LeSAGE, Jr., William DoVel, 1920-
ACCOUNTABILITY--THE RELEASE OF
SCHOOL-BY-SCHOOL COMPARATIVE TEST DATA.

The Ohio State University, Ph.D., 1971
Education, general

University Microfilms, A XEROX Company, Ann Arbor, Michigan
ACCOUNTABILITY - THE RELEASE OF
SCHOOL-BY-SCHOOL COMPARATIVE TEST DATA

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

William Dovel LeSage, Jr., B.S., M.A.

* * * * *

The Ohio State University
1971

Approved by

[Signature]
Advisor
College of Education
PLEASE NOTE:

Some pages have small and indistinct type. Filmed as received.

University Microfilms
ACKNOWLEDGEMENTS

A dissertation is the product of no one individual. This writer owes a debt of gratitude to many persons. Only a few of the most important can be named in this space.

First, I wish to express my sincere appreciation to my friend and major advisor Dr. Anthony C. Riccio. His guiding wisdom and constant counsel were invaluable, not only in this dissertation but in my entire program.

No less valuable were the efforts of Drs. Joseph Quaranta, Harold Reynard, and James Wigtil, who as mentors and members of my committee made this study possible.

Drs. Joseph Davis and Keith Barnes of the Columbus Public Schools are also due my gratitude, Dr. Davis for his encouragement to pursue the doctoral program and Dr. Barnes for his friendship and inspiration.

A particular note of appreciation is due to my friend and colleague Richard Beck who spent many hours in constant review and critique of these efforts.

Finally the entire program would have been impossible without the loving patience and devoted assistance of my wife, Iris.
VITA

October 28, 1920 Born - Huntington, West Virginia

1942. ..................... B.S. in Business Administration, Marshall University, Huntington, West Virginia

1942-1946 ..................... Personnel Officer, United States Army

1946-1952 ..................... Personnel Supervisor, Appalachian Electric Company, Huntington, West Virginia

1952-1962 ..................... President and General Manager, L & M Radio and Appliance Company Huntington, West Virginia

1962. ......................... M.A. in Secondary Education, Marshall University, Huntington, West Virginia

1962-1964 ..................... Counselor, Mad River-Green Local School District, Springfield, Ohio

1964-1967 ..................... Counselor, Columbus Public Schools Columbus, Ohio

1967-1971 ..................... Supervisor, Testing and Evaluation Services, Columbus Public Schools, Columbus, Ohio

June, 1971. ..................... Ph.D., The Ohio State University Columbus, Ohio

PUBLICATIONS

Introduction to the Columbus Testing Profile. Columbus, Ohio: Columbus Public School, 1968.


FIELDS OF STUDY

Major Field Guidance and Counseling, Dr. Anthony C. Riccio

Minor Field Adult Education, Dr. John Ohliger
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>VITA</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF ILLUSTRATIONS</td>
<td>viii</td>
</tr>
<tr>
<td><strong>CHAPTER</strong></td>
<td></td>
</tr>
<tr>
<td>I INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>4</td>
</tr>
<tr>
<td>Purposes of the Study</td>
<td>5</td>
</tr>
<tr>
<td>Definitions</td>
<td>6</td>
</tr>
<tr>
<td>Assumptions</td>
<td>8</td>
</tr>
<tr>
<td>Limitations</td>
<td>9</td>
</tr>
<tr>
<td>Significance of the Study</td>
<td>10</td>
</tr>
<tr>
<td>Design of the Study</td>
<td>11</td>
</tr>
<tr>
<td>Organization of the Study</td>
<td>12</td>
</tr>
<tr>
<td>II REVIEW OF RELATED LITERATURE AND RESEARCH</td>
<td>15</td>
</tr>
<tr>
<td>Introduction</td>
<td>15</td>
</tr>
<tr>
<td>The Rise of Accountability in its Present Concept</td>
<td>15</td>
</tr>
<tr>
<td>Performance Contracting</td>
<td>29</td>
</tr>
<tr>
<td>Voucher Systems</td>
<td>38</td>
</tr>
<tr>
<td>Decentralization</td>
<td>47</td>
</tr>
<tr>
<td>Testing for Accountability</td>
<td>54</td>
</tr>
<tr>
<td>Summary</td>
<td>72</td>
</tr>
<tr>
<td>III DESIGN OF THE SURVEY</td>
<td>79</td>
</tr>
<tr>
<td>Introduction</td>
<td>79</td>
</tr>
<tr>
<td>Organization of the Survey</td>
<td>80</td>
</tr>
<tr>
<td>Section I</td>
<td>80</td>
</tr>
<tr>
<td>Section II</td>
<td>80</td>
</tr>
<tr>
<td>Section III</td>
<td>86</td>
</tr>
<tr>
<td>Section IV</td>
<td>87</td>
</tr>
<tr>
<td>IV SUMMARY OF SURVEY DATA</td>
<td>91</td>
</tr>
<tr>
<td>Section I</td>
<td>91</td>
</tr>
<tr>
<td>Section II</td>
<td>91</td>
</tr>
<tr>
<td>Section III</td>
<td>105</td>
</tr>
<tr>
<td>Section IV</td>
<td>110</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>School Systems Reporting School-by-School Standardized Test Data Showing Grades Reported</td>
<td>92</td>
</tr>
<tr>
<td>2</td>
<td>Derived Score Used to Report Achievement Test Data and Normative Base Used</td>
<td>94</td>
</tr>
<tr>
<td>3</td>
<td>Number and Percentage of Reporting Systems Which Released, Concurrently or Separately. Various Demographic Data.</td>
<td>98</td>
</tr>
<tr>
<td>4</td>
<td>Number and Percentage of Systems Excluding Categories of Students from City-Wide Testing or Inclusion in Summary School Test Data.</td>
<td>105</td>
</tr>
<tr>
<td>5</td>
<td>Number and Percentage of Systems Planning the Reporting of Various Demographic Data in Future Public Accountability Releases.</td>
<td>107</td>
</tr>
<tr>
<td>6</td>
<td>Percentage of Reporting and Non-Reporting Systems Releasing or Considering the Release of Selected Demographic Data.</td>
<td>108</td>
</tr>
</tbody>
</table>
LIST OF ILLUSTRATIONS

Figure                             Page
1  Illustration of Differential Rates of Growth in
   Terms of Grade Equivalent Units for Different
   Percentile Positions on the C.T.B.S. Reading
   Tests ............................. 149
CHAPTER I

INTRODUCTION

Preface

The decade of the 1960's saw the beginning of many drastic changes in the use of standardized test data by the public schools of the nation. For many years, test scores were held sacrosanct by school officials. Students themselves were rarely told their own scores and even then, only in broad generalities or in hedged terms. Similar constraints were applied to releasing test data to parents on their own children.¹

Teachers in all too many instances were not sufficiently trained or skilled in test interpretation to feel confident to make such revelation to their students or the parents. Counselors, who presumably were qualified to make such interpretation, surrounded test data with such a mystique that it was the exceptional parent or student who was given any meaningful access to test results. Administrators from the top down guarded standardized test data, along with other student information, with such zeal that even where competency and willingness existed among teachers and counselor, administrative dicta successfully thwarted their efforts.²


Not only was test data on individuals placed in the "confidential" category, but any accumulated data on a school-by-school or system-wide basis was shrouded in even greater secrecy. In many school districts such data were not even compiled. If information were not available, there could be no disclosure.\(^3\)

The 1960's however, saw the beginnings of dramatic change. Legislatures of several states enacted "public disclosure" laws which forced certain school data, among other information, into the public domain. Legislators seldom act without mandate; so it is evident that some pressures of their constituency were building in the area of stripping secrecy from public affairs.\(^4\)

The 1960's also saw more specific public demands arise for accountability by school systems. Gone were the days when the mere existence of educational institutions was taken as prima facie evidence of successful education. Minority groups, particularly, were faced with unwelcome indications that their children were receiving an inferior education. They began demanding an accounting of the school's performance.

This first disclosure came in the form of the release of achievement data in terms of the average grade equivalent scores. These scores were given for certain grades, city-wide, school-by-school. News media


\(^4\)Gerald Cavanaugh, Address at meeting of Large City Test Directors, Las Vegas, Nevada, April 1, 1969.
had a field day in drawing comparisons between the "best" and "poorest" schools, and minority groups, finding confirmation of their worst expectation, castigated the schools for their failure.

With such a release by one system, the pressure on other systems grew and continues to mount today. Several other large cities have subsequently succumbed to the pressure and many others feel that their turn must come within a matter of time. As large systems yield to demands for accountability, it is a natural consequence that smaller adjacent communities similarly demand information about their own schools' performance.  

Some systems have been more comprehensive in their release of data. They, too, have revealed school-by-school achievement scores but have accompanied them with demographic data on the school and its immediate community. Information which was thought to have a bearing on school performance, such as academic aptitude test scores, attendance, teacher preparation, mobility factors and the like, was included, hopefully, to give a more complete picture of the school scene.

Implicit in the concept of accountability is the probability that one or more of the school's publics will be dissatisfied with the findings. There will be demands for change. These changes may run a broad gamut from minor curriculum revisions or personnel realignments to major organizational revamping or even elimination of the present public school system concept.

---

In most cases there will be no easy answers to the questions; no ready solutions to the problems. Since summary school-by-school test data is the most common tool of accountability, one proposed solution to the inequalities in achievement so revealed has been to discredit the tests themselves and to press for their abandonment. But this is to oversimplify, merely repudiating the symptom and eschewing the causes.

Accountability is upon the educational scene, as well it should be. Educators must be willing to face the issue and "tell it like it is." Yet the problems revealed cannot be solved by the educator alone. There must be concentrated and concerned efforts by all; administrators, teachers, parents, students, communities, taxpayers, and special interest groups, before progress can be made. How can cooperative support be enlisted from such diverse publics?

STATEMENT OF THE PROBLEM

Since these pressures for accountability are comparatively new, educators find little in the literature or in research to provide guidelines for the selection of strategies, techniques, or methods to meet these demands. Each of the school systems faced with the dilemma has been forced to blaze its own path and suffer the pains of trial by fire. Some systems have met the challenge and have come through virtually unscathed. Others have complied with the demands only to

find themselves unmercifully censured by their publics.

As other systems find themselves being pushed in the direction of school-by-school accounting, they too search for guidelines to help them determine the most meaningful and efficacious way to account to their publics. What data should be made public? Who should be involved prior to, during, and after the release of data? What are some possible consequences? How might negativistic attitudes be avoided? These and other questions plague the uninitiated. The presentations must be factual, inclusive, yet designed to enlist the positive and cooperative efforts of all concerned toward educational improvement.

PURPOSES OF THE STUDY

The purposes of this study were:

1. To develop, through a review of pertinent literature, an overview of the recent emergence of the present concept of accountability in education, and concomitant pressures for the release of school-by-school test data.

2. To explore briefly some of the major educational changes espoused by school critics and the implications in terms of testing and accountability.

3. To assess the experiences of large school systems which have released school-by-school test data, attempting to determine which strategies have been most successful in enlisting or maintaining the support of the system's various publics.
4. To ascertain the perceived level of pressure for accountability on school districts not yet reporting summary test information, and to determine their plans or thoughts relative to such release.

5. To use the previous information to develop workable guidelines which may be applied or adapted by districts yet to release such school-by-school test data.

6. To submit these guidelines to a panel of recognized experts in the fields of testing and research and solicit their critiques with the view to strengthening and improving these guidelines.

7. To develop revised guidelines on the basis of their recommendations.

DEFINITIONS

In order to provide a common frame of reference from which to work, the following terms are defined to establish the intended meaning throughout this study.

ACCOUNTABILITY - Historically, school accountability has dealt with the need for schools to show that public funds have been used wisely and prudently. The most current emphasis is that educational outcomes justify the expenditures. That is, that the public received a fair return on its investment in public education.\(^7\) Certain segments of the

school communities go farther in that they are concerned lest particular schools within the district are being slighted to the detriment of their childrens' education. It is in this academic educational outcome sense that the term "accountability" is used in this study.

LARGE SCHOOL SYSTEM - The organization of Large City Test Directors has arbitrarily defined large school systems as those city school districts in cities with over 200,000 population, and county districts (where the county is an operating unit) serving over 50,000 students. There are thirty-eight large city or county systems with student populations in excess of 70,000. It is the LCTD definition which is used in this study.

LARGE CITY TEST DIRECTORS - The Large City Test Directors organization is affiliated with the National Council on Measurement in Education, and includes those individuals directly responsible for testing programs within the large school systems of the country. Organizational patterns vary within the many systems, as do the assigned titles of the individuals directly charged with testing. Testing may be a separate department, combined with the guidance function, related to research, or under the division of instruction. The title varies also, covering the full range from assistant superintendent through director to supervisor. The individuals' direct responsibility for the testing program is the criterion for his being named Test Director.

---


9Ibid.
MINORITY GROUPS - For the purposes of this study, minority groups are those groups which are recognized as minorities or subcultures in the nation as a whole, such as Black, Puerto Rican, Oriental, American Indian, Cuban, Mexican-American, Appalachian White, and the like. They may or may not represent a minority in a particular setting, as in the case of Washington, D.C. where Blacks, a "minority" group, actually represent over 90 per cent, or a definite majority of the city's population.

ASSUMPTIONS

The following assumptions are made in the development of this study:

1. That the trend toward proliferation of demands for accountability will continue. As large school systems have been the prime past targets, other similar systems will bear the immediate brunt of future demands.

2. In the large systems which have made public their school-by-school data, the initial pressures have stemmed largely from minority groups. Since the focal point of minority group concentrations is in large metropolitan areas, a study of the trends in such areas will be meaningful in determining guidelines for meeting these pressures for accountability.
3. Large city test directors, who have to deal daily with the demands, and who are in constant communication with the publics in their individual school communities, are in an excellent position to assess the feelings and reactions within these communities.

4. A study of these large school systems' experiences with accountability pressures, both those reporting and non-reporting districts, should yield valuable data upon which to develop guidelines for meeting such demands in a manner most likely to invoke community support for the schools and their improvement.

5. If guidelines can be developed whereby the needs for responsible accounting are met within the framework of a positive and constructive assessment of educational accomplishment, these guidelines should have broad applicability for all school systems, regardless of size or ethnic composition.

LIMITATIONS

As in any non-experimental research effort, several limitations are apparent. The first is in the review of literature and research. Time dictates that all the literature and research in the area designated for review could not be covered. Since the proliferation of pressures for accountability and school-by-school release of summary test data
was a phenomenon of the 1960's, this temporal factor reduced the span of the review.

The second limitation is the size of the sample of school systems to be surveyed. Since the initial pressures for summary data release have come mainly from minority groups, usually those in the inner city,\(^\text{10}\) the survey of thirty-eight school districts in major population centers yielded an adequate representative assessment of pressures and practices while maintaining a manageable sample.

SIGNIFICANCE OF THE STUDY

The American society is coming more and more to demand excellence and equality of opportunity in the education of its youth. Several factors impinge upon these demands. Some of these are:

1. Technological advances have been accompanied by an increasing demand for education, both in terms of quantity and quality.

2. The rising cost of education with the consequent demands upon the taxpayer, makes him give closer scrutiny toward value received. Is he getting his educational dollar's worth?

3. With the current emphasis on civil rights, minority groups are becoming more and more aware of educational deficiencies and, as a consequence, more imperative in their demands for equitable educational opportunities.

4. Proof of the satisfaction of these demands is dependent upon evaluation, and the evaluations are in turn dependent upon testing.

Present pressures for accountability stem largely from the third factor, but as a consequence of the first public disclosures of school-by-school test data, school publics are taking a closer look at their schools. It is not farfetched to project that in the near future, public accountability for educational excellence will be the rule rather than the exception. It is to be hoped that the experience of those schools which have taken "the step," or will do so in the near future, will provide a meaningful vehicle for all systems to use in taking a closer look at their educational efforts. Educators must face the stark reality that their effectiveness is no longer being taken for granted. They must be able to provide objective proof of their program's worth, or be taken to task by their supporting publics.

DESIGN OF THE STUDY

The method of research used in this study was nonexperimental and developmental in nature. First, a review of literature and research as conducted. This review covered the fields of educational accountability and the release of school-by-school test data, as these two areas are held to be directly interrelated. The review focused upon such factors as:

1. The emergence of educational accountability in its current
concept and some possible circumstances which contributed to its rise.

2. The trends toward school-by-school release of summary test data and some typical reactions to these data by various school publics.

3. Some changes suggested by school critics in light of dissatisfactions engendered by reported performances, and the implications these hold for accountability and testing.

Second, a questionnaire was addressed to the Test Director of thirty-eight of the nation's largest school systems. These systems represent all regions of the country and collectively have significant representation of all major minority and ethnic populations in the country. The questionnaire was designed to elicit information as to the system's policy and practices regarding the release of school-by-school data, and the impact of various elements of accountability upon the school and community.

Third, from the research of literature and from the experiences and recommendations of these Large City Test Directors, this writer attempted to formulate guidelines reflecting those strategies and techniques which have proved most successful in eliciting positively oriented reactions of all concerned, or which best represent the consensus of the Directors and the writers represented in the literature toward that end.
Fourth, once these guidelines, which deal with all phases of educational accountability—testing, pre-release involvements, release techniques, and alternative educational plans or organizations—had been formulated, they were submitted to a panel of experts in the fields of testing and educational research. Members were drawn from among authorities in the field of test publications, professors in educational research and measurement who had large city or state department of education experience, professors in the fields of curriculum and programming, and school testing and research administrators apart from those included in the survey.

ORGANIZATION OF THE STUDY

The problem of the study has been stated; the nature, purpose and significance of the study have been noted; the definitions and limitations pertaining to the study have been posed. In the remainder of this dissertation, Chapter II is a review of literature and research which has a bearing on accountability, the release of summary test data, and alternative educational or organizational plans as they relate to accountability. Chapter III pertains to the development of the survey questionnaire used to accumulate data from the large city school systems surveyed.

