs, benefits, and efficiency assessments). (Check one box in each row.)

	Projected number re- spond- ing from 8,936	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent
Local reports	8,635	119	1.4	484	5.6	2,782	32.2	4,037	46.7	1,212	14.0
State reports	8,486	312	3.7	863	10.2	2,879	33.9	3,313	39.0	1,119	13.2

11. Rate the local and State evaluation reports on the adequacy with which they properly QUALIFY FINDINGS. (notes b and e)

(Qualification of Findings: the extent to which the report properly qualifies the findings and assumptions and identifies those conditions and situations where the findings are not applicable). (Check one box in each row.)

	Projected number responding from 8,936	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-
		ber	cent	ber	cent	ber	cent	ber	cent	ber	cent
Local reports	8,518	217	2.5	518	6.1	2,759	32.4	4,173	49.0	852	10.0
State reports	8,347	451	5.4	709	8.5	2,978	35.7	3,503	42.0	706	8.5

12. Rate the local and State evaluation reports on the CREDIBILITY OF FINDINGS (notes b and e)

(Credibility of Findings: the degree of confidence expressed in the findings through statements of statistical certainty, soundness of method, evidence of replication, consensual agreements, similar experiences, supporting expert judgment and opinions, and reasonableness of assumptions). (Check one box in each row.)

	Projected number responding from 8,936	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-
		ber	cent	ber	cent	ber	cent	ber	cent	ber	cent
Local reports	8,552	277	3.2	546	6.4	2,403	28.1	4,472	52.3	854	10.0
State reports	8,391	616	7.3	779	9.3	2,356	28.1	3,957	47.2	684	8.1

13. Rate the local and State evaluation reports on the adequacy of the QUALIFICATION AND QUANTIFICATION OF MEASUREMENT DATA. (Notes b and e)

(Qualification and Quantification of Measurement Data: the extent to which the evaluation assessments can be qualified and quantified into measurable attributes and parameters that address the problem in measurable, operational, or concrete terms). (Check one box in each row.)

	Projected number	responding from 8,936	Very deficient		Marginal		Adequate		More than adequate		
			Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-	
			ber	cent	ber	cent	ber	cent	ber	cent	
Local reports	8,522	167	2.0	1.071	12.6	2,524	29.6	4,036	47.4	724	8.5
State reports	8,369	298	3.6	932	11.1	2,844	34.0	3,580	42.8	716	8.6

14. How often do your Elementary and Secondary Education Act, title I, local program evaluations adequately report information on the need assessment, number of children, per pupil expenditure, project achievement, and pupil benefit parameters? (Check one box in each row.) (notes b and e)

	Projected number	responding from 8,936	Very often		Generally		As often as not		Occasionally		Seldom	
			Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-
			ber	cent	ber	cent	ber	cent	ber	cent	ber	cent
(1) Information on the manner in which the needs of children are assessed	8,435	3,034	36.0	3,966	47.3	469	5.6	792	9.4	155	1.8	
(2) Information on the number of children in the program	8,422	4,686	55.6	2,692	32.0	415	4.9	474	5.6	156	1.8	
(3) Per pupil expenditures of each program	8,242	2,653	32.2	3,144	38.1	801	9.7	1,094	13.3	550	6.7	
(4) Evidence of qualifiable and measurable achievements	8,435	7,887	93.4	3,973	47.1	999	11.8	498	5.9	79	0.9	
(5) Evidence of qualifiable or measurable pupil benefits	8,331	2,398	28.8	4,324	51.9	829	10.0	633	7.6	146	1.8	

15. From your experience, how often can you draw comparisons between the results of your local Federal program and Federal programs in other localities from State and Federal evaluation reports? (Check one box for each type of report.) (notes b and e)

Projected number re- spond- ing from 8,936	Very often		Generally as not		Occa- sionally		Seldom		No basis to judge			
	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent		
State reports 8,448	629	7.4	1,799	21.3	896	10.6	1,391	16.5	1,852	21.9	1,882	22.3
Federal reports 8,435	258	3.1	1,124	13.3	771	9.1	1,493	17.7	2,316	27.5	2,475	29.3

Section C: Complete this section only if you have a funded title III program.

16. The demonstration of program results should be an important factor in the continuance of program funding from the Office of Education. Consider each of the following types of findings used to assess programs. FROM YOUR EXPERIENCE, which types of results do you think are more likely to impress OE or State education agency program officers? Indicate your answer by rank ordering each of the following alternatives from the most to least impressive. Select the type of results which you think is most likely to impress OE or State education agency officials. Rank this 1st. Do the same for all the remaining categories, ranking them 2nd, 3rd, 4th, 5th, 6th and 7th. (notes b and e)

	Projected number responding from 8,936	1st		2nd		3rd		4th		5th		6th		7th	
		Num-ber	Per-cent												
(1) Improvement in educational management or accountability	2,605	497	19.1	279	10.7	307	11.8	420	16.1	508	19.5	349	13.4	244	9.4
(2) Improvement in school services or facilities	2,554	127	5.0	186	7.3	288	11.3	260	10.2	370	14.5	510	20.0	813	31.8
(3) Student improvement through gain scores on grades or teacher ratings	2,605	237	9.1	298	11.4	433	16.6	352	13.5	499	19.2	387	14.8	401	15.4
(4) Student improvement through gain scores on standardized norm-referenced tests	2,612	1,168	44.7	537	20.5	149	5.7	285	10.9	105	4.0	150	5.8	219	8.4
(5) Student improvement through gain scores on criterion-referenced tests	2,611	148	5.7	884	33.9	627	24.0	172	6.6	136	5.2	435	16.7	208	8.0
(6) Student improvement through gains in the affective domain (e.g., likes, dislikes, interests, attitudes, motives, etc.)	2,612	209	8.0	261	10.0	490	18.7	569	21.8	393	15.0	200	7.6	492	18.8
(7) Improvements in curriculum and instructional methods	2,606	447	17.1	252	9.7	269	10.3	487	18.7	525	20.1	501	19.2	126	4.8