Chapter IV details, summarizes, and interpretes the data obtained from the questionnaires. Chapter V is divided into four sections. The first section draws upon the literature and the questionnaire data to develop recommendations which may serve as guidelines for school systems
entering the era of accountability. Section two reports upon the presentation of these recommendations to a panel of experts in the fields of testing, research, and program development. Their critiques form the basis for amending the recommendations which are then presented in concise revised form. A general summary of the dissertation is then presented in the third section, while the final section contains the conclusions of the study.
CHAPTER II

REVIEW OF RELATED LITERATURE

INTRODUCTION

A review of the literature was conducted to determine the factors which have led to the emergence of this new concept of educational accountability. At the same time, an attempt was made to identify the major trends in educational organization or practice which seem to be arising as a result of educational accounting and to assess their potential impact on future accountability.

The literature reviewed has been divided into three sections. The first section is a review of the development of the current concept of accountability, some causal factors which led to its rise, and some implications for teacher evaluation. The second section studies briefly three organizational manifestations which have resulted largely from public dissatisfaction with school performance, namely, performance contracting, the voucher systems, and decentralization. These are discussed in light of their implication for accountability. The last section in this chapter is concerned with the implications for testing under accountability.

THE RISE OF ACCOUNTABILITY IN ITS CURRENT CONCEPT

Accountability, according to Webster's New World Dictionary, is "the condition of being accountable, liable, or responsible."
Accountability has been an inherent component of public education for as long as public funds have been allocated in support of schools. Yet this accountability has been considered primarily in terms of fiscal responsibility. Checks and balances have been built into the funding mechanisms to ensure that these public monies were properly expended. Local fund budgets are scrutinized by some non-school agent such as the county auditor, while state and federal fund expenditures are similarly screened by auditors at those respective levels.

As late as 1960 the emphasis for accounting remained fiscal. Johns and Morphet define accountability as the need for public school officials to show "where does the money come from and where does it go." This is indicative of the strictly fiscal emphasis at that time.

The term "accountability," however, has only recently entered the general educator's lexicon, and in a context quite different from the traditional fiscal sense. There is in the newer definition the additional requirement that educational outcomes justify the expenditures.

While there is no established standard usage of the term "accountability," there is little doubt about its general meaning and its import for schools. The basic idea that it conveys is that school systems and schools, and more importantly, the professional educators who operate them, should be held responsible for educational outcomes. If this can be done, it is maintained, then there should be favorable changes in

---


the professional performances of school staffs which will be reflected in greater academic achievement, improvement in pupil attitudes, and generally better educational results.³

This proposition, that higher quality education can be obtained by making professional educators responsible for student performance, is what makes accountability so attractive to the public.

According to Barro, there have been at least four major strands of action, or thought, in current education for which established some precedence for the accountability concept.⁴

1. The new, federally stimulated emphasis on evaluation of school systems and their programs.

2. The growing tendency to look at educational enterprises in terms of cost effectiveness.

3. Increasing concentration on education for the disadvantaged as a priority area of responsibility for the schools.

4. The movement to make schools more directly responsible to their clientele and communities, either by establishing decentralized community control or by introducing consumer choice through a voucher system.

These diverse programs for educational reform, when gathered under the accountability banner, reinforce each other in strengthening the pressures for educational change.


⁴Ibid.
The initial and traditional focus of the public school was upon the academic achievement of the student. Over the years the public's expectations of its schools have broadened to include many additional facets of the child's development. With the emergence of the gestalt approach to education, there devolved upon the schools an ever increasing responsibility for development of attitudes, beliefs, and behaviors. While the responsibility for the academics remains with the schools, today's educators recognize the expanded aims of our educational system.

Jarvis, Gentry, and Stephens, in discussing the effective school district, give emphasis to both the cognitive and affective aspects of achievement. Personal and social adjustment are noted by them as being important areas of school responsibility. They further note, however, the difficulty of assessing the school's performance in the affective domain. Only indirect, or very subjective, measures of effect in these areas are suggested, such as inferring student adjustment by satisfactory holding power. The cognitive domain, they emphasize, can be readily assessed by standardized achievement tests.

Greider, Pierce, and Jordan concur, stating that "the problems of identifying and validating the behavioral outcomes of education are yet
to be stated adequately. Defining these outcomes in terms of student performances which can be measured with some degree of objectivity is an equally difficult unfinished task."

Melby⁷ decries the emphasis on academic achievement and fears that further pressures will lead to national testing and as he puts it "... not an educational system but a scholastic establishment." He further deplores the fact that "... our expertise lies in the materials we teach, not in understanding children."

Kelley,⁸ too, deplores the focus upon achievement. He fears that such comparisons of schools based upon academic results will pit community against community, and school district against school. These types of comparison, he relates "... are the real evils of this form of testing. They put people 'on the spot'." Schools then have no choice but to raise the pressures on children. Parents will be up in arms and angry with teachers. Minority groups, he feels, will be further inflamed.

Kelley urges that more consideration be given to the affective areas. There is a "... need to focus on the need for love." More attention should be paid to whether the child is outgoing or withdrawn to the hostile child, the use of freedom, the self concept, and the attitudes of children. He concludes that "if we can learn to assess


our young on these items, results would be positive for teachers and learners."9

Robinson concurs, and while agreeing to the need for accountability, states that "a major share of the debate hinges on the priorities for the schools. Should the schools be chiefly concerned with humanity or with literacy."10 As he queries, who is to say that for a particular child it is more desirable in a particular year for him to progress a given amount in academic achievement than for him to gain a better self concept?11

The validity of the concern represented by Melby, Kelley, and Robinson are not in question. But Kelley himself pinpoints the problem in his phrasing "if we can learn to assess" these affective outcomes. The public is demanding its accounting in a more definitive terms. As Underwood12 propounds, the public looks to the determination of specific, concrete, and measurable instructional objectives. Any innovation must be based upon a firmly established system of measurement. Plans to spend money must be directed at specific and identifiable deficiencies. It is to eliminate guesswork that ". . . nationally normed, standardized tests are used as often as possible."

9Ibid., p. 114.


11Ibid.

Barro\(^{13}\) agrees that schools aim at affecting many dimensions of pupil performance. In principle, he states, we would like to consider all of them when we assess the school's effectiveness, but it is feasible to work with only a subset of educational outcomes, namely those for which objectives are well defined and which we have some ability to measure. He reiterates that because adequate instruments are available to measure the basic skill areas, these will most likely remain the targets of accountability. For measuring the affective areas he falls back on indicators or proxies such as absenteeism, dropout rate, incidence of vandalism, and the like, which are of questionable acceptability by a public demanding exacting accountability.

Dyer\(^{14}\) in discussing the concept of accountability stresses that the professional staff must know as much as possible about the intellectual and personal-social development of the pupils in its charge, and be held responsible for using this knowledge to maximize the potential of its pupils toward certain clearly defined and agreed upon pupil performance objectives.

He cautions against setting goals which tend to concentrate all efforts on a single area of pupil achievement, namely, the "basic skills." He lists three dangers in this practice. One, it encourages the notion that the school is interested only in training in the basic skills when society is concerned also with other human characteristics.

\(^{13}\)Barro, op. cit., p. 199

Two, he fears emphasis on, and success in one area may be at the expense of others, particularly social behavior. The concentration of effort in limited areas would tend to blind people to the interrelatedness of all educational objectives, that is, that pupil development in one area may be dependent upon development in another.\(^{15}\)

But while Dyer goes to great lengths to promote the need for his School Effectiveness Indices to include all facets of development and all external impingements upon the educational process, every time he has to present or cite an example of measurement he is forced to fall back on one of the "basic skills," reading.\(^{16}\)

Wildavsky continues in the same theme, admitting that many environmental and developmental factors affect academic performance. Yet he is primarily concerned with the measure of end result. The child's reading, he relates, may be a function of personality. It may be a result of motivation: children may not read because they see no reason to do so. Whatever complex causes there may be for lack of academic success, and whatever may be the school's responsibility for dealing with the operant factors, the end result is what is measured.\(^{17}\)

Durstine, in presenting a structure to deal with accountability, does not attempt to discuss specific measures of accountability themselves, but stresses the point that there must be "objective measures of accountability." Measures, he conceptualizes as follows:

\(^{15}\)Ibid., p. 211.

\(^{16}\)Ibid., pp. 209-10.

When something passes from one party to another, there may be a need to measure the amount that passes. . . . There are two reasons for keeping track of such (amounts) (1) so that the amount delivered can be compared with the amount agreed upon, and (2) so that the amount delivered can be compared with that received by other recipients of the same things.

With regards to measures of accountability for a public schools system a variety of parties are involved. They include, among others, children, teachers, parents, unions, schools, and the board of education. To create measures of accountability that work in such a situation requires that both the measures and the framework for dealing with them be very carefully designed.18

These measures, he stresses, must be objective, which he defines as "... explicitly stated and measurable in a way that can be defined independently of the person making the observation."19

Lopez sums up the difficulty of establishing broadbased accountability in education. He states that "the measures of accountability so far developed have not even met minimum standards of reliability and relevance. This failure is known as the 'criterion problem'. . . ." He continues that criteria of effectiveness generally lack clear specifications, and that "objective" measures, when examined closely, are usually found to be either non-objective or irrelevant. Subjective measures when given close scrutiny, are usually found to be biased or unreliable.20

It would therefore seem that despite the consensus that schools should be responsible, and therfore accountable, for many facets of

---


19Ibid.

the child's development, there are to date no generally acceptable measures of development other than in the cognitive skills. Even here there are questions of the appropriateness of today's standardized tests for certain segments of our population.

It is outside the scope of this study to dwell at length upon this discussion of appropriateness. This must be subsumed within the framework of determining appropriate objectives of the school and the desired outcomes of the educational process. Wildavsky, LeSage and Riccio, Lennon, and others describe presently used achievement tests as appropriate to today's current academic educational objectives. There may be inappropriate uses of data, but these should not be confused with invalidity of the data themselves for appropriate applications.

Therefore, for the present it would seem that standardized test data will continue to be the prime measure in accountability. It is the only form of data which attains any measure of objectivity and comparability. The "subjective," "indirect," or "proxy," measures proffered even by those supportive of accounting within the affective domain are too subject to other operant variables to be valid indices of the school's effectiveness. As such they will have little acceptance by the school's public.


Since accountability will be based, in great part, upon school-by-school assessment of student performance, there is implicit an evaluation of the professional staff within each school. In fact, this is a large part of what accountability is all about. As Davies stated:

The concept of accountability calls for a revamping of much of our thinking about the roles of educational personnel and educational institutions at all levels. It links student performance with teacher performance. It implies precise educational goals. It forecasts the measurement of achievement.

It means, in effect, that schools and colleges will be judged on how well they perform, not by what they promise. It means that we are moving in a direction we have been contemplating for a long time—shifting primary learning responsibility from the student to the school. It also means a lot of people are going to be shaken up.24

Grieder feels that it is inevitable and desirable that teachers and administrators give a better account of their professional activities, as well as the funds that are devoted to schooling.25

That these educators reflect the public's thinking about holding school staffs accountable for student achievement is supported by a Gallup poll.

One question asked this national sampling of the lay public was whether school systems should hold teachers and administrators responsible for student progress. Sixty-seven per cent answered in the affirmative, twenty-one per cent said no, while twelve per cent gave no opinion.26

---


Darland concedes this rise in public interest in school performance, going so far as to relate that "the American teacher will be the scapegoat of the 1970's. Evidence can be seen in a drive to hold teachers responsible for quality education in our schools." He further concludes that accountability in this sense "... has all the characteristics of a panacea, and one which it appears difficult to fault."27

While Bain foresees this pressure on professional staffs, she posits that "to make the easy assumption that teachers are primarily responsible for the quality of education today is absolutely naive." She sees the educational system as acting as a damping influence on innovation. If teachers are to be held accountable they must have restraints removed. She concludes that if "... teachers are given the power to control the professional aspects of teaching, even the most cynical critics may be surprised."28

One difficulty exists if one attempts to evaluate teacher performance, even within a single school. Though classes throughout the school may be heterogeneously grouped, it would be rare to find all of them exactly equal in all components which might affect achievement.


White recognizes this fact when she states that to measure teacher effectiveness it would be necessary to equalize carefully all classes at the start of each year. The same would hold true between schools, since one cannot compare schools or classes with different starting points. 29

Even if classes could be "equalized" in terms of levels of achievement, this would not give the full picture. Educational attainment is the product of many factors other than a specific teacher. As Dyer indicates, "surrounding conditions" have much to do with achievement. Home conditions, community conditions, school conditions, etc., all impinge upon academic success. 30

Wildavsky considers the difficulties involved when he approaches accountability within the economic concept of "value added." What we want to know, he states, is what impact exposure to a particular teacher or school has had on a child's reading or mathematical ability, compared to what it would have been under different circumstances. 31

This requires some base line knowledge of where a child is in terms of achievement, and where he should expect to be, given his previous rates of achievement and those of other students similarly situated. The question remains as to how much of the students progress

31Wildavsky, op. cit.
(or lack of it) was due, for example, to a reading teacher or to other teachers. How much is due to school versus home environment is also a complex question. He feels, however, that answers can be found to such dilemmas if the problems of causation are approached in the proper spirit.32

Not all teachers see accountability as a threat, but some rather, as a challenge. For example, the New York City Board of Education and the United Federation of Teachers have agreed to the establishment of the New York City Committee on Accountability pursuant to a clause in the negotiated contract which states:

The Board of Education and the Union recognize that the major problem of our school system is the failure to educate all of our students, and the massive academic retardation which exists, especially among minority group students. The Board and the Union therefore agree to join in an effort, in cooperation with universities, community school boards, and parent organizations, to seek solutions to this major problem and to develop objective criteria of professional accountability.33

Even such acceptance of accountability in the generalized sense runs into difficulty when reduced to specifics. If teacher performance is tied to retention, promotion, or salary it often becomes an entirely different matter.

32Ibid.

There has been controversy over the "merit" system for teachers for many years, and discussion of the pros and cons has no place in this paper except to offer the thought that a realistic plan of teacher evaluation within the framework of accountability might provide the vehicle for merit pay. Wildavsky supports this point of view, stating that sanctions for failure to perform are an integral part of any system of accountability. If those who do badly are allowed to continue, the system will not work. If those who do well are not rewarded, there will be no incentive to continue. He favors accentuating the positive, with the principals and teachers showing the greatest progress receiving recognition, promotion, and freedom.\(^{34}\)

Though many teachers and teacher organizations do not support merit pay, it is interesting to note that, in 1969, the Gallup poll showed that 58 per cent of the lay public favored teachers' pay tied to performance, a merit system.\(^{35}\)

**PERFORMANCE CONTRACTING**

Performance contracting, per se, is nothing new to school systems. It has been used for years in such areas as building construction, plant maintenance, etc., but is only recently entering the picture in educational areas.

Performance contracting is one process for which accountability is the product. Lessinger emphasizes the appeal to both liberals and

\(^{34}\)Wildavsky, *op. cit.*, p. 215.

conservatives, as it gives attention to two things that leaders agree are desperately needed in education—assurance of quality and knowledge of results.36

Basically, performance contracting consists of the educational agency awarding a contract to some other agency or organization to achieve specific goals within specific periods of time for specific costs. This other agency or organization may be public, private, or non-profit as will be seen in the following examples.

The initial and most publicized educational performance contract to date is the Texarkana School District (Arkansas) the the Liberty Eylau School District, Texarkana, Texas. The contract is sponsored under Title VIII of the Elementary and Secondary Education Act.37

A number of private companies, some only recently entered into this phase of the education business, bid on the contacts let by the school districts and the Office of Educational Opportunity (OEO). Some of these were McGraw Hill, Radio Corporation of America (RCA), Quality Educational Development, Inc. (AED). The winner of the contract was Dorsett Educational Systems, Inc., of Norman, Oklahoma.38


The emphasis of the Texarkana Project was on dropout prevention. The first requirement was for students to be below national average in the basic skills - reading and mathematics. Two years or more below appropriate grade level was the original criterion for admission to the program, though some "volunteer" candidates were accepted, who, while below level, did not meet this two year criterion.39

The contract called for these students to gain one year grade equivalent in reading and arithmetic achievement in one school year, before any payment would be made to the contractor. Increased gains would result in increased payments.

The attainment of even one year, (average) progress could be considered as real success, as these students, in their past school history, had gained only between .33 to .5 years of grade equivalency in reading and arithmetic during each school year.40

There were several innovations in the Texarkana program. Teachers were hired by Dorsett from outside the regular school systems. These were certificated but unemployed teachers, usually from within the community. The pay was at the local salary scale except for bonuses for longer days. A mixture of hard and software was used in the Rapid Learning Centers.41 Another variation from the regular school was the

39Ibid., p. 510.


41Elam, op. cit., p. 512.
"livingroom" atmosphere. Carpets, casual furniture, and small groups contributed to what Dorsett referred to as a "positive, middle class atmosphere."  

First results seemed almost miraculous. Randomly selected students were tested at varying intervals and such results were reported as:

1. Eighty students gained one grade level in eighty hours of instruction.\(^43\)

2. Fifty-five students gained 2.2 grade levels in reading and 1.4 grade levels in math in sixty hours of instruction.\(^44\)

3. Average gain by selected students showed more than two grade level gains in reading and more than one grade level in math after forty-eight hours of instruction.\(^45\)

But soon some of the glowing reports were offset by creeping doubts. Despite the fact that testing was conducted by the school system, not by the contractor, the evaluator concluded in his final report that from 30 per cent to 100 per cent of the questions in the May, 1970 tests were "contaminated" because the questions had been taught in the classroom prior to the tests. Dorsett countered with

\(^{42}\)Dorsett, op. cit.

\(^{43}\)Filogama, op. cit., p. 53

\(^{44}\)Ibid.

an estimation of only 6.5 per cent contamination. The project, funded by OEO for five years, was renegotiated and eight companies, including Dorsett, bid on the second phase.  

The Texarkana Board of Education decided not to renew the Dorsett contract, charging that the company was "teaching to the test." Instead the contract was awarded to Educational Development Laboratories, a division of McGraw Hill.

This experience points up two limitations which need to be considered within the province of performance contracting.

1. As Wildavsky warns, testing must be done by one other than the unit accountable. Tests must not be available to the teachers prior to testing and an observer from the testing organization must be present.

2. Before payments are agreed upon there should be a way to establish if "true" gains are made. The Texarkana proposal originally contained a penalty clause to guard against temporary achievement spurts or "Hawthorne" gains which would not remain. If


Initial gains would disappear after six months an adjustment would be made. This clause was not included in the final contract.\textsuperscript{49}

The controversy over the value of performance contracting may be settled by a study to be made, by the Rand Corporation, of all existing and proposed projects. The sixteen month, $300,000 study financed by a Health, Education and Welfare grant will produce a guide to school districts considering the use of performance contracting.\textsuperscript{50}

In the meantime the performance contract marches on. More than 250 school districts have written the Office of Economic Opportunity asking how they, too, might get on the band wagon.\textsuperscript{51}

Philadelphia has contracted with Behavioral Research Laboratories to improve the reading ability of 15,000 children at least one year in grade equivalency, for $600,000 or $40 per successful student. The firm is to provide all materials, teachers, paraprofessionals, and interpret the program to the community.\textsuperscript{52}

Gary, Indiana is turning over an entire elementary school to the same company on a four year contract, in this instance to bring students up to or above national norms in reading and arithmetic. Fee for success is to be $800 per pupil—the district’s present per pupil expenditure.\textsuperscript{53}

\textsuperscript{49}Elam, \textit{op. cit.}, p. 509.
\textsuperscript{50}"Performance," \textit{op. cit.}, p. 3.
\textsuperscript{51}"Startling," \textit{op. cit.}, p. 123.
\textsuperscript{52}"Performance," \textit{op. cit.}, p. 3.
\textsuperscript{53}ibid.
State education officials in Virginia and seven of its local districts are planning to use contracting to increase the reading level of some 2,500 disadvantaged students. The contractors, still unselected, would receive $85 per pupil. A Dallas program will concentrate on 960 students, in five high schools, who rank in the bottom quarter of their class. These projects will use Title I funds. The education contract goes to a New York firm, New Centuries, while the motivational and occupational aspects will be handled by Thiokol Chemical Corporation.54

Learning Consultants, of Medina, Ohio, is to improve the reading of 2,160 ninth and tenth graders in Flint, Michigan for $310,000, while Model Cities funds in Providence, R.I., will be used to teach reading in four public and two private schools. The contractor is as yet unnamed.55

Performance contracting has caused much controversy among educators. The American Federation of Teachers launched a national campaign against what it called "educational gimmickry" such as performance contracting. The charge is that industry thinks that it knows more about educating children than teachers and parents do.56

John M. Lumley, legislative head of NEA, told a senate subcommittee

54 Ibid.

55 Ibid.

that performance contracting programs tend to weaken the structure of public schools and discredit them in the eyes of the public.57

But new "firsts" have been achieved. Teachers are signing performance contracts. In two districts (Mesa, Arizona and Stockton, California) the Office of Economic Opportunity signed contracts with the local NEA affiliates. Twenty-four Mesa teachers and twenty-two Stockton teachers will work with 1,200 low-income students in grades 1-3 and 7-9 who are performing below grade level.

OEO is injecting an "incentive" into these plans. Teachers will receive salary bonuses of 5 per cent to 6 per cent of their base salary if all the students raise their skills by two grade levels in either reading or mathematics. If the goal is achieved in both, the bonus could be 10 per cent to 12 per cent.58

The salary bonus may be used by the teacher for materials, prizes, games, etc., or kept by the teacher. A teacher whose students fail to increase their performance by at least one grade level will receive his salary -- but no bonus.59 It must be remembered that these students, typically have not gained one grade level during a school year.


59Ibid.
In another "first", both teachers and administrators stand to earn a $500 salary bonus in a new type of performance contract that will be tested in the Dallas Independent School District. The $75,000 contract was proposed by the Dallas School Administrators Association at the invitation of the Superintendent. The offer was originally made to the Dallas Classroom Teachers Association, but was rejected by them because of its merit pay aspects.60

The proposal calls for setting up a developmental reading laboratory in one inner city school. In a similar school another laboratory using the same "up-to-date, motivational instructional methods" will be installed, but without incentive pay. A third traditional school will serve as a control. The degree of success will be measured by tests devised and controlled by Educational Testing Service. Results will determine the amount of the bonus.61

The Dallas school board originally approved the proposal without the pay incentives in deference to the Classroom Teachers Association's opposition, but the incentive factor was reinstated after talks with the specific teachers involved.62

The concept of performance contracting is gaining in popularity. A recent study by the American School Boards Association determined that two thirds of the members polled found at least some favor in the concept.

---

61Ibid.
62Ibid.
of performance contracting. Webb theorizes that board members are more directly responsive to public opinion than are educators, and that:

An aroused and incredibly diverse public is making complex vocal demands on its schools and insisting on a measurable accounting from the people it selected to be immediately answerable for what the schools accomplish or fail to accomplish.63

He goes on to recount that the nation's public school children number forty-five million, with some thirty-five billion in tax dollars being spent annually on education with "virtually no measurable results." To the question as to why board members favor performance contracting, he offers the postulate that they feel teachers have turned from a commitment to children as the primary responsibility to a commitment to their own occupational interest.64

He summarizes that performance contracting may be seen as a tool for "boardsmanship" in insisting upon teacher performance "or else."65

VOUCHER SYSTEMS

Another alternative to the "or else" is the voucher system. Termed probably the most radical proposal for achieving better education through accountability, the voucher concept would allow competing publicly financed schools to coexist and would permit parents to choose schools for

64 Ibid. p. 29.
65 Ibid. p. 30.
their children. Usually the plan would be coupled with a proposal for financing by means of "educational vouchers," though this is not the only possible mechanism.66

The rationale behind the "consumer-choice" solution is that there would be direct accountability by the school to the parent. There is built in an automatic enforcement mechanism: a dissatisfied parent could move his child, and therefore the funds, to another school.

This market approach is based upon a plan suggested by Professor Milton Friedman of the University of Chicago. He claims that schools are essentially monopolistic in that they provide services for a captive audience. Since most parents and their children have little choice but to attend local schools, no matter how poor their performance, the students are forced into a system ill suited to their needs. The proponents of the plan feel that if parents were given free choice over alternative schools, and schools were required to compete for students, substantial increases in educational effectiveness would result.67

While this voucher system is a radical departure from the current public education system, it is not altogether new. Plans to give money

66Barro, op. cit., p. 197.

to parents and allow them to pay for their children's schooling go back into the early days of our country's history and have been proposed and discussed since then.68

There are a number of implications for educational institutions and accountability implicit in this revived exploration into the voucher concept.

1. What will be the effect on parochial schools, and what are the implication of public funds being provided for their support through vouchers?
2. What implications are there for private schools, both with respect to their accreditation and to their use to circumvent integration?
3. Upon whom will devolve the responsibility to judge the quality and effectiveness of these alternative schools?
4. What effect will diverting support to non-public institutions have upon the public schools?
5. What are the implications for public schools in terms of the neighborhood school concept?
6. What are the implications for accountability and the comparison of school's performance in terms of student achievement?

One fear among public school educators is that public funds will go to support church regulated schools and the teaching of religion.

Though there might be some philosophical or even legal implications in this "separation of church and state" controversy, there are two proposals which might have some ameliorative effect. These proposals were the result of a study made by the center for the Study of Public Policy and accepted by the Office of Economic Opportunity.

1. The proposed plan suggests that 50 per cent or more of the seats in a voucher school be filled "at random" or by lottery. Thus a church related school must accept a fair share of all applicants, regardless of their church affiliation or non-affiliation.

2. The voucher could be specified to apply only to secular education, and could be redeemed at only, say, 80 per cent of tuition by church related schools.\(^{69}\)

It is felt that these reductions in funds plus the loss of controls over a significant portion of their enrollment, many administrators of church related schools will not be overly anxious to participate in the voucher support system.

With respect to private schools, two fears are most prevalent. One is relative to the "get-rich-quick" operator who could expect certain parents to pay a bonus to assure their child's acceptance. The other involves the possibility of the school's being used to circumvent integration. The first fear could be avoided by requiring that the

\(^{69}\text{Ibid.}\)
voucher be accepted for full tuition. Participating schools would also be required to be open for audit by the voucher agency. The quota system mentioned previously would insure that a goodly portion of the student body be accepted on a random basis, thus preventing planned exclusions. The proposal goes so far as to suggest that the final enrollment must at least "mirror" the ethnic makeup of the applicant body.  

The voucher plan might prove a saving force for private schools, now as a whole facing financial difficulties. During the 1969-70 school year, the National Association of Independent Schools reported that 53 per cent of its numbers finished in the red. These schools operate largely from tuition increases.

Yet despite these difficulties, Robinson reports an increase of over 700 private schools in the past three years. Each day, he goes on, sees two or three new schools enter the field, though about one each day fails. Over one thousand private schools exist in California alone.

The National Association of State Boards of education has passed a resolution opposing the use of public funds for non-public schools.

---

70 Ibid.


The members voted in favor of banning aid to non-public schools "... under any circumstances which would jeopardize the welfare, stability or support of the system of public education." This wording left it up to the individual states what, in their opinion, would constitute jeopardy of public education.73

The NASBE, though, is split on the issue. Proponents of aid to non-public schools cite recent Supreme Court decisions and the 1965 Elementary and Secondary Education Act, which allows aid for specific purposes to non-public school children. Opponents, the stronger element, argue that public schools would be weakened and that segregation academies would be encouraged.74

On the other hand, Bernard J. Donovan, former superintendent of the New York City Schools, feels that vouchers could lead to the "... blossoming of the public schools, not their doom."75

The 1969 Gallup poll asked the public their opinion on an educational voucher system. Of those expressing an opinion the split was fairly even, 46 per cent against and 43 per cent favoring such a plan. Those whose children attended public schools were against vouchers 49 per cent to 41 per cent, while parochial school patrons favored vouchers

74 Ibid.
48 per cent to 40 per cent. With as little publicity as voucher plans have been given in the general media, these poll results are quite revealing.76

Another poll question concerned whether public and private schools should be allowed to coexist. Here 72 per cent responded that new communities should provide public, private, and parochial schools. Where areas already have all three, the favorable reply jumped to 84 per cent. When asked to rate the quality of education offered by the different types of institutions, 32 per cent favored public schools, 24 per cent picked private schools, 21 per cent chose parochial schools, and 20 per cent ranked them equal in quality.77

The voucher concept espoused by the Office of Economic Opportunity forsees the establishment of an Educational Voucher Agency (EVA) at the local level. This Agency would resemble a board of education, though its exact composition is not spelled out. This group would receive all federal, state, and local monies designated for education. It, however, would not actually have a hand in operating schools.78

These monies would be distributed in voucher form to parents who would in turn select from among existing private, public, or parochial

76Gallup, op. cit., p. 102.


schools, or from new schools which might be formed. This would, according to Jencks, eliminate the monopoly of public schools.\textsuperscript{79}

The regulations prohibiting extra tuition and the restriction on student exclusion mentioned above are reinforced. There would be more flexibility with regard to staffing and curriculum, though the EVA would oversee the various schools meeting eligibility requirements before they could become voucher recipients. Jencks indicates further that the EVA would have the responsibility of collecting and disseminating information to the community upon which parents might base their evaluation and choice of institution.\textsuperscript{80} What specific information would be obtained and released was not spelled out, but, if the present demands for accountability were extended, this could mean the release of school-by-school achievement scores, plus such other data as would assist parents in comparing schools' performance.

In the meantime, a Maryland plan in the making calls for a voucher worth from fifty to two hundred dollars to aid parents wishing their children to attend non-public schools.\textsuperscript{81} Preplanning grants have been awarded by the Office of Economic Opportunity to four school districts.


\textsuperscript{79}\textit{Ibid.} p. 50.

\textsuperscript{80}\textit{Ibid.} p. 51.
They are in Alum Rock, California; Gary, Indiana; San Diego, California; and Seattle, Washington. 82

There is another suggested modification which would eliminate the non-public school controversy yet retain the competitive element. Douns has suggested that cities merge several existing attendance areas so that all students within a given section of the city can attend any schools within that boundary. This modification is basically a free enrollment plan, but with the difference that schools would compete for students, and teachers and resources would be shifted as enrollment in less successful schools dwindled. Principals and staff would have an incentive to maximize the important educational outputs desired by residents of the merged attendance areas, or face a loss of clientele and resources. 83 Many present plans place limitations on pupil movement to pupils "... with special interests or talent... (or) difficulties..." 84 Others place limitations to protect the existence, at an efficient level of all schools.

The caution is that only a school system whose board and superintendent have a great confidence in their staff's ability to meet the challenge of competition, and to adapt to meet it, will be willing to undertake such an experiment. 85

83 Levin, op. cit. p. 81
84 Swanker, op. cit., p. 244.
85 Ibid.
DECENTRALIZATION

Decentralization is another manifestation of the search for accountability. Several of our large cities have had some degree of decentralization for a number of years, but today the trend is toward the proliferation of the concept and toward an even greater degree of abdication of control by central administration.

The National School Boards Association recognizes the need for effective community involvement in school operations. Involvement at a meaningful level is difficult in the large metropolitan areas with their many and diverse publics when total control is centrally focused. A special decentralization committee was instituted by the NSBA's Council of Big City Boards of Education to look into the challenge of community involvement. 86

The committee came up with three ground rules for community involvement.

1. The only basic purpose of schools is effective education of children.

2. The community and parents must always have the right to review this process.

3. The method for this review must depend more on local conditions than on any formula. 87


87 Ibid.
The committee continued, stating that community involvement did not necessarily mean that a school district must be broken up into separate entities nevertheless, varying degrees of subdivision are embodied in virtually all plans yet devised. Barro agrees, concluding that while decentralization and community involvement are conceptually distinct, they are usually lumped under one heading because of their close linkage in recent years. 88

Administrative decentralization, in which decision making authority is shifted from central administrators to local area administrators or to school principals, can contribute greatly to accountability. This shift of control greatly favors professional responsiveness to local conditions, and local initiative is fostered. When control is shifted to local units, these units can then be held more accountable by their clientele and the local community. 89

Featherstone and Hill maintain that central offices cannot meet the needs of varied populations and their myriad educational needs. It is impossible for a uniform educational program to serve the best interests of these varied publics. They caution, however, that some central control will continue to be needed to maintain overall balance. 90

88Barro, op. cit., p. 197.

89Ibid.

New York City was the first major U. S. city to move for decentralization in the full sense. Other cities have made modified adjustments within the broad concept of decentralization, but the New York model is the most complete.

As the result of dissatisfaction expressed by some segments of the public upon the 1964 release of school-by-school achievement data, New York organized three "Demonstration Districts," including the controversial Ocean Hill-Brownsville district. In 1969 the New York State Legislature finally agreed upon a decentralization plan which would divide New York's 1.1 million pupil system into thirty to thirty-three partly autonomous districts. The law did not give as much power to local districts as community control advocates had hoped for. Key power was to remain with a seven man school board composed of an elected representative of each of the city's five boroughs, and two members appointed by the mayor. The districts were to have their own elected boards.91

The three experimental districts were broken up and their areas redistributed as appropriate, within the newly formed districts. It was to be up to the new community districts to decide whether the experimental educational programs started by the demonstration districts were to be continued.92

---

The New York plan entered the 1970-71 school year without incident. The final number of local districts was thirty-one. Each district was headed by a locally elected board and a locally selected superintendent.93

Other large systems have also moved, or are moving, toward some degree of decentralization, but none yet as extensive as the New York model.

Detroit, with its two hundred and ninety thousand students is being divided into seven to eleven districts with between twenty-five and fifty thousand students.94

Los Angeles schools have been divided into four zones which will be directed by district superintendents.95

Washington, D.C. also has moved toward decentralization having established two pilot districts; the Morgan Community School District and the Anacostia Community Demonstration District. Each has its own local board of education and policy board.96

Chicago, too, has a four school demonstration district in being, designated as the Woodlawn Project. Its novel feature is a tri-partite

board representing the Woodlawn Community, the Chicago School Administration, and the University of Chicago. 97

Portland, Oregon has a plan-in-making for decentralization. In announcing the plan to the Portland Board, superintendent Robert W. Blanchard stated:

I am recommending the decentralization of the system into four pre-school through grade 12 administrative areas. . . . The need is imperative to provide a better response to community and staff needs and concerns. . . . and to allow more direct and productive involvement of the community and staff in school affairs. 98

Blanchard continued with the projection that each area have citizen representation in the form of an advisory school board. 99

With the increasing trend toward decentralization Brownell makes several recommendations for school systems considering such a move. 100

1. Population within the district should be diverse, providing for integration when and where possible.
2. Parent organizations and citizen committees should be given a voice and opportunity to influence school operations.
3. Community control should extend to each school,

99 Ibid.
with local responsibility and authority given within established policy.

His concept places the class as the basic unit of community concern followed by the school, then a constellation of schools. The district he envisions is a constellation of two or more high schools and their feeders. These districts would be operated by a locally elected board which would have all powers except those reserved to the city-wide board. 101

The city-wide board would basically have the following responsibilities: 102

1. Determine board policy and do city-wide planning
2. Be responsible for overall evaluation.
3. Determine long range needs and plant planning.
4. Seek resources (local, state, federal).
5. Allocate funds on formula.
6. Stimulate intra and inter district research.
7. Represent and coordinate city-wide interests with city, state and federal governments.
8. Hear appeals on district complaints.
9. Project competence as the basis for appointment, promotion, and tenure.