17. Regardless of the types of results which are likely to impress OE or State education agency program officials, what types of results are likely to impress you, an educator? Indicate your answer by rank ordering each of the following alternatives with the same procedure used in the previous question. (notes b and e)

	Projected number responding from 8,936	1st		2nd		3rd		4th		5th		6th		7th	
		Num-	Per-												
		ber	cent												
(1) Improvement in educational management or accountability	2,609	337	12.9	253	9.7	346	13.2	299	11.5	263	10.1	455	17.4	657	25.2
(2) Improvement in school services or facilities	2,602	183	7.0	149	5.7	421	16.2	246	9.4	286	11.0	735	28.2	583	22.4
(3) Student improvement through gain scores on grades or teacher ratings	2,587	283	11.0	224	8.7	293	11.3	437	16.9	638	24.7	479	18.5	234	9.0
(4) Student improvement through gain scores on standardized norm-referenced tests	2,597	697	26.8	255	9.8	227	8.7	451	17.4	216	8.3	425	16.4	327	12.6
(5) Student improvement through gain scores on criterion-referenced tests	2,579	309	12.0	663	25.7	438	17.0	262	10.2	354	13.7	170	6.6	382	14.8
(6) Student improvement through gains in the affective domain (i.e., likes, dislikes, interests, attitudes, motives, etc.)	2,568	469	18.3	509	19.8	551	21.4	448	17.4	195	7.6	163	6.3	234	9.1
(7) Improvements in curriculum and instructional methods	2,594	711	27.4	576	22.2	298	11.5	390	15.1	364	14.0	195	7.5	59	2.3

APPENDIX II

APPENDIX II

	Number of projected responses from 8,936 (note b)	Responses	
		Number (note b)	Percent (note c)
18. Are your local evaluations of federally funded programs usually performed by external or internal evaluators? (Check one.)	2,623		
___ Internal (e.g., local education agency staff)		1,445	55.1
___ External (e.g., consultants)		392	14.9
___ Both		786	<u>30.0</u>
			<u>100.0</u>
19. Which of the following types of tests did your local education agency administer for its title III evaluation? If your local education agency did not test title III students, skip questions 19 and 20.	1,745		
___ Standardized norm-referenced tests (note d)		1,380	79.1
___ Criterion-referenced tests (note g)		424	24.3
___ Other tests (please specify)		515	29.5
If you do not employ standardized norm-referenced tests, skip question 20. If you do, continue.			
20. How did you report the results for the standardized norm-referenced testing?	1,319		
___ Raw scores		400	30.3
___ Grade equivalents		981	74.4
___ Percentiles		765	58.0
___ Quartiles		73	5.5
___ Stanines		283	21.5
___ Other (please specify)		64	4.8

APPENDIX II

APPENDIX II

**NOTE:** In the following five questions you will be asked to rate several aspects of the Elementary and Secondary Education Act, title III, local and State evaluation reports that affect the degree to which these documents satisfy your policy, management, and programmatic information needs. You are asked to provide overall judgments on the adequacy of the quality, informational content, and utility of the evaluation reports. Do this by considering each of these attributes: very deficient, deficient, marginal, adequate, and more than adequate. Check the box which most appropriately reflects how you feel about the respective local and State evaluations with regard to the particular aspects in question.

21. Rate the FOCUS AND SCOPE of the local and State evaluation reports. (notes b and e)

(Focus and Scope: the adequacy with which the report covers the essential and related material and the appropriateness of the emphasis and treatment given to the relevant topics, details, and high and lower priority information). (Check one box in each row.)

	Projected number responding from 8,936	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent
Local reports	2,463	35	1.4	44	1.8	531	21.6	1,396	56.7	458	18.6
State reports	2,411	16	0.6	162	6.7	607	25.2	1,221	50.6	405	16.8

22. Rate the local and State evaluation reports on the PRESENTATION OF REQUIRED MANAGEMENT INFORMATION NEEDS. (notes b and e)

(Presentation of Required Management Information Needs: the extent to which the report presents the information needed to evaluate and update policies by those who transfer policy decisions into plans, budgets, program implementation, curriculum, operational oversight, resource allocations, forecasts, status assessments and reports, educational accountability, costs, benefits, and efficiency assessments). (Check one box in each row.)

	Projected number responding from 8,936	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent
Local reports	2,471	58	2.3	147	5.9	558	22.6	1,333	53.9	376	15.2
State reports	2,420	45	1.9	204	8.4	672	27.8	1,132	46.8	367	15.2

23. Rate the local and State evaluation reports on the adequacy with which they properly QUALIFY FINDINGS. (notes b and e)

(Qualification of Findings: the extent to which the report properly qualifies the findings and assumptions and identifies those conditions and situations where the findings are not applicable). (Check one box in each row.)

	Projected number responding from 8,936	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent
Local reports	2,416	36	1.5	77	3.2	720	29.8	1,236	51.2	147	14.4
State reports	2,358	14	0.6	169	7.1	740	31.4	1,150	48.7	236	12.1

24. Rate the local and State evaluation reports on the CREDIBILITY OF FINDINGS. (notes b and e)

(Credibility of Findings: the degree of confidence expressed in the findings through statements of statistical certainty, soundness of method, evidence of replication, consensual agreements, similar experiences, supporting expert judgment and opinions, and reasonableness of assumptions). (Check one box in each row.)

	Projected number responding from 8,936	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent
Local reports	2,451	31	1.3	118	4.8	616	25.1	1,282	52.3	404	16.5
State reports	2,398	47	1.9	181	7.5	625	26.1	1,281	53.4	266	11.1

25. Rate the local and State evaluation reports on the adequacy of the QUALIFICATION AND QUANTIFICATION OF MEASUREMENT DATA. (notes b and e)

(Qualification and Quantification of Measurement Data: the extent to which the evaluation of assessments can be qualified and quantified into measurable attributes and parameters that address the problem in measurable, operational, or concrete terms). (Check one box in each row.)

	Projected number	Very deficient		Marginal		Adequate		More than adequate			
		Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-		
Local reports	2,446	39	1.6	120	4.9	810	33.1	1,121	45.8	358	14.6
State reports	2,371	62	2.6	145	6.1	778	32.8	1,031	43.5	355	15.0

26. How often do your title III, local program evaluations adequately report information on the need assessment, number of children, per pupil expenditure, project achievement and pupil benefit parameters? (Check one box in each row.) (notes b and e)

	Projected number	Very often		Generally		As often as not		Occasionally		Seldom	
		Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-	Num-	Per-
(1) Information on the manner in which the needs of children are assessed	2,358	692	29.3	1,015	43.0	125	5.3	299	12.7	227	9.6
(2) Information on the number of children in the program	2,365	1,045	44.2	857	36.2	131	5.5	111	4.7	221	9.3
(3) Per pupil expenditures of each program	2,328	640	27.5	936	40.2	212	9.1	265	11.4	276	11.8
(4) Evidence of qualifiable and measurable achievement	2,329	562	24.2	1,070	45.9	262	11.3	229	9.8	266	11.4
(5) Evidence of qualifiable or measurable pupil benefits	2,347	642	27.3	957	41.2	242	10.3	264	11.2	232	9.9

27. From your experience, how often can you draw comparisons between the results of your local Federal program and Federal programs in other localities from State and Federal evaluation reports? (Check one box for each type of report.)

	Projected number	Very often		Generally as not		Occasionally		No basis to judge					
		Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent				
State reports	2,428	112	4.6	523	21.5	110	4.5	518	21.3	424	17.5	742	30.6
Federal reports	2,433	72	3.0	309	12.7	209	8.6	331	13.6	567	23.3	945	38.9

Section D: Complete this section only if you have a funded Title VII program.

28. The demonstration of program results should be an important factor in the continuance of program funding from the Office of Education. Consider each of the following types of findings used to assess programs. FROM YOUR EXPERIENCE, which types of results do you think are more likely to impress OE or State education agency program officers? Indicate your answer by rank ordering each of the following alternatives from the most to least impressive. Select the type of results which you think is most likely to impress OE or State education agency officials. Rank this list. Do the same for all the remaining categories, ranking them 2nd, 3rd, 4th, 5th, 6th, and 7th. (notes b and e)

	Projected number		1st		2nd		3rd		4th		5th		6th		7th	
	responding from 0,936	re-	Num-	Per-												
(1) Improvement in educational management or accountability	453		97	21.4	62	13.6	44	9.6	56	12.4	68	15.0	102	22.5	25	5.4
(2) Improvement in school services or facilities	453		36	8.0	38	8.4	33	7.2	75	16.6	65	14.3	117	25.7	90	19.9
(3) Student improvement through gain scores on grades or teacher ratings	450		54	12.0	5	1.1	33	7.3	61	13.6	62	13.7	73	16.1	163	36.2
(4) Student improvement through gain scores on standardized norm-referenced tests	454		203	44.8	151	33.2	44	9.7	16	3.4	6	1.2	6	1.3	29	6.4
(5) Student improvement through gain scores on criterion-referenced tests	444		19	4.2	101	22.7	63	14.1	51	21.1	74	16.6	73	16.5	22	5.0
(6) Student improvement through gains in the affective domain (i.e., likes, dislikes, interests, attitudes, motives, etc.)	444		37	8.3	25	5.5	66	14.8	96	21.6	106	23.9	29	6.5	86	19.4
(7) Improvements in curriculum and instructional methods	454		23	5.1	71	15.5	160	35.2	50	11.0	85	18.7	38	8.4	28	6.1

29. Regardless of the types of results which are likely to impress OE or State education agency program officials, what types of results are likely to impress you, an educator? Indicate your answer by rank ordering each of the following alternatives with the same procedure used in the previous question. (notes b and e)

	Projected number responding from 8,936		1st		2nd		3rd		4th		5th		6th		7th	
	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent
(1) Improvement in educational management or accountability	441		19	4.3	67	15.1	67	15.1	95	21.5	53	12.0	54	12.1	88	19.9
(2) Improvement in school services or facilities	441		20	4.5	7	1.6	93	21.0	49	11.0	107	24.3	70	15.8	96	21.8
(3) Student improvement through gain scores on grades or teacher ratings	441		89	20.1	36	8.1	29	6.6	36	8.2	79	17.8	69	15.7	104	23.6
(4) Student improvement through gain scores on standardized norm-referenced tests	441		131	29.6	39	8.9	69	15.7	58	13.1	66	15.0	51	11.6	28	6.2
(5) Student improvement through gain scores on criterion-referenced tests	441		61	13.8	87	19.6	49	11.1	48	10.8	37	8.3	95	21.6	65	14.8
(6) Student improvement through gains in the affective domain (i.e., likes, dislikes, interests, attitudes, motives, etc.)	441		107	24.3	121	27.4	36	8.5	25	5.6	34	7.6	32	7.2	86	19.5
(7) Improvement in curriculum and instructional methods	441		73	16.5	76	17.1	101	22.9	84	19.1	84	19.1	19	4.3	5	1.0