101 Ibid.
102 Ibid.
Some raise questions as to whether decentralization will accomplish its avowed purpose. Bushkin decides that community involvement could be the answer to many problems of ghetto schools, but warns that it is necessary to create a structure for giving parents a real voice in their schools. Too many plans, he relates, give only "lip service" to community involvement. The parents, particularly in black communities, want "... authority to make real educational change."\(^{103}\)

Despite the reluctance of teacher associations and Unions, Bushkin feels that subdistricts should have the authority not only to determine budget and curriculum, but to hire and fire staff.\(^{104}\) This position is similar to that proposed by Brownell, except that his eighth and ninth board reserved points would provide the teacher with appeal from unfair and arbitrary practices.\(^{105}\)

Featherstone and Hill caution against decentralization by edict or protest. They feel that time must be allowed to train staff and develop competency in the community. Concerning the inner cities of our urban complexes, they warn: "In the heart of our cities as in the heart of underdeveloped countries, the path to enlightened self government may be long and difficult."\(^{106}\)


\(^{104}\) Ibid., p. 56.

\(^{105}\) Brownell, op. cit.

The effectiveness of decentralization has yet to be fully evaluated, especially on the scale and to the degree of the New York model, but as a Harlem board member, David Spencer stated, "We are just trying to show that there is another way. . . . Because the way that has been shown to us all these years ain't been working!"

**TESTING FOR ACCOUNTABILITY**

If, as would seem to be the case, accountability is to be based largely upon student performance as measured by achievement tests, what type of tests should these be? Reference is made to criterion-referenced tests versus norm-referenced tests. Questions arise as to the appropriateness of various derived scores for reporting achievement. Statistical concepts come under scrutiny when achievement gains are made the basis of evaluating a school's effectiveness or a contractor's payment.

Most, if not all standardized achievement tests currently used by school systems to evaluate performance are of the norm-referenced variety. Popham and Husek define norm-referenced measures as:

> . . . those which are used to ascertain an individual's performance in relationship to the performance of other individuals on the same measuring device. The meaningfulness of the individual score emerges from the comparison. It is because the individual is compared with some normative group that such measures are described

as norm-referenced. Most standardized tests of achievement can be classified as norm-referenced.\textsuperscript{108}

Publishers of the commonly used standardized achievement tests attempt to have their normative samples represent a cross section of the universe for which their instrument is designed. The sample should have approximately proportionate representation of geographic areas, urban/rural population, ethnic groups, public, private, and parochial schools, etc. Cronbach\textsuperscript{109} cautions against tests without adequate or representative samples, but with the recent introduction of computer technology, the sampling techniques have greatly improved. The national sample cross section would never be exactly representative of any specific school system, though it can serve as a reference point. For certain uses and comparisons, other normative data provided by some publishers might prove useful.

CTB/McGraw Hill, for example, provide for their Comprehensive Tests of Basic Skills, not only the normative data on the total national sample, but also provide data based upon major geographical sections, and upon the type of system -- rural, suburban, small town, etc. Of particular interest might be their table of Large City Norms. This is based upon samples drawn from school districts with student populations in excess


of 95,000. The sample cities are all ones with significant "core-city" populations. 110

There has been much discussion as to the appropriateness of standardized tests for measuring the minority populations within large cities. Cities such as New York, Los Angeles, and others have been forced to eliminate group intelligence testing at lower grades. 111, 112 Arguments have blossomed, pro and con, over the appropriateness of these actions. 113, 114 Many studies have been conducted to determine the effect of some use (or misuses) of "intelligence" test data. 115, 116, 117

110 Large City Norms, CTBS, Monterey, California, 1969.


But while intelligence, or more properly, academic aptitude tests are held to be useful devices for assessing the abilities of students particularly in the early years in school, the focus of this study is principally upon achievement tests.

The better designed achievements tests on the market today are likewise valid measures of currently defined educational objectives.

Cronbach discusses content interpretation in light of a universe of tasks defined as the instructional objectives. The validity of content, he relates, is judged entirely apart from all the persons to be tested. From an absolute point of view, the score on a task indicates that the individual does or does not possess all the abilities required to perform successfully. As this applies to various segments of the population, low scores may reflect any of a number of impingements upon their preparedness for the task, including the many disadvantages often cited, such as cultural differences, home environment, poverty, etc.

As long as our society puts a premium on the skills and knowledges being tested, the test validly serves to indicate the student's relative standing in relation to the possession of those skills and knowledges.

---

118 Goslin, op. cit., p. 132.

The burden is then placed on the school to intervene in any areas necessary or possible to effect the desired achievement outcome.\footnote{120}{Thomas J. Fitzgibbon, \textit{Content, Constructs, Criteria, and Confusion}, (New York: Harcourt, Brace, Jovanovich, Inc., 1970), p. 5.}

Norm-referenced achievement tests can be valuable in establishing benchmarks at periodic evaluations of performance as children proceed through their schooling. Where they may prove less valuable is in measuring smaller increments of performance over comparatively short periods of time. Tyler\footnote{121}{Ralph W. Tyler, "Testing for Accountability," \textit{Nations Schools}, Vol. LXXXVIII (December, 1970), p. 37.} is concerned about the emphasis accountability will put upon short time measures. Performance contracting is a current example of this emphasis, and he states that the norm-referenced tests currently available are not suited to the task.

Tyler's concern is valid. The average achievement test is constructed so that the final version contains items that from forty to sixty per cent of the children were able to answer. There are very few questions that represent things that are learned by either slow learners or by more advanced students. This is not a weakness of the test for the purpose for which it was designed, but does make it inappropriate for certain uses, including measurement of small increments of performance.\footnote{122}{Ibid.}

Another limitation of the present norm-referenced standardized tests in measuring limited elements of achievement in the comparatively small...
number of items targeted at the objective areas within each test. The Comprehensive Tests of Basic Skills, published by CTB/McGraw Hill, is typical of the field.

The test of reading comprehension purportedly measures eight objective areas at the upper elementary level. Test items within these objective areas vary in number from five to eight. A total of forty-five reading comprehension items is used to evaluate and rank students in grades four, five and six. Of necessity, items must be included to cover the difficulty ranges over the three year span. Such a test can, with reasonable reliability, measure achievement differences over periods exceeding a year, but are of somewhat less value as year or less measures.123

In this same CTBS reading test, the standard error of measurement (SEm) is almost four raw score points above and below the attained score. Even at the middle range of the test's target grades this band covers almost one year of grade equivalency. With such a gross measure it would be difficult to determine if a measured difference, after only a brief period of instruction, was the product of that instruction or a chance variation.124

Lennon125 comments upon the same problem in measuring the differences


in performance, or gain scores, as they relate to performance contracting. Even with tests having a very satisfactory reliability, for example .90, the reliability of gain measured over relatively brief periods is distressingly low. The error of measurement of the gain score may very easily equal or exceed the amount of gain normally to be achieved on the short-term intervention.

Hogan\textsuperscript{126} agrees, citing an example measured gain in a reading program. The tests used as pre and post measures each had a reliability coefficient of .90, and the correlation between pre-test and post-test scores was .75, however, the resulting reliability of the difference score was .60. His conclusion is that, though quite reliable tests were used, the resulting measure of academic gain based upon those tests was relatively unreliable.

Durost\textsuperscript{127} cautions against using norm-referenced tests as before and after measures of gain over relatively short period of time. He, too, concurs that such use "... requires the tests to do something for which they were not constructed."

Lennon\textsuperscript{128} stated that the unreliability of individual gain scores posed severe problems within performance contracts. The instinctive


reactions, he stated, was to let compensation be based upon average gains for the group. His warning was that this might be taken as a "cop-out," and let the contractor avoid responsibility for poor students. The contractor, he feared, would permit failure by a significant fraction of the group, hoping for this to be offset by good gains from those more amenable to instruction.

Another factor limiting the usefulness of norm-referenced tests for measuring the effectiveness of short periods of instruction is the differing instructional sequence in various schools. The usual standardized test is designed for use in schools throughout the nation and is based upon a composite of instructional patterns used with varying texts and curricula. As such, they may be more, or less, appropriate for a system using a certain text or a particular sequence of instruction.  

It has therefore been recommended that criterion-referenced tests be utilized for these year-or-less measures. Popham and Husek define criterion-referenced measures as:

... those which are used to ascertain an individual's status with respect to some criterion, i.e., performance standard. It is because the individual is compared to some established criterion, rather than other individuals, that these measures are described as criterion-referenced. The meaningfulness of an individual score is not dependent on comparison to others.  

---

129 Tyler, op. cit., p. 38.

130 Popham and Husek, op. cit.
Programmed learning is an excellent example of the criterion measure approach. Whether a student is using a linear or branching program, the pace is determined by the time taken by the student to adequately assimilate the material. Only when the criterion of mastery is reached is the student permitted to proceed.131

Tyler suggests a reading test, for example, for primary children who enter without having learned to distinguish letters and sounds. At the end of the year he suggests that they be tested on letter recognition, association of letters with sound, and word-recognition of one hundred most common words. A large number of items would be presented to measure the first named objectives, which together with the one hundred words, would adequately evaluate the year's progress. A mastery or proficiency cutoff score would provide the criterion for satisfactory performance.132

Tyler's suggested one hundred word minimum for beginning readers is a far cry from the thirty or forty word samples used to measure vocabulary knowledge, even at high school levels, on most standardized reading tests.

The Office of Economic Opportunity is again injecting the stimulus for overcoming the problems of developing effective criterion-referenced


132 Tyler, op. cit., p. 39.
instruments. A $6.5 million dollar performance contracting experiment will include criterion-referenced testing as a portion of the criteria for payment. Testing is to be handled by Battelle Memorial Institute of Columbus, Ohio.133

The first task in the development of these tests was the agreement upon curriculum objectives. Once objectives are determined, test items are constructed to assess the level of mastery. The education companies are working closely with Battelle to develop specific tests to measure each phase of each individual program being developed within the experiment. O.E.O. expects over five hundred criterion referenced tests to be needed for the six week periodic testing in all of the many reading and math programs within the several projects.134

Millman135 admits the difficulties in developing criterion-referenced instruments, particularly the problem of defining the universe of tasks and the proficiency standards which must be met.

Criterion-referenced instruments may well have their place in accountability, but not as a total replacement of norm-referenced tests. There are several reasons why this is true.

1. As indicated in the O.E.O. experiment, construction


134Ibid.

of criterion referenced tests is a complex task and must be adjusted to the specific curriculum of the program. The need for many of the different tests when several school systems are working in common math and reading areas is an indication.  

2. If test publishers develop a series of criterion referenced tests for, by way of example, fifth grade mathematics, then there must be the complex task of adjusting the performance cutoff to the one system's own curriculum. The other alternative is to adjust the curriculum to the test objectives. This latter could lead to a national curriculum.

3. Since the criterion-referenced test must be more comprehensive, as Tyler suggests, the time for administration of a test measuring objectives covered over a long period of time, such as an entire school year, would be unreasonable.

It is likely that when accountability demands frequent measures of performance, criterion referenced instruments will be used. Although few criterion referenced tests are presently used, there are indications

136 Ibid.

137 Tyler, op. cit., p. 39.
that major test publishers are exploring the market. Tyler envisions a crash program criterion-referenced test development, and the O.E.O. is providing the "seed money". Though criterion-referenced testing will be used, as defined also by O.E.O. as "interim performance objective tests," there will be reliance on the more traditional norm-referenced tests used to measure overall gain over longer periods of time.

Another reason that the norm-referenced test will likely retain a prominent place in accountability is that there will be continued public demand to know who the schools are performing in comparison to "national standards," or how their schools compare to schools in other cities. Unless national curricula are established and nationally used criterion-referenced tests are developed to evaluate performance within these curricula, the reliance for comparability must be upon the norm-referenced instruments.

Another problem which faces the accountable school, and even more particularly the school system considering performance contracting, is defining the normal growth to be expected of children. The most widely used means used to express academic growth is the grade equivalent score.

---

140 Ibid.
Durost\textsuperscript{141} questions if grade equivalents are ever the proper statistical unit for indicating change from beginning to the end of a training period. He cites as a deficiency of these grade equivalent (G.E.) scores that they are usually interpolations or extrapolations of scores derived at one, or at best two, administrations during a school year. They do not take into account the lack of uniformity in achievement throughout the year, or the effect of the summer vacation. He concludes that "... grade equivalents, for the enormous range of scores above or below the successive averages, must be obtained by some artificial means."

A further deficiency of grade equivalent scores becomes evident when one attempts to define "normal growth." Downie\textsuperscript{142} cautions that grade placement norms are not "standards." They represent the average performance of the norm group, yet performance contracts typically call for children to gain a year of grade equivalency in a particular period, or to be at appropriate grade equivalent by some specific time. These expectations, for each child to gain one year of equivalency for each year of schooling, are not uncommon, even among teachers, according to Downie.


Hogan\textsuperscript{143} deplores the use of the grade equivalent concept in establishing a rate of "normal growth." If each child is expected to attain a one year gain in grade equivalency for each school year of instruction, this is to set the \textit{average} as the norm for all. Less able students would be faced with unrealistic demands, while more able students would be inadequately challenged.

His suggested approach would be that each student should be expected to maintain his relative rank in comparison to the normative group in order to evidence "normal" growth. A student who, at one grade, ranked at the tenth percentile on the test's norm tables should maintain this same relative percentile rank in succeeding grade levels. The same should hold true of students at varying points within the percentile ranks, with those students near the fiftieth percentile being the only ones to be "at grade level" and gain one year of grade equivalency year to year.\textsuperscript{144}

A note of concern has been expressed by some in discussing the use of tests in accountability. This concern has to do with the statistical treatment of the data, not so much as to its inappropriateness, but because of the frequent overlooking of common phenomenon.

\textsuperscript{143}Hogan, \textit{op. cit.}

\textsuperscript{144}Ibid.
One phenomenon mentioned earlier is the standard error of measurement (SEM). Womer\textsuperscript{145} describes the SEM as being of less consequence in the interpretation of group test results, such as in comparing mean scores between groups, than in interpreting individual scores. In group data, errors tend to cancel out. Some performance contracts make payment to the contractor dependent upon individual student gains — but do not provide penalties for individual student losses. In these instances the contractor could profit from the chance variation, and has the SEM operating entirely in his favor.\textsuperscript{146}

Another statistical phenomenon with implications for accountability is that of the regression effect. Generally speaking, when scores are either very high or very low on the initial test, they can be expected to regress toward the mean on the second testing. High scores on the first test tend to be somewhat inflated by test errors, and low scores tend to be somewhat depressed. Thus, if a person who initially received a very low score on a test were to take a parallel form of the test immediately afterward, there is a good possibility that the score would go up. Similarly, a person who scored very high on the first test would probably score lower the second time.\textsuperscript{147}


\textsuperscript{146} Supra, pp. 33-34.

If a given student population is subjected to both the initial test and the retest, the mean should remain relatively stable, as the regression effect on the upper and lower segments would tend to cancel out. However most Federal programs and performance contracts tend to deal with those students at the lower end of the spectrum. The shifts in score due to regression would therefore be generally upward. A contractor paid for measured gain would stand to profit not only from the results of his instruction but also from the statistical artifact.\footnote{148}

Hogan\footnote{149} suggests a possible solution. Instead of using the same test both to screen students into the program and as a pre-test, first select students with one test, then use a second instrument as the pre-test. The regression would operate between these two tests, but would not contaminate the later post-test, which would be a third instrument. Any regression between the pre and post test would involve the entire instruction group and would tend to be self-correcting as to the mean gains.

A final concern related to testing is the fit between the test and the objectives of the educational program. Cronbach\footnote{150} relates content validity to the universe of tasks defined as instructional objectives.

\footnotetext{148}{Lennon, \textit{op. cit.}, "Accountability".}

\footnotetext{149}{Hogan, \textit{op. cit.}}

\footnotetext{150}{Cronbach, \textit{op. cit.}}
But what are to be the objectives? This is the first decision which must be made, but by whom?

Traditionally it has been the professional educator who stated the school's objectives. In theory he selected objectives which reflected the needs and desires of the community, but, particularly in an urban complex with its pluralistic society, this is a difficult task. Merriman\footnote{Howard O. Merriman, "Accountability-Willingness and Responsiveness," Paper presented at the Wisconsin Association of School Boards' Annual Convention, Milwaukee, Wisconsin, January 27, 1971, p. 9.} states that it is essential to consider the entire educational community, which he delineates as all those persons in a community who have an interest in education—parents, teachers, students, administrators, and often, just interested citizens. Each of these categories he emphasizes, represents a set of educational decision makers.

Klein\footnote{Stephen Klein, "Instructional Objectives," Address at Invitational Measurement Conference, Memphis, Tennessee, April 26, 1971.} agrees that individuals in these categories must be involved, but feels that some framework must be established by the educator. The stating of objectives in \textit{measurable behavioral} form is essential if the school is to be held accountable for the attainment of the objectives. His suggestion is that educators develop a comprehensive universe of objectives for which the school could become responsible. The various categories within the educational community as
outlined above could then each rank these objectives in priority order. While unanimous agreement would probably never be reached as to priority, at least there would be a point of departure for negotiation, and the objectives would be in a form amenable to program implementation and evaluation.

The objectives included in Klein's "Q Sort" would be comprehensive both within the cognitive and affective domain. If objectives are stated in measurable behavioral terms, then all concerned would be aware of how the attainment of the objectives would be evaluated, and any limitations that would exist as to objectivity, reliability, etc.153

Lennon154 comments upon the matter of even nominally valid tests being more or less inappropriate for a particular program. He cites the example of the many "reading" tests on the market. All may measure some facet of reading quite correctly, yet they vary greatly with respect to subtest composition and relative emphasis on component skills. All would be defensible on rational grounds as samples of the reading domain, but it does not follow that each of them is equally valid or, indeed, that any of them is valid as a measure of the particular reading objectives in a given program or performance contract.

153Ibid.

154Lennon, op. cit., "Accountability."
SUMMARY

Accountability would seem to be with us to stay. Until better assessment techniques are developed to evaluate school impact in the affective domain, and the public comes to accept these evaluations as valid indices of school effectiveness, the principal instrument of accountability will be the achievement test.

As past portrayals of achievement have been based on national norms, any deviation from past practices may well be interpreted as subterfuge. Criterion referenced tests can, and will, serve many systems in their attempt to evaluate performance, or measure achievement gains over shorter periods of time. The final evaluation of performance must, however, be in conventional terms. How does Johnny read in comparison with others in his same grade nationally?

When the nation's first "literacy survival test" showed that at least 13.5 million Americans could not read well enough to function effectively in today's society, further pressure was placed upon our schools to insure that present in-school youth will not become a part of tomorrow's statistics.\(^\text{155}\)

The newly formed National Reading Council is planning an annual survey to establish a "reading index" --an annual report on reading achievement--that can be checked yearly.\(^\text{156}\)


\(^{156}\)\textit{Ibid.}\n
Such studies, and those conducted under the auspices of Adult Basic Education and other Federally funded ventures into basic literacy, place a great emphasis upon the failure of our schools to educate your youth satisfactorily in the basic skills necessary in our society.

Is it any wonder, then, that the emphasis of criticism against public schools is that students are not receiving an adequate education? That public sentiment is building is evidenced by action taken by state legislatures and some school systems to insure that schools do not turn out functional illiterates.

Denver, Colorado has for some time used the Proficiency and Review tests as a determinant of the achievement level necessary to graduate from the system's high schools. Approximately 96 per cent of the seniors meet the requirements, but the remaining four per cent must remain for additional assistance in the basics if they wish the diploma.157

St. Louis, too, commenced using the Proficiency and Review tests in 1970. A Chamber of Commerce study found industrial employers in arms over job applications filed by high school graduates. Board member Schafly reflected the community's opinion when he stated that it was a cruel hoax to hand a high school diploma to boys and girls and turn them

157 Gerald Cavanaugh, Address at the Large City Test Directors' Meeting, Las Vegas, Nevada, April, 1969.
loose on the job market when they are spelling at the fifth grade level.158

California was the first state to mandate a level of achievement for high school graduates. Beginning in 1972, high school graduates must prove eighth-grade competency in reading and mathematics before being permitted to graduate. Schools may use various means of assessing students' competency. One way is through the statewide tests in reading and mathematics which are already required by law at various stages in their schooling. Locally selected exams based upon state-adopted text books, or enrollment in diagnostic and remedial courses are also accepted means of certifying competency.159

Colorado has also joined the list of states experimenting in the assessment of student learning. Others are Michigan, Pennsylvania, Iowa, Kentucky, and New York. Massachusetts is considering such an experiment. These state assessments are outside the National Assessment program, which is designed merely to assess the national progress in education but not to establish standards or proficiency levels. The individual state ventures into assessment have, as a possible goal, establishment of statewide proficiency standards much in the order of the California precedent.160


These state and city minimum graduation criteria are being done one better in other systems. Superintendent Mark Sheed of Philadelphia has adopted "no-nonsense, systemwide reading goals that makes us accountable to the pupils, their parents, and the public." The goals demand not only that students who leave the system at age 16 or older read on at least eighth-grade level, but that all children completing sixth grade read on at least the 5.5 grade level.\(^{161}\)

In all of the above cited instances, nationally normed standardized achievement tests are used in determining if students meet the grade level proficiencies.

Performance contracting, too, has its basis in meeting various proficiency levels. These levels are usually expressed in such terms as "students to gain one year or more in reading and arithmetic" in a certain period of time,\(^{162}\) or students to be performing at grade level within a certain period of time.\(^{163}\) These measures of performances can only be expressed in terms of nationally-normed standardized tests, or instruments equated to such tests.

Under voucher systems there arises a question. If parents are dissatisfied with their children's achievement in public schools, and

---


\(^{162}\)Dorsett, op. cit.

\(^{163}\)"A New First," op. cit.
thereupon transfer their children and the funds to private or parochial schools, how are they to know if the achievement record of these schools is better?

It is conceivable, though not spelled out, that state accrediting agencies could require proficiency testing in all schools prior to the granting of diplomas. California laws apply only to public schools, but those requirements could be expanded. Jencks forsees the Education Voucher Agency as the body to assess the performance of schools participating in the voucher system. It would be an "educational audit" and the EVA would then present the findings to parents upon which their judgements could be based.164

Decentralization could also pose a problem to the public seeking comparative evaluation of schools, or wishing to assess their schools' effectiveness. New York had that problem upon establishment of the three original "demonstration districts". These came about as the result of public dissatisfaction with performance of the schools under the original organizational pattern. Immediately, these demonstration districts, and the Ocean Hill-Brownsville district in particular, discarded old texts and curricula and instituted new and innovative programs. The board and administration of this district defied all city and state mandates and refused to participate in any city-wide testing programs.

---

164 Jencks, op. cit., p. 49.
As a result, it became impossible to evaluate the effectiveness of the new programs in comparison to previous achievement or to the other city programs.\(^{165}\)

It must therefore be considered in any form of decentralization plan that some measure of continuity be maintained in comparability, both to maintain public confidence in the accountability, but more importantly, to provide means of assessing the effectiveness of the decentralization process in meeting the needs of the school system's publics. This is the theme behind Featherstone and Hill's caution that there must be some central control in programming,\(^{166}\) and Brownell's including of overall evaluation as a duty reserved to central administration and board.\(^{167}\)

Criterion referenced tests have their place in the scheme of evaluation, but not the exclusion of standardized norm-referenced tests. Criterion instruments can meet the needs of evaluating local curricula of a system or of an innovative variation within a segment of a decentralized system. Again, New York encourages local evaluation by various means, and by such instruments as locally developed criterion referenced tests, but city-wide evaluation of performance, too, is needed. This is accomplished by utilization of standardized instruments.\(^{168}\)

\(^{165}\)Samual D. McClelland, Interview at Invitational Measurement Conference, Philadelphia, April, 1969.

\(^{166}\)Featherstone and Hill, *op. cit.* p. 44

\(^{167}\)Brownell, *op. cit.*

\(^{168}\)McClelland, *op. cit.*
The determination of what constitutes "normal gain" or expectancy for a particular child or group is another area which needs attention. Whether schools or contractors continue to use grade equivalency as an expectation, adopt Hogan's percentile rank maintenance proposal, or develop more appropriate schema, there remains the problem of justifying the method to the public. 169

Statistical concepts must be considered in their applicability, particularly to experimental programs or performance contracting, where statistical artifacts can distort the true picture of performance, reduce credibility of the accounting, or work to the profit of the contractor.

Test must be used which suit the purpose of measuring the attainment of objectives, whether they be norm-referenced or criterion referenced. Each type has its strengths and its limitations. Of perhaps more importance is the fit of the test to the objectives which the school establishes in conjunction with the several segments of the school community. 170

Chapter II has been a review of the literature as it pertains to accountability, testing and the ramifications inherent in such related areas as performance contracting, voucher systems, are decentralization. Chapter III will establish the rationale behind the development of the survey questionnaire which was directed to the Test Director in thirty-eight major cities.

169 Hogan, op. cit.
170 Klein, op. cit.
CHAPTER III

DESIGN OF THE SURVEY

INTRODUCTION

A survey was made of thirty-eight of the largest school districts in the United States to determine their policy and practice in areas related to the emerging concept of accountability. Since standardized testing is the central theme of today's accountability, the survey was conducted by means of a questionnaire directed to the Test Director in these large systems. In numerous instances the data gathered by means of the questionnaire was supplemented by additional information gained from personnel interviews and/or correspondence.

The actual title of the Test Director varies from system to system as does the area of administrative responsibility wherein testing is placed. The Test Director may be accorded the rank of assistant superintendent, coordinator, director, supervisor, or specialist, though even the hierarchial arrangement of these titles is by no means consistent, system to system. The responsibility for testing is likewise diversely lodged within the provinces of research, guidance, pupil personnel, instruction, or psychological services, depending upon the system. In any event, the Test Director is considered to be the administrative officer in the system having direct responsibility for city-wide standardized testing programs. This individual has been identified by the
association of Large City Test Directors, an affiliate of the National Council of Measurement in Education.  

ORGANIZATION OF THE SURVEY

The survey instrument, which may be found in Appendix A, was divided into four sections to facilitate answering by respondents. Sections I and IV were answered by the Test Director of all participating systems. Section II was appropriate only if the system had already released, to its public, summary school-by-school test data. Section III was directed to those systems which had not as yet committed themselves to the release of such summary data.

Section I

This section contained but a single question, which asked whether the system had, or had not, established a practice of releasing standardized test results to the public on a school-by-school basis. If the response to this question was in the affirmative, the respondent was directed to continue with Section II. If the answer was in the negative, the Test Director was instructed to move on to Section III.

Section II

If the system had already released school-by-school test data, the Director was then asked in question two to check those grades, one through twelve, for which achievement test data were publicly released. Much of the current emphasis upon the evaluation of educational output is directed toward fixing minimum criteria for high school graduation

The Denver use, for several years, of the Proficiency and Review Tests, the recent turn of the St. Louis school system to these same tests, and the graduation criteria established for California schools, are examples of the public pressures for guarantees of educational standards for high school graduates.

The trend must turn, however, toward earlier checkpoints of progress. The challenge, in performance contracts, for students to achieve "at grade level" in their elementary years, and the Philadelphia goal of improved performance by the end of sixth grade, are indications of this trend.

Systems must therefore be prepared with base line data on achievement at various levels if they are to stand accountable for improvements in their educational efforts. The survey enabled an assessment of the degree to which systems have committed themselves to accountability at earlier-than-graduate levels.

Of further interest was the form in which these data on achievement were released and the normative base for these data. Questions three and four, therefore, asked the type of derived score in which achievement

---

2 Gerald Cavanaugh, Address at the Large City Test Directors' Meeting, Las Vegas, Nevada, April, 1969.


5 Supra, pp. 34-35.

data were reported and the normative base used. As long as consistency of form and base are maintained within any one system, the goal of responsible accountability can be attained, though public understanding may suffer through use of less familiar derived scores. Diverse reporting forms and normative bases can be a great handicap in drawing comparisons between system, or in making comparisons with national samples. The Akron, Ohio Test Director recently encountered such a difficulty when attempting to evaluate another system's performance contract as it would apply to the Akron student population. The other system's contract was based upon grade equivalents and national norms, while Akron's data were in terms of stanines, and Ohio norms. Only a very crude conversion schema could be devised.

Intelligence, or more appropriately, academic aptitude test scores, are considered as valid predictors of academic success at a given point in time. If achievement is viewed in relation to expectancy, the perspective may be different. For example, if two schools within a system attain comparable levels of achievement, but the students in one school evidence lower academic aptitude, then their measure of success would be relatively greater. Question five asked whether academic

7 Howard B. Lyman, Intelligence, Aptitude, and Achievement Testing (Boston: Houghton Mifflin Company), pp. 41-45.


aptitude summary test data were released concurrently with achievement data. As in the reporting of achievement scores, the type of derived score used to depict academic aptitude is important, so this question was included as number six.

The child's ability to achieve in school is the product of many factors, some school related, others related more to his socio-economic or ethnic background. If a school's success is to be evaluated, particularly with the view toward making needed changes, modification, or adjustments, then some of these impinging variables should be considered when taking a total perspective of that school.

Some of these school-related variables which are felt to have some possible effect on student learning are the teacher's level of preparation, teacher experience, and the student/teacher ratio. The racial composition of the school population can have dual implications. The special need to deal with problems of minority group students may be compounded if those groups, or some significant portion thereof, are non-English speaking or bi-lingual.

Parental levels of education, socio-economic factors, mobility rates, and attendance, are all felt to have some influence on school achievement, and to that degree should be considered by those seeking to effect improvements.


Some systems which have released school-by-school achievement data have made concurrently available to the public various of these demographic data. In question seven, each Test Director was asked to indicate which, if any, of the above or other demographic data was released along with achievement test information. In a further question, the Director was asked which of these (or other) data that were not released would, in his opinion, have contributed to a better public understanding of the total school situation.

Prior to the release of test data, there were varying degrees of involvement of concerned parties. Question nine attempted to determine whether the system had made any pre-release efforts to prepare the news media, administrators, teachers, or significant representatives of the community to better understand and interpret the data. If such in-service programs or briefings were not held, a further question asked the Director if, in his opinion, such programs or briefings would have been of value in promoting understanding.

The news media are frequently cited as sensationalizing the test score releases, and playing up school-by-school disparities. The eleventh and twelfth questions related to the system's first and subsequent year's experience with the treatment accorded the release by the media.

School staffs may feel threatened by the evaluation of their performance implied in the public release of school-by-school test data. With public sentiment strongly in favor of holding teachers and

12Large City Test Directors, Minutes of Meeting, New Orleans, March 24, 1970, p. 2.
administrators responsible for student achievement, this is a valid con-
-cern for school staffs. In question thirteen, the Test Director was
asked his assessment of staff reaction to the implied threat.

There have been varied public reactions to the confirmation of the
fact that some school's students fair poorly on achievement tests. Some
school communities are completely negativistic, finding no good in tests,
texts, curricula, staff, or buildings. Other communities, no more sat-
isfied with results, lend positive support to effect needed change. In
between are those who are critical but inactive, and those so apathetic
that they voice no concerted opinion. Question fourteen sought to de-
termine the Test Director's evaluation of community response.

It has been noted by some Large City Test Directors that when one
system released school-by-school test scores, the smaller adjacent com-
munities came under increasing pressure for accountability. In
order to assess the prevalence of this phenomenon, the Director was asked, in
questions fifteen and sixteen, to assess the level of pressure for
release of achievement data in adjacent communities, and whether any of
those systems had already followed the lead in releasing school-by-
school data.

The appropriateness of current standardized tests for certain
students has been questioned, but, to the extent that they are relevant

14 Large City Test Directors, Minutes of Meeting, Chicago, February
9, 1970, p. 3.
to the schools established objectives, they have been judged valid. 15, 16

There may be within the school's population certain students who are or should be excluded from standardized testing programs, or at least be omitted from comparative data. Which students should or should not be excluded from the data base has been a source of concern among Large City Test Directors. 17 Educable Mentally Retarded (EMR) students fit the category of students not participating in the regular curriculum who could validly be excluded from participation in city-wide standardized testing. The inclusion or exclusion of certain segments of the student population could greatly influence summary statistics and either raise or lower mean data levels. The practice of various systems toward these exclusions could influence the comparability of data between systems, so question sixteen was included to determine which, if any, students were omitted from their reported data.

Section III

The third section was directed to those systems which had not as yet released school-by-school standardized test data. The initial question, number eighteen, asked for the Test Director's assessment of the level of pressure within the system's community for the release of test scores, school-by-school. Such pressure is seen as an indicator


17 Large City Test Directors, Minutes of Meeting, Chicago, February 9, 1970, p. 2.
of the demand for accountability.

The Director was next asked, in question nineteen, whether the system was contemplating the release of such test data, either soon or in the foreseeable future. Question twenty then asked which (if any) types of data were being considered for release. A check list was provided which included achievement test scores, academic aptitude test scores, along with several items of a demographic nature. Provision was made for the entry of other items not included in the check list.

The final question in this section asked the Director's evaluation of staff concern over the implied evaluation of their performance should summary test data be released.

Section IV

All Test Directors were requested to respond to this last section. The questions were intended to assess the system's position in regard to some of the accountability-related manifestations explored in Chapter II, such as decentralization, performance contracting, voucher systems, criterion-referenced testing, and norm-referenced testing. Question twenty-two asked whether or not the particular system had adopted any plan of decentralization. If the response was negative, the following question asked whether there were community pressures for decentralization as a means of making schools more responsive to community needs.18

Performance contracting was the theme of the next two questions. Question twenty-four asked whether the system was already party to a

performance contract while the next question (if the previous answer was NO) asked the Director's opinion as to whether there were pressures within the school system or community for the adoption of some form of performance contracting.

The move toward voucher systems has been largely the product of dissatisfaction with the performance of the public schools. Some proponents of vouchers may favor the move for racial or religious related reasons rather than dissatisfaction with the public school's performance per se. Of interest, therefore, was any evidenced perceptible shift away from public schools which might presage a future constituency for educational vouchers. Question twenty-six sought the Test Director's opinion as to whether there had been an appreciable shift toward parochial or private schools attributable to a lack of confidence in the public schools performance.

Questions twenty-seven and twenty-eight dealt with criterion-referenced testing. The Test Director's opinion as to the need for such testing was solicited, along with a determination as to whether his system was currently making none, some, or extensive use of criterion-referenced tests.

As there has been growing opposition in some quarters to the use of standardized tests as we know them today, the Test Director was next...

---


asked, in question twenty-nine, if in his professional opinion, some standardized, norm-referenced testing would be needed and retained in his school system. The response section provided the opportunity to relate the need and use to one or more of the following areas: public accountability, system planning, student counseling and staff evaluation.