APPENDIX II

APPENDIX II

	Number of projected responses from 8,936 (note b)	Responses	
		Number (note b)	Percent (note c)
30. Are your local evaluations of federally funded programs usually performed by external or internal evaluators? (Check one.)	426		
___ Internal (e.g., local education agency staff)		205	48.1
___ External (e.g., consultants)		91	21.3
___ Both		131	<u>30.7</u>
			<u>100.0</u> %
31. Which of the following types of tests did your local education agency administer for its title VII evaluation? If your local education agency did not test title VII students, skip questions 31 and 32.	394		
___ Standardized norm-referenced tests (note d)		357	90.7
___ Criterion-referenced tests (note g)		148	37.5
___ Other tests (please specify)		91	23.0
32. How did you report the results for the standardized norm-referenced testing?	396		
___ Raw scores		132	33.4
___ Grade equivalents		288	72.7
___ Percentiles		204	51.6
___ Quartiles		48	12.0
___ Stanines		35	8.8
___ Other (please specify)		28	7.1

APPENDIX II

APPENDIX II

**NOTE:** In the following five questions you will be asked to rate several aspects of the Elementary and Secondary Education Act, title VII, local and State evaluation reports that affect the degree to which these documents satisfy your policy, management, and programmatic information needs. You are asked to provide overall judgments on the adequacy of the quality, informational content, and utility of the evaluation reports. Do this by considering each of these attributes: very deficient, deficient, marginal, adequate, and more than adequate. Check the box which most appropriately reflects how you feel about the respective local and State evaluations with regard to the particular aspects in question.

33. Rate the FOCUS AND SCOPE of the local and State evaluation reports. (notes c and e)

(Focus and Scope: the adequacy with which the report covers the the essential and related material and the appropriateness of the emphasis and treatment given to the relevant topics, details, and high and lower priority information). (Check one box in each row.)

	Projected number re- spond- ing from 8,936	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent
Local report	363	0	0	11	3.0	146	40.1	175	48.1	32	8.7
State reports	314	24	7.7	23	7.2	143	45.6	120	38.3	4	1.3

34. Rate the local and State evaluation reports on the PRESENTATION OF REQUIRED MANAGEMENT INFORMATION NEEDS. (notes b and e)

(Presentation of Required Management Information Needs: The extent to which the report presents the information needed to evaluate and update policies by those who transfer policy decisions into plans, budgets, program implementation, curriculum, operational oversight, resource allocations, forecasts, status assessments and reports, educational accountability, costs, benefits, and efficiency assessments). (Check one box in each row.)

	Projected number re- spond- ing from 8,936	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent
Local reports	363	0	0	17	4.7	135	37.1	163	45.0	48	13.2
State reports	314	18	5.7	28	8.8	139	44.2	121	38.4	9	2.9

35. Rate the local and State evaluation reports on the adequacy with which they properly QUALIFY FINDINGS. (notes b and e)

(Qualification of Findings: the extent to which the report properly qualifies the findings and assumptions and identifies those conditions and situations where the findings are not applicable). (Check one box in each row.)

	Projected number re- spond- ing from 8,936	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent
Local reports	363	0	0	17	4.6	127	34.9	183	50.5	37	10.1
State reports	314	23	7.3	29	9.3	165	52.6	95	30.1	2	0.6

36. Rate local and State evaluation reports on the CREDIBILITY OF FINDINGS. (notes b and e)

(Credibility of Findings: the degree of confidence expressed in the findings through state- ments of statistical certainty, soundness of method, evidence of replication, consensual agree- ments, similar experiences, sup- porting expert judgment and opinions, and reasonableness of assumptions). (Check one box in each row.)

	Projected number re- spond- ing from 8,936	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent	Num- ber	Per- cent
Local reports	362	1	0.3	28	7.7	143	39.4	138	38.2	52	14.4
State reports	314	24	7.7	25	7.8	142	45.1	122	38.8	2	0.6

APPENDIX II

APPENDIX II

37. Rate the local and State evaluation reports on the adequacy of the QUALIFICATION AND QUANTIFICATION OF MEASUREMENT DATA. (note b and e)

(Qualification and Quantification of Measurement Data: the extent to which the evaluation assessments can be qualified and quantified into measurable attributes and parameters that address the problem in measurable, operational, or concrete terms). (Check one box in each row.)

	Projected number responding from 8,936	Very deficient		Deficient		Marginal		Adequate		More than adequate	
		Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent
Local reports	349	2	0.4	3	0.9	138	39.6	165	47.3	41	11.8
State reports	295	25	8.3	21	7.1	106	35.9	127	43.1	17	5.6

38. How often do your Elementary and Secondary Education Act, title VII, local program evaluations adequately report information on the need assessment, number of children, per pupil expenditure, project achievement, and pupil benefit parameters? (Check one box in each row.) (notes d and e)

	Projected number responding from 8,936	Very often		Generally		As often as not		Occasionally		Seldom	
		Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent
(1) Information on the manner in which the needs of children are assessed	369	146	39.4	161	43.6	4	1.1	45	12.1	14	3.8
(2) Information on the number of children in the program	368	198	53.7	106	28.8	2	0.5	50	13.5	13	3.5
(3) Per pupil expenditures of each program	367	89	24.3	103	27.9	69	18.8	79	21.4	28	7.6
(4) Evidence of qualifiable and measurable achievements	369	135	36.6	152	41.2	28	7.6	40	10.8	14	3.8
(5) Evidence of qualifiable or measurable pupil benefits	368	148	40.1	123	33.4	28	7.6	54	14.5	16	4.3

APPENDIX II

APPENDIX II

39. From your experience, how often can you draw comparisons between the results of your local Federal program and Federal programs in other localities from State and Federal evaluation reports? (Check one box for each type of report.) (notes b and e)

	Projected number responding from 8,936	Very often		Generally		As often as not		Occasionally		Seldom		No basis to judge	
		Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent
State reports	361	40	11.1	72	19.9	12	3.3	104	28.8	37	10.2	97	26.9
Federal reports	374	22	5.9	39	10.4	34	9.1	95	25.4	88	25.5	97	25.9

a/We requested that the questionnaire be completed by local education agency officials familiar with the local agency's evaluation efforts conducted for Elementary and Secondary Education Act, titles I (regular, migrant, neglected and delinquent, etc.), III, and VII, and the local agency's own assessment efforts. The questionnaire was divided into four sections: section A, National Assessment of Educational Progress and local education agency testing programs; section B, Elementary and Secondary Education Act, title I; section C, title III; and section D, title VII evaluation efforts. We suggested that the questionnaire be taken apart and distributed to persons most familiar with each part, such as the local education agency director of research, planning, and evaluation for section A, the respective local education agency directors of title I for section B, title III for section C and title VII for section D. Some of the questions from the questionnaire have been omitted from this report. Questions pertaining to the National Assessment of Educational Progress and local education agency's own testing programs are shown in GAO's report to the Congress (HRD-76-113), dated July 20, 1976.

b/In April 1975 we sent the questionnaire to a national statistical sample of 832 local school districts. By June 1975 we received responses from 710 school districts or 85 percent. The numbers shown above represent the number of local school districts in the Nation--out of the 11,666 in the defined universe with 300 or more pupils--to which our local questionnaire sample responses have been projected. We projected a total of 8,936 local education agencies responding instead of 11,666 for technical reasons--based on the weighting and the response rates across the various strata in our sample, this method allows us to obtain the most accurate percentage breakdowns on the answers given. Where the number of responses to each line item do not total to the "Number of projected responses" column, it is due to rounding.

c/This column shows for each question the percentage of projected respondents choosing each specific answer.

d/Standardized norm-referenced tests are tests which purport to assess the individual student's ability or achievement in broad subject areas as compared to the rest of the test population (e.g., the Metropolitan Achievement Test (MAT) or the Wechsler Intelligence Scale for Children (WISC)).

e/The percent columns show the percentage of projected respondents to each line item choosing each category. Where the percentages on each line do not add to 100.0, it is due to rounding.

f/Total does not add to 100.0 due to rounding.

g/Criterion-referenced tests are tests specifically constructed to measure students' attainment of specific educational objectives or proficiency with specified curriculum material. These tests, which may be standardized, usually provide a specific and operational description of the level and type of task performance or behavioral measures used as a criterion to indicate attainment of the educational objectives. For example, the student must be able to compute the correct product of all single digit numerals greater than zero with no more than five errors.



DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE  
OFFICE OF THE SECRETARY  
WASHINGTON, D.C. 20201

**JUN 15 1977**

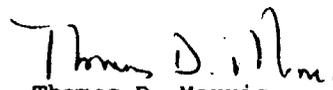
Mr. Gregory J. Ahart  
Director  
Human Resources Division  
United States General  
Accounting Office  
Washington, D.C. 20548

Dear Mr. Ahart:

The Secretary asked that I respond to your request for our comments on your draft report entitled, "Problems and Needed Improvements In Evaluating Office of Education Programs." The enclosed comments represent the tentative position of the Department and are subject to reevaluation when the final version of this report is received.

We appreciate the opportunity to comment on this draft report before its publication.

Sincerely yours,

  
Thomas D. Morris  
Inspector General

Enclosure

Comments of the Department of Health, Education, and Welfare on the Comptroller General's Report to the Congress entitled "Problems and Needed Improvements in Evaluating Office of Education Programs," February 22, 1977, B-154031 (1)

General Department Comments

The subject GAO report critically assesses the Office of Education's program evaluation mechanisms, concludes that opportunities exist for improvement, and recommends that the Secretary of HEW direct the Commissioner of Education and the Director of the National Institute of Education to take a number of corrective actions. Several changes would strengthen the report. There needs to be more careful consideration of the costs of evaluation. While there is an assertion (p. 50) that "the amount of funds spent to evaluate State and/or local education agency Title I, III, and VII elementary and secondary education programs is substantial," on p. 33 the amounts are respectively 0.2% of Title I and 1.8% of Title III. Twenty cents per \$100 is not "substantial."

While the quality of data the report appears to expect would require significant additional resources and this expenditure is not possible within existing budgets, it would also be high in relation to the possible payoffs through improvements in the programs. Are the increased costs of providing better information to State and local decision-makers likely to result in sufficient improvements to justify the expenditure? Do State and local decision-makers have the incentives and does the state-of-knowledge permit improvements or economies at the local level equal (at least) to the increased costs of evaluation improvements? If the costs of improved evaluation are not recoverable in improved productivity or efficiency, calling for more is a questionable recommendation. As another example, the call for the NIE to support more research on criterion-referenced testing is acceptable, but this is, as recognized in the report, a long-term effort, and it is not clear whether GAO believes NIE deserves additional appropriations for such an effort. NIE can hardly divert substantial resources from its present \$70 million.

A final point deserves mention. In spite of the reports of three of the four key congressional committee staff members that OE's evaluations "often have been completely ineffective or have had little impact on legislation" (p. 21), there is a growing body of professional opinion that OE's studies have, over the past ten years, been responsible for many major changes in existing legislation. In retrospect, even though apparently many of the reports did not reach "Congressional decision-makers" in a timely manner, the reports were solid pieces of work, widely discussed and debated, and by the time the next piece of legislation came up, the climate of opinion about particular programs had changed.