Many proponents of accountability decry the extent to which the school's effectiveness in developing the cognitive skills is given undue priority over school effectiveness in the affective domain. Most of these writers also admit to the difficulty of assessing student progress in areas such as attitude, self-image, and the like. Question twenty-nine asked of the Director, to determine whether his system was making any attempt to utilize or develop measures of school effectiveness in affective areas.

The final question, number thirty, asked the Director whether, to his knowledge, the local teachers' organization, be it association or union, formally acknowledged the need for educational accountability. Teachers are vitally concerned with the seeming threat which accountability poses. The refreshing note regarding the entrance of the New York City Teachers' Union into an accountability partnership with the school system raised the question as to whether this phenomenon had occurred elsewhere.

---

21 Supra, pp.
22 Supra, pp.
This chapter has dealt with the design of the survey questionnaire and established the rationale, where appropriate, behind the inclusion of certain questions. Chapter IV will tabulate the results of the survey and discuss the implications of these results in relation to the literature.
CHAPTER IV

SUMMARY OF SURVEY DATA

The response to the survey questionnaire sent to the Test Director of thirty-eight of the largest school systems was 100 per cent. A listing of the responding test directors and the cities they represent is included in Appendix A. In addition to the information gleaned from the survey, several directors made added or expanded comments, either in accompanying or separate letters, or in personal interviews with this writer. These latter data will be included at appropriate points in the summary.

Section I

In response to the first question, whether or not the system had already released summary school-by-school standardized test data, thirteen of the reporting systems answered in the affirmative, while twenty-four responded in the negative. The thirteen reporting systems are listed in Table I, below. Only one of these systems, Oakland, California, indicated that there was a good possibility that they would discontinue the school-by-school reporting practice in 1971.

Section II

Question two sought to determine the grade level or levels for which school-by-school achievement test data were released. Table I lists the reporting systems and indicates the grade levels for which
achievement scores are publicly released.

**TABLE 1**

**SCHOOL SYSTEMS REPORTING SCHOOL-BY-SCHOOL STANDARDIZED TEST DATA SHOWING GRADES REPORTED**

<table>
<thead>
<tr>
<th>System</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltimore</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chicago</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Columbus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denver</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detroit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Los Angeles</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oakland</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philadelphia</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Diego</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Francisco</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tulsa</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Washington</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>6</td>
<td>3</td>
<td>12</td>
<td>3</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

The California decision to mandate specific tests statewide and to insure that high school graduates meet minimum achievement levels is reflected in the reporting pattern of the four California cities. Los Angeles Public Schools and Oakland Public Schools report only the
grades for which state mandated tests are given, namely grades 1, 2, 3, 6 and 12. San Diego and San Francisco Public Schools report these same grades, but also independently add measures of achievement in junior and senior high school. These California Public Schools represent four of the six nationally reporting twelfth grade performance.

The number of grades reported range from two in Columbus Public Schools to all twelve in the Denver Public Schools. The trend toward accountability at earlier-than-graduate levels is reflected in the average of 5.8 grade levels reported per system. Of particular interest is the reporting of sixth grade achievement by all but one system. Eighth grade is the second most common reporting point with ten systems including this grade in their release data. Third grade, as an intermediate elementary check point, is included by all but three systems.

The next two questions related to the type or types of derived score used in reporting achievement test results and the normative base or bases of these data. Table 2 depicts the replies.

One facet of the California state mandate on tests is that all system data on these tests be reported annually to the State Department of Education. While there is no requirement that individual school districts release their own test results, either by system total or school-by-school, the summary state data is made available to the public each year.

Three of the four California reporting districts include this state normative data in their release, along with local and national
TABLE 2
DERIVED SCORE USED TO REPORT ACHIEVEMENT
TEST DATA AND NORMATIVE BASE USED

<table>
<thead>
<tr>
<th>Derived Score</th>
<th>Normative Base</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grade Equivalents</td>
</tr>
<tr>
<td>Baltimore</td>
<td>x</td>
</tr>
<tr>
<td>Chicago</td>
<td>x</td>
</tr>
<tr>
<td>Columbus</td>
<td>x</td>
</tr>
<tr>
<td>Denver</td>
<td>x</td>
</tr>
<tr>
<td>Detroit</td>
<td>x</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>x</td>
</tr>
<tr>
<td>Oakland</td>
<td>x</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>x</td>
</tr>
<tr>
<td>New York</td>
<td>x</td>
</tr>
<tr>
<td>San Diego</td>
<td>x</td>
</tr>
<tr>
<td>San Francisco</td>
<td>x</td>
</tr>
<tr>
<td>Tulsa</td>
<td>x</td>
</tr>
<tr>
<td>Washington</td>
<td>x</td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
</tr>
</tbody>
</table>

comparative data. The state data included is based in all instances on the preceding year's statewide summary. The Oakland format (See Appendix B p. 193) is typical of these three systems. Oakland
presents a data sheet on each individual school. One table is in terms of national grade equivalents showing state, Oakland, and the individual school's performance level. Another table compares the Oakland Public Schools and the specific school to the state norms in terms of percentile ranks.

San Francisco Public Schools does not include the state achievement level in its comparison, using only district and individual school performance in terms of nationally normed grade equivalents. The further comparison with state mean achievement would be simple for any interested party, since the state data is readily available to the public.

There are a number of variations in the format used for release of school-by-school test data. Some school systems, such as Oakland, Chicago, and Columbus provide an individual reporting sheet for each individual school. Others such as Detroit, Philadelphia, and San Francisco merely list schools followed by the pertinent statistical data. Examples of these typical schools' reporting formats appear in Appendix B.

Most systems use the median test score as the basis of their school-by-school reporting. Detroit, and Philadelphia differ in that they report on the basis of the school mean score.

For the most part those systems employing a sheet-per-school format go a step farther, reporting in addition to the mean or median, first and third quartile points for each school (plus state and system quartiles where appropriate). Those systems which use the
listing technique merely report the achievement school-by-school based on the median or mean score. An exception to this latter practice is Detroit. Although this system reports in a school lists format, they include in addition to the mean score the range of scores within the school (highest and lowest attained score). Also the range of scores in the "middle 2/3" is included. This shows the scores bracketing the mean as determined by one standard deviation above and below the school mean.

Columbus uses a similar approach, but the portrayal is in graphic form. Bands upon a scale divide the range of scores into four parts, with breaks indicating the position of the quartile points. (Examples of the Detroit and Columbus format are in Appendix B).

These two approaches, Columbus and Detroit, tend to deemphasize the mean scores. They accentuate the fact that there is a greater range of pupil performance within any single school than there is between the highest and lowest schools in the system. The inclusion of quartile points used by other systems is a step in the same direction.

Questions five and six related to the release of intelligence or aptitude test scores, school-by-school, and the type of derived score used in such reporting. The thirteen systems which gave out test scores were almost evenly divided. Six systems which report achievement data school-by-school also include academic aptitude scores in the report. Six other reporting systems release achievement data only. The thirteenth, San Francisco, does not report
academic aptitude school-by-school, but does release the city median "IQ" for two grades, six and twelve.

Of the six systems releasing aptitude test scores, three report both in terms of IQ and percentile rank. One uses only IQ while another uses percentile rank only. Columbus differs completely in that the report is in terms of standard scores (T-Scores). In addition the Columbus graphic portrayal technique is consistent in that range and quartile points for academic aptitude are also shown (See Appendix B. p. 190).

Virtually all systems releasing achievement scores school-by-school also make public various demographic data on the school and/or community. In some instances these data were included as part of the test release as, for example, Columbus and Oakland. In other instances the data were released as a separate report. The Chicago report, which is prepared separately by the Department of Operations Analysis, is an excellent example of the separate format (See Appendix B. p. 187).

The various data which systems made publically available, either accompanying the test score release or separately, are shown in Table 3.

Only one system, Detroit, released no demographic data. This is one of the systems which also did not disclose academic aptitude. San Diego released only one item of demographic nature and that was the racial, or non-white component of the system. All other systems which report school-by-school achievement also made public, in some manner, two or more of these demographic items.
<table>
<thead>
<tr>
<th>Item</th>
<th>Number of Systems</th>
<th>Per Cent of Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student mobility</td>
<td>8</td>
<td>62</td>
</tr>
<tr>
<td>Attendance data</td>
<td>7</td>
<td>54</td>
</tr>
<tr>
<td>Teacher preparation</td>
<td>6</td>
<td>47</td>
</tr>
<tr>
<td>Teacher experience</td>
<td>5</td>
<td>38</td>
</tr>
<tr>
<td>Teacher mobility</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>Student/teacher ratio</td>
<td>6</td>
<td>47</td>
</tr>
<tr>
<td>Economic data (ADC, etc.)</td>
<td>5</td>
<td>38</td>
</tr>
<tr>
<td>Racial composition</td>
<td>9</td>
<td>69</td>
</tr>
<tr>
<td>New to city</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Above age in grade</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Classroom teacher ratio to total professional staff</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>Special fund staff</td>
<td>2</td>
<td>16</td>
</tr>
</tbody>
</table>

Student-related data were most commonly revealed, with such information as racial composition, student mobility, and attendance being most popular. Next in order came school-related factors such as teacher preparation and experience, and student/teacher ratio. Two systems reported teacher experience in the building (teacher mobility).
The only community related item yet found in test or test related releases was of economic nature and varied with the system. Columbus Public Schools reported, for example, the incidence of Aid to Dependent Children (ADC) in the school's student population, while Oakland focused on median family income, which was then related to city medium family income. In other cities, Chicago Public Schools for example, only one familiar with federal or state special program requirements (usually based on welfare, ADC, or other economic criteria) could determine to some degree the economic level of the school community.

Two items appear in the Columbus report which were not made available in any other city's test release. There were "students new to Columbus city schools," and "over-age-in-grade" students. With the completion of a student record file on computer memory, the impact of mobility or immigration upon student and school scores will be more nearly determinable in the near future.

The Columbus School Profile format has a provision for additional information which will be included in the 1971 spring release. These data will include percentage of low-income families, percentage of sub-standard housing, data on population mobility and immigration, percentage of parents who are high school graduates, and data on intact families. These economic, educational, and social factors are all ones which the literature and research (Supra, pp. 20-21) finds related to children's school achievement.
Only one school system, San Diego, which in its first release cited only racial composition, indicated that student mobility, attendance, and family economic status should and would be included in future releases.

Question nine concerned the involvement of certain elements of the school or community in pre-release in-service or briefings. In response, three of the thirteen releasing systems had held no briefings for any individuals outside of the central office. Eight systems held some manner of briefing for members of the news media, though one director admitted that it was "accidental", while another stated that the briefing was "limited".

Principals and other administrators were given pre-release in-service briefings in five systems. Such briefings were designed to prepare these individuals to answer subsequent questions by their staff and members of the local school community.

Two systems, Columbus and Denver, actually made an effort to communicate to teachers the meaning and interpretation of the data prior to the release. Columbus and Tulsa were similarly the only systems which attempted to involve significant members of the community in briefings prior to the release of the data to the general public.

Only one system did not respond to this question. The Los Angeles test director indicated that the first public release had taken place prior to his appointment to the position.
All test directors who stated that briefings had been held by their systems for any of these various publics indicated that the effort had been worthwhile. Several directors further stated that it was their personal conviction that not enough had been done, nor enough individuals included in pre-release preparations. Only one director in a system which did not conduct any pre-release briefings indicated that he felt they would have not been helpful.

The eleventh question related to the treatment of the release by the news media. Those systems which made the release in the form of school-by-school listings usually found these releases reproduced directly in the newspaper. Where the release was more comprehensive, or on a sheet-per-school bases as in Columbus and Oakland, the news media either compiled their own school-by-school listings, or extracted certain data for purposes of drawing comparisons.

Four categories (plus an "other" option) were provided for directors to check as their evaluation news coverage. Though the claim is generally made that news coverage has been "sensationalized", or "negative", (Supra pp. 2 - 3) eleven of the thirteen directors stated that in general coverage by the news media of their test score releases had been factual and unbiased. Two Test Directors, those from Detroit and San Francisco, felt that the releases were treated factually, but were sensationalized. Baltimore, Chicago, and Denver were among those categorizing the coverage generally factual and unbiased, but each indicated that coverage by some specific news reporters leaned toward the sensational or even outright distortion.
One Director alone, indicated that certain reports were slated strongly in support of the city's schools. Examples of the news coverage of several city's test data release may be found in Appendix C.

Columbus Public Schools experience with the news media seemed typical of the general trend, according to exchanges with other test directors. Community or minority oriented newspapers and/or electronic media were generally more critical than their counterparts representing more advantaged areas. The city-wide media varied with the overall policy of the individual paper or station, or of the individual reporter or commentator.

When the Columbus release was made in separate school-by-school format, the media did not attempt to compile an alphabetical or rank order listing, but did extract "typical" inner city schools to contrast with the highest achieving schools. All data reported was factual, in context, but cited in contrast as noted above.

Three of the reporting systems, Chicago, San Diego, and Tulsa have had only one year's experience with school-by-school reporting and so were unable to respond to question eleven. All other directors believed that the news coverage of their second or subsequent year's releases had been accorded the same treatment as the first with one exception. Detroit, one of those finding its first releases sensationalized, reported that the second release was accorded a more positive treatment.

Question thirteen dealt with the directors' evaluation of staff reaction to the threat implied in the release of achievement data.
One director declined to evaluate staff feelings. One, the Oakland director, replied that there were strong staff reactions to the releases. It might be noted at this point that Oakland was the system considering dropping the release practice, though there was no indication given that these facts were directly interrelated. Two directors, those from Columbus and Denver, noted parenthetically that the negative teacher reaction was principally in lower achieving schools, though this is probably a universal phenomenon.

Relative to the reaction of the school communities of the lower achieving schools to the reported test results, New York, Philadelphia, and Washington reported the full range of community reactions among their many schools. Two systems' directors declined to make an assessment of individual school communities. Two systems reported an apathetic reception of their school-by-school releases. These were Baltimore and San Diego. Detroit and Tulsa reported their schools' communities were critical, but passive, though the Detroit director noted isolated instances of extremely negativistic reaction. Columbus, Denver, Oakland and San Francisco report that the school communities of individual low achieving schools were generally critical, but supportive of system school efforts to effect improvements. The test directors of those cities noted that individual school communities, consisting primarily of parents and other interested parties, were critical but supportive, certain segments of the general community were the vocal critics and negativists. These segments were principally minority group organizations actively devoted to social reform on all fronts.
Asked whether they assessed any increased pressure on adjacent communities as a result of their systems’ releases, four directors did not feel qualified to answer. Three felt there was little or no pressure for release in adjacent communities while five sensed a growing pressure in the surrounding areas. Only one, the Denver test director, noted strong pressures for test score release in communities adjacent to his system. Two of the three systems with only one year’s release experience reported pressures in adjacent communities for similar presentations of school effectiveness.

Three of the reporting districts, Los Angeles, Oakland, and San Diego, stated that adjacent communities had already commenced the school-by-school release of test data. This may be a concomitant of the California state mandated test program, perhaps fortified by the precedent of four California systems. Denver reported that as a result of the strong pressures noted above, at least one adjacent system was expected to join the ranks this year.

Question seventeen was intended to determine which, if any, students were excluded in districts from the city-wide testing program and/or inclusion in the summary test data. The response of the directors is noted in Table 4.

It will be noted that the exclusion of EMR students and those with vision difficulties is a common practice employed by all reporting systems. Nine systems exclude the emotionally or neurologically handicapped, though in several cities, including Columbus, these students are allowed to participate in the city-wide testing program, but these scores are not included in the summary.
TABLE 4

NUMBER AND PERCENTAGE OF SYSTEMS EXCLUDING CATEGORIES OF STUDENTS FROM CITY-WIDE TESTING OR INCLUSION IN SUMMARY SCHOOL TEST DATA

<table>
<thead>
<tr>
<th>Category</th>
<th>Number Systems</th>
<th>Per Cent Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educable mentally retarded</td>
<td>13</td>
<td>100</td>
</tr>
<tr>
<td>Bilingual (at least 50 per cent of class time)</td>
<td>4</td>
<td>31</td>
</tr>
<tr>
<td>Emotionally handicapped</td>
<td>9</td>
<td>69</td>
</tr>
<tr>
<td>Physically handicapped</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>Neurologically handicapped</td>
<td>9</td>
<td>69</td>
</tr>
<tr>
<td>Blind or partially sighted</td>
<td>13</td>
<td>100</td>
</tr>
</tbody>
</table>

Section III

Twenty-five of the large cities had not as yet released achievement test results school-by-school. The test director of these cities were asked their evaluation of community pressure for such release. Ten directors felt that there was little or no pressure at the time, while fifteen directors indicated the presence of some community pressures. Seven directors with whom this writer has held personal conversation have indicated that the principal sources of pressure were minority group organizations. Another director interviewed pinpointed the press as the principal source of demands for release of test data. None of the twenty-five test directors cited the presence of strong pressures in their school community.
When asked if their school system was planning the release of test data, only two directors indicated an imminent release. These test directors represented Kansas City and Miami. Miami had reported some pressure for test score releases, but Kansas City is planning such a move despite a virtual absence of demand. Twelve of the cities foresee that they will release test data at a later date. Eight of these cities are ones where there is demand building, but four are ones where there is little or no pressure at this time. Eleven cities do not foresee any future release of test data, even though five of these report the existence of some pressure for release.

When asked in question twenty which (if any) data they felt would be released if the practice were adopted, ten of the eleven directors which had indicated no intention of releasing data opted not to respond. One director admitted to contingency planning, so his system's thoughts were included. Another system which will "probably later" release data had not formulated plans, so was not included in the tabulation. The data which the fourteen remaining respondents felt would be included in their systems' releases is tabulated below in Table 5.

Some interesting similarities and differences appear when a comparison is drawn between the demographic data actually released by those reporting systems and the data projected for release by those not yet reporting school-by-school achievement. Table 6 presents the comparative percentages of these two categories of cities relative to selected demographic items.
### TABLE 5

**NUMBER AND PERCENTAGE OF SYSTEMS PLANNING THE REPORTING OF VARIOUS DEMOGRAPHIC DATA IN FUTURE PUBLIC ACCOUNTABILITY RELEASES**

<table>
<thead>
<tr>
<th>Item</th>
<th>Number of Systems</th>
<th>Per Cent of Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achievement data</td>
<td>14</td>
<td>100</td>
</tr>
<tr>
<td>Attendance data</td>
<td>8</td>
<td>57</td>
</tr>
<tr>
<td>Academic aptitude data</td>
<td>6</td>
<td>43</td>
</tr>
<tr>
<td>Student mobility</td>
<td>9</td>
<td>64</td>
</tr>
<tr>
<td>Students new to system</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>Students above age-in-grade</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>Racial composition</td>
<td>4</td>
<td>29</td>
</tr>
<tr>
<td>Economic data (ADC, etc.)</td>
<td>6</td>
<td>43</td>
</tr>
<tr>
<td>Parents education level</td>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>Teacher preparation</td>
<td>4</td>
<td>29</td>
</tr>
<tr>
<td>Teacher experience</td>
<td>9</td>
<td>64</td>
</tr>
<tr>
<td>Teacher mobility</td>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>Student/teacher ratio</td>
<td>9</td>
<td>64</td>
</tr>
</tbody>
</table>

The reporting and non-reporting systems are within five percentage points of agreement on the release of five demographic items. These are academic aptitude, teacher mobility, economic data, attendance data, and student mobility. Of most dissimilarity are the reversed emphasis on teacher preparation and teacher experience. A greater number of
presently reporting systems tend to present data on teacher preparation as opposed to teacher experience, while the non-reporting systems seem to weight experience as a more significant factor than level of training.

TABLE 6
PERCENTAGE OF REPORTING AND NON REPORTING SYSTEMS RELEASING OR CONSIDERING THE RELEASE OF SELECTED DEMOGRAPHIC DATA

<table>
<thead>
<tr>
<th>Data Category</th>
<th>Per Cent of Reporting Systems</th>
<th>Per Cent of Non-Reporting Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic aptitude</td>
<td>46</td>
<td>43</td>
</tr>
<tr>
<td>Teacher preparation</td>
<td>47</td>
<td>29</td>
</tr>
<tr>
<td>Teacher experience</td>
<td>38</td>
<td>64</td>
</tr>
<tr>
<td>Teacher mobility</td>
<td>16</td>
<td>21</td>
</tr>
<tr>
<td>Racial composition</td>
<td>69</td>
<td>29</td>
</tr>
<tr>
<td>Economic data (ADC, etc.)</td>
<td>38</td>
<td>43</td>
</tr>
<tr>
<td>Attendance data</td>
<td>54</td>
<td>57</td>
</tr>
<tr>
<td>Student mobility</td>
<td>62</td>
<td>64</td>
</tr>
<tr>
<td>Student/teacher ratio</td>
<td>47</td>
<td>64</td>
</tr>
<tr>
<td>Students new to system</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>Students above age-in-grade</td>
<td>8</td>
<td>14</td>
</tr>
</tbody>
</table>

The most divergent, and probably most significant factor is the reporting of racial composition. This was the point of greatest agreement with the presently reporting systems, being presented by 69 per cent of those systems. Only 29 per cent of the systems anticipating
data releases are planning to include the racial factor in their presentations.

Finally, more non-reporting systems favored the inclusion of student/teacher ratio data than was the case with reporting systems. Slightly less than half of the latter had released such statistics as part of their test related communications, but 64 per cent of the systems yet in the planning stage project inclusion of this ratio as a pertinent factor.

The final question in Section III asked the director's assessment of staff opposition to the release of school-by-school achievement data due to the implied evaluation of their efforts. Four of the twenty-five directors did not feel qualified to assess staff feelings. Five staffs were felt to have expressed no opposition, but these same five were cities with no plans for release. Were system policy to change, there could be a change in staff position. Thirteen city staffs were assessed as expressing opposition to accountability releases of school performance. This group included the two cities planning immediate releases and six of those considering later release.

The only strong staff opposition reported was noted in Portland and carried the emphasis of "very". Portland noted some community pressure but has no projected plans for release, perhaps because of staff opposition. Miami's director sensed "some" staff opposition but noted a "strong" stand by teacher organizations. The Mobile director stated that the release of test data was seen by some staff members as a racially related issue, and support or opposition had developed along these lines.
Section IV

Question twenty-two asked whether the system had adapted any plan of decentralization. The response was evenly divided, with nineteen reporting the implementation of some form of decentralization, while nineteen replied negatively.

There was no practical way of determining the exact degree of decentralization, or its effectiveness in meeting community needs. The added comments of several directors, either in the survey form or in personal conversation revealed no plan nearly as comprehensive as that of New York City. (See Supra, p. 49)

Philadelphia is divided into eight districts, but little autonomy is granted the district superintendents and their staffs other than in the mechanics of administration. St. Louis, too, is divided into districts, and district superintendents are appointed. Local "advisory" groups assist the superintendent in local planning. Innovation is encouraged, but only within not-to-flexible guidelines.

The six districts in Miami are merely administrative subdivisions. Policy and curricular controls are tied to central administration. Only routine administrative decisions may be made at district level, and even these within a restrictive framework. Milwaukee's director calls their decentralization a "token plan." He states that there has been no change in central administration. Contrasting high school districts are designated as clusters and are combined into Program Service Areas.

Pittsburg also reports that the decentralization plan adapted in September, 1970, is "modified" and still essentially centralized. The
director cites two difficulties in implementing an effective plan of decentralization. One is the comparatively small city size — near 75,000 students, and inter-geographic area student mobility.

The directors of four of the decentralized systems report strong pressures for more effective plans to place greater control at the local levels. Here administrative subdivision is not seen as responding to local community wishes or needs. These four systems are Detroit, Los Angeles, Oakland, and St. Louis.

The directors of the nineteen cities with no decentralization were asked whether there was pressure in their school community for decentralization as a means of meeting local needs.

Six directors responded that there were no pressures on their systems for decentralization. Eleven directors indicated that there was some demand in their communities, while two, those of Baltimore Public Schools and Clearwater Public Schools, felt that there were strong pressures for decentralization. The Clearwater director further related that decentralization was "in the planning stage" in his system.

Two other systems in the "some pressure" category also are moving toward decentralization. The Tulsa director stated that their "aim is to move in this direction." Kansas City's director has been told by her superintendent that "it (decentralization) is to happen in the near future."

In summary, outside of the New York City plan where locally elected boards have a voice in policy matters, no other decentralization plans seem other than administrative conveniences.
In response to question twenty-four, the test directors indicated that only three cities had actually entered into performance contracting. These systems were Dallas, Oakland, and Philadelphia. (The Dallas and Philadelphia contracts are described - Supra, p. 34). No details are yet available on the Oakland contract.

Five other cities, Atlanta, Minneapolis, Nashville, Seattle, and Wichita are under pressures to explore performance contracting, according to their test directors. The Minneapolis pressure is not so much from the community as from the Office of Economic Opportunity which wishes their participation in experimental studies.

When asked whether there was no preceptable shift toward parochial or private schools attributable to a lack of confidence in public school performance, six directors responded in the affirmative. Four of these are understandable in light of the controversy over desegregation in the South. These systems were Atlanta, Miami, Jacksonville, and Mobile. The feeling was expressed well by the Mobile director who reported that there was fear in some quarters that the entry of large numbers of blacks into the schools would lower achievement levels.

The trend noted in the other two cities, Baltimore and Oklahoma City, is similarly race oriented as there are redistricting and/or bussing pattern problems which have implications for dissatisfaction. Seattle's director injected an opposite note, but by virtue of the fact that the city's parochial schools are closing. Pittsburg, too, is noting a decrease in both private and parochial school enrollment. Shifts in that
city are not away from public schools, but some requests are reported to follow each particular school disturbance.

Question twenty-seven related to the need for more criterion-referenced tests as better measures of achievement, particularly over short periods of time. Only 17 per cent of the directors felt that there was no particular need for such instruments, although two of these directors admitted that there were some thoughts about their use within their system.

Although 83 per cent of the directors feel that criterion-referenced tests are needed, only 35 per cent of the systems employ such tests, none extensively. Seattle uses some criterion-referenced measures in the mathematics area only. Atlanta makes some use of criterion measures in curriculum improvement studies, while Portland finds such tests an aid in diagnosis of student learning problems.

All directors feel that there will continue to be need for standardized norm-referenced testing. The uses which they foresee for such data do vary. Four principal uses for test data were listed on the questionnaire, with space provided to insert any others which the director might feel appropriate. Table 7 shows the number and percentage of systems whose directors feel will need, and continue to use, norm referenced testing for various purposes.

Six directors report no felt need for testing for accountability purposes. These same directors are in systems with no plans for releasing test scores to the public, so the answers are consistent. Only two directors, those from San Francisco and St. Louis, expressed no need for standardized testing for system planning. Both directors,
however, are ones interested in criterion-referenced instruments. Three of the directors indicated no need for standardized testing for use in student counseling. This is particularly interesting since one is an assistant superintendent also charged with guidance responsibility, and the other two have backgrounds as school psychologists.

**TABLE 7**

NUMBER AND PERCENTAGE OF SYSTEMS WHICH WILL CONTINUE USING NORM REFERENCED TESTING FOR SPECIFIED PURPOSES

<table>
<thead>
<tr>
<th>Use</th>
<th>Number of Systems</th>
<th>Per Cent of Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public Accountability</td>
<td>32</td>
<td>84</td>
</tr>
<tr>
<td>System Planning</td>
<td>36</td>
<td>95</td>
</tr>
<tr>
<td>Student Counseling</td>
<td>35</td>
<td>92</td>
</tr>
<tr>
<td>Staff Evaluation</td>
<td>13</td>
<td>32</td>
</tr>
</tbody>
</table>

Half the directors reported little or no activity in assessing affective factors, while the other half noted some attempts to establish such measures. Almost universally these attempts were in association with federally funded programs or in special projects. No system reported any broad or general application, except Miami, whose director stated that the Maryland Scale, "Self Concept as a Learner," had been used a few times system-wide. Several other directors made added comments such as "takes resources," "extensive in-services needed for staff," "we don't know how to get at this validly," and "this is being seriously studied."
The final question asked the directors' evaluation of their systems professional organizations views on accountability. The directors of Atlanta, Clearwater, Denver, Honolulu, Houston, Miami, Seattle, and Wichita report that their association or union had agreed to the concept of accountability, though not always in contractual terms as in New York.

Dallas organizations have agreed to accountability, but the director reports that there is disagreement concerning the newly initiated performance contract experiment. San Diego's director relates that their professional organization is in agreement with accountability as long as they are given a greater voice as to the uses which are made of the data.

Chapter IV has summarized the information gained from the survey questionnaires returned by the thirty-eight large city test directors. Chapter V will draw upon these data and the literature reviewed in Chapter II, and attempt to formulate recommendations for systems not yet faced with the problems of accountability. These recommendations will be presented to several national authorities in the field of testing and research and amended in accordance with their suggestions.
CHAPTER V

RECOMMENDATIONS AND CONCLUSIONS

All of the indications are that the accountability concept in education is here to stay. The earlier demands of the schools' publics for information about how well the schools were educating their children has taken an added turn. Much of the impetus to the emergence of the concept in its present form stems from questions regarding the cost-effectiveness of various educational programs. The ever increasing requests for greater funding of educational institutions have raised questions concerning whether these additional expenditures really result in improved education.

Until recently, most of the appraisal of education and educational programs has been in terms of input. For example, these imputts were stated in terms of instructional materials, teachers' salaries, class sizes, or the institution of programs designed to reduce the number of dropouts.

The new approach holds that those whose task it is to provide education for children should be held accountable for their efforts in terms of the output in the educational program. For example, these outputs should be in terms of improved student learning, changes in student behavior, or reduction in the number of dropouts.
Accountability in terms of input, or monies expended, provides little evidence of pupil performance resulting from the educational program. Accountability in terms of output allows the assessment of programs in terms of changes in pupil behavior. If cost-effectiveness is to be evaluated, then change in student behavior must be interpreted in terms of expenditure, that is, how much change in student learning or achievement results from specified expenditures.

Accountability—By Whom and for What?

The whole matter of accountability must be examined within the framework of four basic parameters: accountable to whom, by whom, for what and in what manner.¹

There has been accountability in education for years, but primarily in the fiscal sense. The board of education has been accountable to the taxpaying public for the wise and prudent expenditure of the tax dollar. The "to whom" under the new concept of educational accountability is now expanded to include being accountable to the parent and student for satisfactory achievement of the goals established by the schools.

The "by whom," as noted above, formerly was the board of education. The board was charged with the legal responsibility for wisely handling the public funds. Under the new concept of accountability, not only

does the board retain its fiscal responsibility, but the Board, central administration, principals and classroom teachers are all held accountable for student performance.

The "for what" has similarly expanded. The fiscal component of accountability has been joined by the need to account for accomplishment of the schools' mission in educating children.

The final parameter, "in what way," must be expanded from the earlier concept of dollar accounting for educational input, to include accounting for educational output in terms of units of student learning and changes in student behavior.

**Components of Accountability**

In order to apply the new concept of accountability to education, it is necessary that certain components be present. There must be measurable objectives in terms of pupil behavior. There must be programs, in most cases a learning program, which are designed to lead to the achievement of these measurable behavioral objectives. There must be an evaluation of how well the program has led to the achievement of these objectives, together with an assessment of the cost of the program. And finally, there must be a systematic feedback of information to those responsible for program policy so that appropriate changes may be made to improve the program. 

---

Many of the features outlined above have been used by educators for years. What is new is the integration of all of these components into a unified whole, or what might be called a "systems approach." Such an approach, it is contended, promotes more efficient planning and carrying out of an educational program, and is also conducive to the determination of the cost-effectiveness of a program.

In addition to the above noted components, any practical implementation of the concept of accountability would need to contain certain conditions. These might include hours of pupil exposure to programs, specified time limits to achieve certain objectives, or the specified group of students to be affected.

Thus, in summary, accountability would contain requirements that there be evidence of the achievement of clearly stated objectives, by a specified group of students, using well defined procedures, within a given period of time, together with a determination of the cost of the program.

General Recommendations

The educational objectives of any school system are dependent in large part upon the needs and desires of the residents of that particular community. Accountability will not gain public acceptance unless it is accountability for the attainment of objectives which are held by the community to be of value. The larger the community the more complex will be the task of determining its needs and desires. In our highly pluralistic society a major problem is establishing priority goals
and objectives. Schools must be responsive to many different audiences, with their varying value structures, life styles, and economic and educational concerns. Unless care is taken to assess their several needs, and reconcile them within the framework of educational planning, we are doomed to fail some one or more segments of the community.  

Recommendation I

Viable communications must be established to gain input from all segments of the educational community. This should include the establishing of realistic educational goals and objectives. These goals and objectives may of necessity be varied or modified by sub-district or by school. This communication may be accomplished, as locally feasible, by central committees, by local advisory groups, by complete decentralization, or by some combination of these approaches.

All goals of education must be stated as a series of measurable behavioral objectives. The literature pertaining to accountability places as much emphasis on "measurable" as it does on clear statement of objectives. If a school system is to be judged on its effectiveness, then the objectives must be stated in such a manner that it is possible to determine whether or not these objectives have been met. This must

hold true whether the objectives are within the cognitive or the affective domain.4

Recommendation II

After the educational goals of the district, sub-district, or school have been determined, the entire educational program pertaining to both the academic and socio-personal development of students must be reduced to a series of specific, measurable, behaviorally-stated objectives.

If a school system is to commit itself to being accountable to its publics, it must be prepared to give evidence of the success or failure of the programs instituted to attain the goals and objectives which have been adopted. To provide such evidence the system must have, or develop the capacity for program evaluation. If the system is large, it may be feasible to utilize or develop its own research and evaluation staff. This has been the path taken by the many systems represented in the Large City sample surveyed.

If the system is small, and incapable of supporting an evaluation staff, there are two options which may be taken. Both, according to the literature, have proved successful when used in conjunction with special projects or Federal programs. One is to enter into a contract with

an outside agency to provide the evaluative service. The other is to enter into a joint venture with one or more small systems to form an evaluative component on a shared cost basis.\(^5\)

The literature further cautions against the program implementor also being assigned the task of program evaluation. There could be the claim that the evaluation was overly biased by the relationship, even if this were not the case. The essence of accountability is credibility and any factor which could lessen the credibility should be avoided.

Recommendation III

An effective evaluation component must be established, either within or without the system, whether integral, contracted or shared. This component can then certify to the school and the community as to the effectiveness of the programs in attaining the stated objectives.

The evaluators should be sufficiently divorced from the programs themselves so that a reasonable degree of objectivity may be obtained.

Evaluation of the attainment of academic instructional objectives is much easier than measurement of personal-social growth. The literature points both to the need for, and the dearth of reliable instruments in

the affective areas. This need is particularly great if schools are not to be forced to over emphasize the academics to the neglect of affective areas merely because of accountability requirements. The directors whose systems are making efforts to evaluate personal-social growth (50 per cent) further note that special and Federal projects are providing the "seed money" to stimulate research in these areas. Hopefully, much good can come from these beginnings.