As a result of such retrospective analysis, the approach to evaluation which assumes there are certain "decision-makers" and that effective evaluations provide timely data to them is increasingly being questioned. Rather it appears that effective evaluations affect the broad political climate within which particular decisions are made. The large scale OE evaluations have stood the test of time in this view of evaluation far better than any of the alternatives proposed in the report appear likely to. (To underscore this point witness the large number of citations of these large-scale OE studies in the report itself.)

In short, the report is well done within its limits. The recommendations do not give adequate recognition to the additional financial burden they entail, nor to whether the tradeoffs in improved program quality are likely to justify the additional expenditures. And there is no discussion of the limitations of the view that evaluation ought to serve decision-making in a direct and linear sense.

#### GAO Recommendation

That the Secretary of HEW direct the Commissioner of Education to more strongly emphasize the purpose of providing information to the Congress when planning, implementing, and reporting on evaluation studies. In particular, more attention should be given to timing the studies so that they closely coincide with the legislative cycle, (See GAO note p. 124) and briefing congressional committee staff more frequently.

#### Department Comments

We concur with the general conclusion that Congressional needs can be better served than we are now doing. First, we concur that it is clearly desirable that evaluation studies should be timed to more closely coincide with the legislative cycle. It is not the case, however, as GAO has apparently concluded, that this obviously important problem has not received attention. Rather, the frequent failure to coincide studies with the legislative cycle has been due to such factors as: inadequate funds to initiate evaluations at the right time; delays in the procurement cycle due to uncertainties in the appropriation process; increased difficulty in getting studies and data collection instruments cleared; and failure to anticipate difficulties and delays in the data collection process in the field. Nevertheless, we believe these problems can be more effectively addressed and, in order to do so, we have initiated a series of reviews of our studies which focus on predicted production dates for findings and recommendations vis-a-vis critical dates for input to legislation renewal.

Second, we concur that congressional committee staff should be briefed more directly and fully on the findings of evaluation studies. This need has not received the attention it deserves, but we have recently made the decision to institute such briefings on all major evaluation studies, and expect to get this new procedure underway in the coming weeks.

(See GAO note p. 124)

GAO Recommendation

That the Secretary of HEW direct the Commissioner of Education to better define the program objectives to be evaluated as required by the General Education Provisions Act. This includes translating the legislative purposes of each program into specific qualitative and measurable program objectives, and clearly stating these objectives, and the progress made toward achieving them, in the annual evaluation report.

Department Comments

We do not concur. In most cases, legislation fails to state a program's objectives with sufficient clarity that they are readily susceptible to evaluation. But very often this failure of the legislation to be clear and precise on program objectives is the price paid through political compromise for getting the legislation passed at all. The Office of Education proceeds at considerable peril in trying to further specify legislation. It has in fact been criticized on several occasions for going further than the Congress intended, and of "trying to legislate by means of regulation." Furthermore, in many cases it has been the Congress' specific intention to avoid specification of program objectives and to leave such judgments and decisions up to State and local officials.

For example, in turning down a proposed amendment to concentrate seventy-five percent of compensatory funds on basic skills Title I of ESEA, the Senate Committee said:

"The Committee believed it inappropriate for the Federal government to sustain its judgments on appropriate Compensatory Education Programs for that of State and local officials." (Senate Report 93-765, 1974, p. 30)

On the same question the House observed:

"The Committee feels strongly that the Local School Agency is the appropriate level to determine the special needs of educationally deprived children and should be primarily responsible for determining approaches to meeting those needs." (H Report 93-805, 1974, p. 20-21)

For all these reasons, we believe there are very definite limits on the Office of Education's authority and ability to increase the clarity and specificity of program objectives.

GAO Recommendation

That the Secretary of HEW direct the Commissioner of Education to improve the implementation of evaluation results by giving greater attention and priority to procedures such as the issuance of Policy Implication Memorandums designed to insure implementation of those results.

Department Comments

We concur. The need to improve the utilization and implementation of evaluation reports is a major problem which all Federal agencies and the Congress jointly face, and we further concur that increased efforts should be devoted to its solution. The Policy Implications Memorandum (PIM) is an invention of OE's evaluation office, and while its potential for improving the utilization of evaluation findings is considerable, GAO is correct in observing that it has not been used nearly as extensively in OE as it should have been. Efforts are currently underway to expand the production and use of the PIM. We are now conducting periodic reviews of the production schedule for PIMs and emphasizing their high priority.

GAO Recommendation

That the Secretary of HEW direct the Commissioner of Education to assess whether State and/or local evaluation reports for Title I and VII of the Elementary and Secondary Education Act can be improved so that they supply officials at Federal, State, and/or local levels with the reliable program information they need for decision making.

GAO Recommendation

That the Secretary of HEW direct the Commissioner of Education to review the types of State and/or local program information collected on programs authorized by Titles I, III, and VII of the Elementary and Secondary Education Act.

Department Comments

We concur with the general thrust of GAO's recommendations in this area, but most of the actions that GAO recommends are already underway, many by legislative mandate.

With respect to Title I, new evaluation requirements (Section 151 created by P.L. 93-380) directed the Commissioner to develop evaluation models and standards and to provide technical assistance to the States and local districts in order to improve the quality of local Title I evaluations and to yield comparable data which could be aggregated to State and National levels. All of this is being carried out: OE has interviewed personnel in policy making roles at the Federal level (both Congressional

and HEW staff) to determine the kinds of information they felt should be included in the annual State and local evaluation reports; an advisory group of personnel from the different operating levels of the Title I program (including parent representatives) indicated the kinds of information they thought should be included in the reports; and evaluation models and their associated reporting forms were developed and reviewed by each State agency and three of its locals to determine the kinds of problems they might have in using them. As a result of these efforts, a limited core of essential information was identified as being desirable at the Federal level. This core of information is substantially less than what has frequently been contained in State and local reports in the past. It will become the Federal evaluation requirement when regulations for this portion of the legislation are published.

OE has sponsored a series of workshops for State and local evaluation staff in the use of the standard evaluation models and reporting system. We have established ten technical assistance centers--one for each of the HEW regions--to provide technical assistance and training for State and local staff on a continuing basis. We have prepared training manuals and guidebooks for widespread dissemination to current and future users of the models and reporting system. Once the models and system are in place across the nation, their use should result in data which can be aggregated across States and across school districts. And, once they are in place, OE will be better able to assess whether or not the resultant data are sufficiently free of systematic errors to support aggregations to the State and national levels in a satisfactory manner. If they are not, then a determination can be made as to whether technical problems existed that could be overcome or whether different kinds of studies needed to be done to satisfy Federal, State, and local reporting requirements.

With respect to the recommendation on ESEA Title VII local evaluation reports, we believe they can be improved and we have taken the following steps to do so. Recently published Title VII regulations strengthen the requirements for evaluation of LEA bilingual projects. In addition, NIE and OE have a joint project underway to upgrade the technical expertise of persons responsible for local evaluations. While we believe it is worthwhile to improve local evaluations, the extent to which such evaluations will be useful at the Federal level is not yet evident. Certainly the problems of aggregating across local bilingual evaluations are more severe than in Title I, if only because of the multiplicity of languages involved.

(See GAO note p. 124)

Thus, as regards the much needed and legislatively mandated actions to improve State and local evaluations and reporting under Titles I, IV, and VII, we believe GAO's understanding is incomplete (See GAO note p. 124) that the problems they refer to are well understood by both the Office of Education and Congress and that appropriate actions to deal with them are already well underway, and in some cases near completion.

GAO Recommendation

That the Secretary direct the Director, National Institute of Education, to consider the need for funding additional research on (1) criterion-referenced tests and other alternatives to standardized norm-referenced achievement tests for uses which include program evaluation and (2) the nature and extent of racial, sexual, and cultural biases in standardized tests and how such biases may be reduced.

Department Comment

We concur. Several groups within the Institute presently have ongoing research programs in criterion-referenced testing and test bias. More emphasis will be given to these programs in FY 1977, FY 1978, and FY 1979 if appropriations for the Institute are increased significantly above present levels.

Under NIE leadership, the Center for the Study of Reading at the University of Illinois and the Center for the Study of Evaluation at UCLA are doing research that will lead to alternatives to standardized achievement tests in reading comprehension and writing. In addition, statistical research is in progress on how to assess the probability of making errors in classifying students on such tests, how to set the length of such tests and how to determine passing scores.

The SOBER-Espanol project funded by NIE is developing criterion-referenced tests to assess competency in reading Spanish for grades K-6. The SOBER system allows teachers to create "tailor-made tests" by matching prepared test items to reading objectives. Currently, K-3 tests are being published and distributed through Science Research Associates. Grades 4-6 will be published later this year.

NIE is also supporting research and development on criterion-referenced testing and other alternatives to norm-referenced achievement tests for the purpose of educational exit testing and occupation entry selection.

In December 1975 NIE held a conference for test developers, test critics and others to consider methods of identifying and eliminating bias in reading achievement tests. Since that time, NIE has funded two grants on test bias: one is a project to detect and eliminate the motivational causes of test bias and the other is a project to debias the language found in a widely used, standardized achievement test. And finally, NIE has an in-house project which applies new methods of qualitative data analysis to the intractable problem of detecting bias in test items.

NIE is also supporting work on sex bias in the assessment of a person's occupational interests and biases in educational exit and occupational entry testing--the latter being of particular importance given the Griggs v. Duke Power decision of the U.S. Supreme Court. The products of these studies include the NIE Guidelines on Sex-Fair Vocational-Interest Measurement and the Abt kit on the interpretation and usage of sex-fair vocational-interest tests.

GAO Recommendation

That the Secretary direct the Director, National Institute of Education, to improve dissemination of available NIE-funded information, which is intended to help in selecting the most appropriate standardized tests, thereby increasing State and local education agency officials' awareness and use of this information.

Department Comment

We agree that there is a need to improve the dissemination of NIE-funded material intended to help educators select appropriate standardized tests. The Institute's policy is not to force its products on school personnel, but we do intend to make school personnel familiar with the products that are available. In the case of our consumer's guides to standardized tests, that is, the test evaluation books produced for NIE by the Center for the Study of Evaluation at UCLA and mentioned in this GAO report, our Basic Skills Group and our Dissemination and Resources Group are collaborating in the dissemination of these products. Several approaches are or will be used. First, the test evaluation books are listed in the Institute's Catalog of NIE Education Products. This catalog has been offered free to each superintendent of schools and each district director of curriculum in the country. Five thousand copies have been distributed to superintendents. A similar distribution will now be made to the district directors of evaluation, who presumably will be particularly interested in the test evaluation books. In addition, we plan to use the newly formed Lab and Center R&D Exchange to give school personnel more information about the acquisition and use of test evaluation books (e.g., through brochures or workshops). This dissemination network has the potential of reaching 50 percent of the country's school systems. Still another dissemination network that will be used in a similar way is the one formed by our seven R&D utilization contractors (five State education agencies and two nongovernmental organizations). If the GAO report is correct in identifying a need for the test evaluation books--and we think that it is--these approaches should give the books much wider dissemination than they have had up to this point.

NIE is also sponsoring the dissemination of test information in more specialized areas. The Education and Work Group has supported consumer guides to standardized tests in career education and occupational preparation. Dissemination of products will also be built into the further work of this group to improve testing in career education. The Educational Equity Group has funded American Institutes for Research to develop a catalog reviewing assessment instruments for children of limited English-speaking ability at the K-6 levels. Descriptive information (author, publisher, research data available, etc.) and analyses of the appropriateness of the tests for use with bilingual children will be provided. All information will be comprehensible to educational practitioners. This contract will also identify those areas and levels for which existing tests are inadequate or nonexistent.

(See GAO note below)

GAO note: Deleted comments pertain to material presented in the draft report which has been revised or not included in the final report.

ADDITIONAL SUGGESTIONS BY OE CONFEREESFOR IMPROVING TESTING AND EVALUATION

As discussed in chapter 6, the Office of Education sponsored a special 4-day conference on "Achievement Testing of Disadvantaged and Minority Students for Educational Program Evaluation" in May 1976. The Office invited about 50 experts in testing, program evaluation, and related fields, including university and other researchers, and representatives of leading test publishers. Federal and local education agencies, as well as education and other interest groups, were also represented. The conference focus was on large-scale program evaluations of elementary and secondary school compensatory and desegregation programs--programs on which OE concentrates much of its effort. The purpose of the conference was to identify, define, and analyze the many problems associated with using standardized achievement tests in these programs, and was to develop interim and long-term solutions.

In addition to the suggestions discussed in chapter 6, conclusions and recommendations from the five small working groups formed at the conference's close included the following:

- In the context of educational program evaluation: development, standards, administration, and use of standardized tests must be accounted for and monitored. Alternative approaches which should be explored include: (1) Federal legislation to establish a monitoring body with enforcement powers to oversee the testing practices of test developers, State and school district educational systems, researchers, and others, (2) an independent monitoring agency sponsored by major test developers composed of minority and organizational representatives, and (3) an enforceable testing code of ethics, including mandatory withdrawal of services by test producers in established cases of misuse of tests and test information.
- There is need for federally funded studies to increase understanding of the nature and extent of biases in tests and how such biases might be reduced. Studies of this problem might include detailed studies of individual pupils in interviews or computer simulations of bias models.

- There is no clear consensus on test bias definitions nor clear technical procedures for identifying a biased question or test. There is, nevertheless, enough documentation of public concern--including calls for cessation of testing--and empirical data to justify change and development of guidelines. It is imperative that the testing community join with other interested groups to agree on the steps to be taken to develop and use tests that are judged to be fair.
- Specific guidelines are also needed for test administration and assessment of bilingual groups.
- There clearly needs to be an extension of the present professional testing standards of the American Psychological Association to cover the use of achievement tests in program evaluation.
- OE should support developing a procedures manual for determining the appropriateness of using achievement tests in program evaluation and properly selecting, administering, scoring, and interpreting data from such tests. Such a manual should include the degree that other information must be used together with achievement tests to adequately describe program outcomes. This should be the first (and more immediately accomplishable) stage of a longer term effort to produce procedures manuals that address means other than achievement tests for collecting program evaluation data. For the longer term effort, more experimentation (field-testing) is needed to investigate alternative data collection modes to firmly establish them.
- Publishers of standardized tests should give explicit step-by-step instructions in their users' or technical manual about how to use their tests correctly for various purposes and how to avoid misuse. For example, these purposes may include needs assessment, diagnosis and prescription, or project evaluation.
- The use of grade-equivalent scores on standardized tests should be eliminated.
- More tests ought to be developed for diagnosing educational problems and prescribing remedies.

--The work statements for evaluations in Federal agency requests for proposals are not always adequate. Examples can be cited in which technical approaches have been overspecified by persons who perhaps have not fully understood either the technical or the practical problems involved. More time needs to be allocated for writing requests for proposals, and more professional review of them must be accomplished. Detailed technical and procedural specifications should never be included in work statements unless such specifications are the consensus of a panel of national experts in the field.

PRINCIPAL OFFICIALS OF THE  
DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE  
RESPONSIBLE FOR ACTIVITIES DISCUSSED IN THIS REPORT

	<u>Tenure of office</u>	
	<u>From</u>	<u>To</u>
<b>SECRETARY OF HEALTH, EDUCATION, AND WELFARE:</b>		
Joseph Califano	Jan. 1977	Present
David Mathews	Aug. 1975	Jan. 1977
Caspar W. Weinberger	Feb. 1973	Aug. 1975
Frank C. Carlucci (acting)	Jan. 1973	Feb. 1973
Elliot L. Richardson	June 1970	Jan. 1973
<b>ASSISTANT SECRETARY FOR EDUCATION:</b>		
Mary Berry	Apr. 1977	Present
Philip Austin (acting)	Jan. 1977	Apr. 1977
Virginia Y. Trotter	June 1974	Jan. 1977
Charles B. Saunders, Jr. (acting)	Nov. 1973	June 1974
Sidney P. Marland, Jr.	Nov. 1972	Nov. 1973
<b>COMMISSIONER OF EDUCATION:</b>		
Ernest L. Boyer	Apr. 1977	Present
William F. Pierce (acting)	Jan. 1977	Apr. 1977
Edward Aguirre	Oct. 1976	Jan. 1977
William F. Pierce (acting)	Aug. 1976	Oct. 1976
Terrel H. Bell	June 1974	Aug. 1976
John R. Ottina	Aug. 1973	June 1974
John R. Ottina (acting)	Nov. 1972	Aug. 1973
Sidney P. Marland, Jr.	Dec. 1970	Nov. 1972
Terrel H. Bell (acting)	June 1970	Dec. 1970
James E. Allen, Jr.	May 1969	June 1970
<b>DIRECTOR, NATIONAL INSTITUTE OF EDUCATION:</b>		
John Christensen (acting)	July 1977	Present
Emerson J. Elliott (acting)	Jan. 1977	July 1977
Harold L. Hodgkinson	July 1975	Jan. 1977
Emerson J. Elliott (acting)	Oct. 1974	July 1975
Thomas Glennan	Oct. 1972	Oct. 1974