Even in the more familiar areas of the cognitive skills, there needs to be careful consideration. There are many readily available commercial standardized tests which purport to measure "reading" or "arithmetic" skills. Yet even within these instruments there is great variance both in content covered and in the emphasis placed on various aspects of content. For this reason, one test may be a more valid instrument than another for measuring what is taught in a specific program. Only a careful examination of the content of both the program and the test can determine their congruency.7

Recommendation IV

Selection of evaluative instruments or tools must be made with the utmost care. Program content and the stated behavioral objectives must be carefully examined,

6 Supra, pp. 18-21.

and equated with the available instruments in order to obtain ones with the best possible fit. Not only must the academic areas be so treated, but all possible effort must be made to evaluate objectives within the affective domain.

Accountability, within its new context, must be stated in terms of output, but program evaluations must also relate these outputs to other factors. Some of these factors are school related. Sixty-two per cent of the reporting systems included in their release data one or more factors related to the school itself, such as student mobility, teacher preparation, class size, academic aptitude levels, etc. Other factors, also related in the literature to student learning, are outside the school, but concern the community. Almost 70 per cent of the releasing systems reported one or more of these community-related factors such as racial composition, ADC rate, or family median income.

The systems not yet reporting followed a similar pattern in presenting their considered judgement about including demographic data. Sixty-four per cent project the release of one or more school-related factors. Fewer intend, at this time, to present community-related data, though almost half feel that some community factors are to be given a place should they become committed to accountability.
The relevance of these in and out of school data to the study of school achievement problems is well documented in the literature. The principle problem will be the acceptance of the relationships by the public. If such data is seen by a concerned community as an attempt by the schools to escape their responsibilities, the release will backfire. The schools must take special pains to educate their publics in the significance of any factors they chose to include.

Hopefully, all segments and sectors of the public will be involved in school and community planning. The two must go hand in hand. If classroom ratios are excessive for example, and are felt to have a detrimental effect on achievement, this is a taxpayer concern, as additional dollars will be needed to hire additional staff. If teacher mobility is considered a problem, then system policy or teacher organization transfer demands must be reconsidered. If racial ratios are thought to be causal of some difficulty, their redistricting or bussing might be considered, or perhaps long-range attention needs to be considered by fair housing authorities.

Recommendation V

Each factor which is thought to have a bearing on student achievement, whether school or community related, should be evaluated for possible release. If it is the considered

---

8Supra, pp. 27-28.

judgment of those preparing the data, that the release of the specific data can be turned to a positive end to bring school and/or community attention to bear with the view of improving or correcting the condition, it could be an item worthy of inclusion.

The focus of some elements of accountability has been upon establishing standards for high school graduation, even to the point of legislative mandate. This may be very defensible in terms of quality control of the school product, but would be completely indefensible if schools were to permit students to complete twelve years of education—only to be refused graduation—unless there were planned points of intervention at several earlier stages in their school career.

The literature showed one system which has already established a sixth grade checkpoint. Consideration might also be given to a pre-graduation check at grades ten or eleven, so that if remediation were necessary, there would still be time. It is a usual fact that reading is not a formal subject past grade eight, and the same holds true for arithmetic past grades nine or ten, with the exception of students whose program requires additional mathematics.

The survey of reporting systems showed the minimum release as two grade levels; grades six and eight. That system’s testing program


encompasses grades one through ten, but the release is restricted to the two noted. The most frequently reporting system included all twelve grades, but the average was 5.8 grades reported.

There are certain periods within a child's school career in any particular system where there are natural transition points. The elementary to junior high stage is an obvious example, but system patterns vary. There may be the "middle school" type of organization. Specialists in basic skills may be concentrated at certain levels, or compensatory education programs available only at certain points. All of these circumstances should be considered when establishing a testing program with accountability as one of its goals.

Accountability is not an end in itself. Its purpose is to focus attention, in the school, in the system, and in the community, upon needs for improvement. Any point in the students' careers where success or failure to meet school objectives can call attention to need for program replacement, modification, improvement, or change should be considered as a point for the focus of accountability.

Recommendation VI

Accountability is not just for visibility but must serve a purpose. The purpose is to bring attention to any need for change. To this end the school should be accountable for performance at each major juncture in the total school program, and at as many intermediate points as it is felt the data could be meaningfully employed.
There has been raised, in some communities, the question of whether the releases are completely factual, or whether the data has been screened to make it look better than it is. One factor has been the claim that certain "poor" students were not tested, or, if tested, their scores were not included in the school results.¹²

Schools and school systems must be geared to accommodate virtually all children. Some very few students, by virtue of careful screening, are considered "excludable" from public schools. In Ohio, for example, children who are certified by school psychologists to score less than a 50 I.Q. on an individual psychological examination, using either (or both) the Stanford Binet or Wechsler, are excludable. Such action is usually taken only after the child has been very carefully tested and retested.

Children scoring above the 50 IQ point, up to and including 70 IQ or 80 IQ, according to varying state standards, on individual psychological examination, may be classified as Educable Mentally Retarded. Special programs are provided for these children, so it is only logical that they not be held responsible for the same academic content or progress as other children. In addition, classes for EMR are not structured grade wise, so inclusion of test data from such classes would not be feasible.

¹²Robert Walker, Minutes of LCTD Meeting, Detroit, Michigan, 1968, p. 2. (Mimeoographed)
Other special classifications of children are similarly provided for in ungraded or special classes. The directors surveyed were asked how the data on various categories of "special" students were handled.

All thirteen releasing cities reported that EMR and visually handicapped children were either not tested, or the data excluded from summary statistics. Sixty-nine per cent of the systems also excluded students considered as emotionally or neurologically handicapped. Four systems reporting from areas with major concentrations of Spanish speaking students, noted the exclusion of those students for which English was the second language. Thus it may be said that the exclusion practice was very extensive, but for very justifiable reasons only, and that data on the various systems could therefore be considered comparable.

One practice noted by several systems has merit, in that it provides for the psychological support of the children affected. In too many instances children in "special" classes are made to feel "different" and are excluded from many normal school experiences. Testing is a school-wide activity, and when "special" students are obviously excluded, there may be a further psychological blow. Their inclusion in testing by these cities has a two-fold benefit. The children are made to feel "included," and the data can be useful to their individual teacher. Their scores need not be, and are not included in the summary data.

---

Recommendation VII

EMR and other special students may or may not be included in the testing program, according to building or program conditions. If the content of their program is not congruent to the test content, then the data may be validly excluded from inclusion in school-by-school or system summary data.

If identifiable elements of the community are included in the planning stages of school goals and objectives, it is but a further projection of that same philosophy if representatives of these elements be included in the evaluation and pre-release stages. Since it should be an obvious need to include significant numbers of the community in developing an understanding of data prior to the release, it is difficult to understand why only two of the thirteen releasing districts reported such involvement.\(^\text{14}\)

Recommendation VIII

All segments of the community should be included in the formulation of school objectives, and should be involved in relating school and community factors to school performance. Therefore, these same segments should be represented in the evaluative stages so that they may help interpret the data to their local community.

\(^{14}\)Luvern L. Cunningham, "A Report to the Columbus Board of Education," The Ohio State University, June, 1968, p. 8.
Many individuals are not part of any cohesive group within the community. They must rely for the receipt and interpretation of data upon the mass media. Specific media, whether press or electronic, may be the voice of some particular community group or city sector. Their editorial policy may reflect that group's or sector's beliefs or biases. Other media policy may depend upon political affiliation or reporter opinion. According to the Survey, most media did attempt to present factual information, even though in instances editorialized.

If schools are to expect the proper coverage of their releases by the media, it will behoove schools to provide these media representatives with the full picture. To this end they must include the media in all stages of planning, evaluation, and release. All that schools do will not reflect success. These failures must also be reported fairly. Admission of failures will lend greater credence to reports of successes.\textsuperscript{15} there must be no credibility gap in the schools' program of accountability, and the media will be among the first to recognize and air any duplicity.

Recommendation IX

The representatives of the various news media should be included in the planning for accountability. If they are fully aware of the meaning and intent of all data included, the "power of the press" may be an invaluable aid in working for the equitable solution of school and community problems related to education.

\textsuperscript{15}Merriman, op. cit., "Case Study", p. 8.
Over half of the directors surveyed indicated that school staffs felt some measure of threat in the implication that their efforts would be evaluated as the result of school-by-school data releases. Some teacher organizations, both association and union, have taken positions on accountability. However, the consensus of these groups seems to be that they want to be a part of accountability, not merely subject to it.

The literature has revealed that the public wants teachers to be accountable for their performance, even to the point of differential or merit pay. The literature further revealed that much professional thinking has to do with the reward/penalty phase of school staff accountability, whether in the traditional school setting or in innovative situations such as those created under performance contracts. (Supra, pp. 26-29)

Teachers form one of the publics within the school community which must be involved in all phases of accountability. Unless they have their fair share of input into all phases of planning, their fullest understanding and cooperation cannot be expected. If there is an element of unwarranted threat, there will be compensating reaction which can take many forms, even to the point of deliberate or unthinking contamination of evaluative data. (Supra, pp. 32-33)

Teachers, further, are on the firing line, so to speak, with parents. This is no less true of other building staff including principals and counselors. All of these individuals should be involved in the
original planning of objectives. Counselors, particularly, should have input into the consideration of socio-personal objectives. Total staff should be fully aware of the reasons for inclusion of various demographic data in the release, for upon them will fall a considerable burden of public interpretation.

Recommendation X

Teachers, principals, and counselors should be given a voice in formulating general and local objectives, both cognitive and affective. Their knowledge of the local school community and familiarity with local problems can be invaluable in helping to state local needs. They must further be involved in the evaluation and release stages, for understanding will both reduce threat, and make them more effective ambassadors to the local community.

There are numerous options as to the release format which may be employed or the manner in which data are reported. The examples shown in Appendix B are representative of major variations. Released data may consist only of achievement test scores, listed school-by-school and depicting the mean or median point. Academic aptitude test mid-points may be given, along with various other demographic data. These additional data may be released as part of the achievement reporting or may be presented in separate documents.
Some schools surveyed chose not to merely report the mid-scores but to include the range and/or quartile points within each school. This type of release tends to de-emphasize the midpoint, for too often parents and others may think of this central score as representing all of the student population. There must be awareness that half of the students score at or above a median score, and this positive aspect must be kept in mind. Of course the other half score at or below the median, and the median itself may well be below city or national norms, but the entire perspective must not be lost.

The survey and the literature both revealed that grade equivalents and IQ were by far the most common derived scores used for reporting achievement and academic aptitude respectively. These two types of scores are known by professionals to have serious weaknesses, but the reason stated most often for their continued usage is that people, both inside and outside of education, are familiar with grade equivalents and IQs, although there is considerable misunderstanding and misinterpretation of the concepts behind these derived scores.16

Only one of the thirteen surveyed systems which have released achievement data did not report in terms of mean or median grade equivalents. That one system reported in percentile rank only. Six of the systems used both grade equivalent and percentile, with only one system including, in addition, a stanine score.

16Durost, op. cit., pp. 299-300.
Half of the systems which also reported academic aptitude reported both in terms of IQ and percentile rank, while one each used only IQ or percentile. Only one system deviated, in that it reported in a standard T-score, showing school and system range and quartile points.

Thus grade equivalent was almost universally used as the reporting score, while the IQ concept is diminishing in favor.17

Until such time as there is greater public understanding of the meaning of other types of derived scores, and as a consequence greater acceptance, then grade equivalents will probably be the test means of reporting achievement test scores. The inclusion, also, of percentile rank may be a way of bridging the gap in public understanding.

This writer shares a bias with Downie18 toward standard scores. These relate student performance on test to the student's age or grade peer group, and not to the interpolated or extrapolated performance of other student groups. Unlike percentiles, which satisfy this same requirement, standard score intervals represent equal increments of performance rather than mere rankings. However, to date, there has been no appreciable acceptance by either professionals or lay publics of the standard score concept. Witness, only one system uses standard scores alone to depict academic aptitude, while one other uses stanines


in addition to grade equivalent and percentile scores to portray achievement. These two releases take a step, albeit a small one, toward educating the public to an understanding, and acceptance of standard scores.

All systems surveyed related their individual school data to the national normative base of the instruments used. They also included, among their data, the local system mean or median, as appropriate. In the one state where a mandatory state testing program permitted the development of state norms based upon the preceding years, the majority of systems also related local data to state achievement levels.

The principle criteria of any format of accountability are understanding and acceptance. Public understanding can be best achieved by the inclusion of representatives of the various segments of the community in planning. This statement merely reinforces the repeated theme, involve the community.

The report may be tabular or graphic, but the school-by-school individual report offers the best means of communicating to individual school communities. All data pertinent to that school are included in a single presentation. National, local and state (if appropriate) comparisons may be drawn, but there is less of a tendency to compare one school to another. The attention is thus focused on the needs, the strengths, and the weaknesses in that particular school and community.
The units of measure used to report data must be understood, and, again must be the subject of community education. For example, people feel they understand grade equivalents, but must be trained in the concept of standard scores or even percentile ranks. The planning for the reporting scheme must accompany the statement of objectives. If the objectives state "bring students up one grade equivalent," or "up to grade level," then reporting performance in terms of percentile rank would be inappropriate.

Recommendation XI

The release format and reporting units should be planned at the onset of the program of accountability. The format should be kept as simple as is consistent with the data to be presented, and should report in the measurement terms used to state the behavioral objective. Involvement of the various publics at all stages of planning and release will maximize the understanding and therefore acceptance of the release as a sincere and credible report of accountability.

Decentralization

Decentralization has been presented in the literature as a means of permitting individual schools and/or clusters of schools to meet more fully the needs of local communities. ¹⁹ Especially in the large

¹⁹ Supra, pp. 46-47.
urban complexes, where are pluralistic society evidences its greatest range of diversity, is there a great need for the review of school curricula.

It should be obvious that no one fixed program of instruction or pattern of courses can fully meet the needs of students of varied ethnic minorities or from different socio-economic backgrounds. Nor has it proved feasible for a central administration to determine and provide adequately for the needs of the many segments of the community. These administrators are too far removed from the local community to respond to local needs.20

A number of systems, both according to the literature and the survey have adopted some degree of decentralization, but for the most part this has been an administrative gesture rather than a philosophical response. Management principles of the "span of control" dictate that no central administrator can effectively "manage" a large number of individual schools. Sub-districts are organized, with local assistant administrators assuming the direct supervision of schools within the sub-district. For the most part these reorganizations are for effective administration and bear no resemblance to a plan for community control.

Only one system of those surveyed has effected any meaningful change. Its decentralization plan provides for the election of local boards which have authority, though not unlimited, over personnel and curriculum. Such a plan can make local schools responsive to local needs, yet maintain a needed measure of central surveillance.

Recommendation XII

Decentralization should be considered as a means of meeting local schools needs. If the plan does not provide for locally elected boards, there must be provision for broader flexibility in local policy and program, and the local community must be involved in a meaningful way in planning and decision making.

If a plan of decentralization is adopted, accountability must be maintained to the greater district-wide community. To this end, the central board must maintain an overall evaluative perspective. As shown in the literature, several "demonstration districts" assumed such autonomy that valid evaluation of the measure of success or failure of their innovative programs was impossible. Innovation as an end in itself has no justification. Innovation which results in improved academic achievement or greater measures of socio-personal growth is what is needed. Only by continued evaluation may the effect of innovation be determined.

There may be variations of program content and emphasis school-by-school, but the theme of the basic skills, for example, runs throughout. Reading and arithmetic are common denominators found in all systems reporting school-by-school, under decentralization. Basic skill areas are the ones in which the public has evidenced the most universal concern.

Recommendation XIII

Though the system may decentralize, certain elements of central control must remain. One of these is the component of evaluation. Accountability dictates that the public be kept aware of the results of program change or innovation, and the effectiveness of any investment in educational reform. The data base, for comparative purposes, must maintain consistency through the system, though special or different instruments may be needed for selective intra-subdistrict or intra-school evaluation.

Performance Contracting

Performance contracting looms large upon the educational horizon as a panacea for many school ills. Funding of experimental contracts with Federal funds has enabled many districts to implement innovative programs without risking local funds. The programs are usually focused upon problems which plague the schools, such as dropouts, low achievers, and the like. Boards find the performance contract a potential tool in
negotiation. The concept has many appeals to many people. Yet there are implications which may not be readily apparent. It would behove the system contemplating the entry into performance contracts to be aware of these.

First, there are certain regulations in virtually all states which require school boards to advertise for bids before contracts are awarded. The flagrant violation by one district of this percept was noted in the literature. Another legal implication dealt with the question as to whether or not a board could indefinitely contract for a prime service which its own staff could reasonably perform. A judicial opinion found in favor of the limited contract which embodies the "turnkey" concept.

A further legal implication is concerned with the board's retention of policy control. One cited violation revolved around the fact that the contractor, not the board, established the objectives for the programs to be implemented. There has been no court ruling, as yet, but the warning is there.

Recommendation XIV

If a system is considering entry into a performance contract with any organization, public or private, the board and its administrative staff should first establish, in specific

---

terms, the parameters of the services sought, and
the program objectives. These objectives should be
in measurable behavioral terms. Appropriate legal
steps should then be taken to obtain bids for these
services from qualified contractors.

The intent of OEO in sponsoring performance contracts is to stim­
ulate innovation in education. As stated previously, the focus of
these contracts has been upon areas which are traditional problems to
schools.

Some of the innovations noted in the literature have involved
teaching machines, programmed learning, new "classroom" settings,
expanded use of paraprofessionals, and greatly reduced student-teacher
ratios. Other approaches have been more "teacher oriented" and had to
do with motivational efforts and techniques of instruction, backed by
minor changes in material. (Supra, pp. 31-32)

If school systems find the techniques and materials used by the
contractor work, then the ideal termination of the contract is the
"turnkey" phase. In this phase the program is turned over to the
school as a continuing operation. Unless the successful program is
continued, and expanded into other schools or areas, the monies ex­
pended in the experiment are largely wasted.23 There are two prime
causes for successful experiments not being continued by school systems.

23 Reed Martin and Charles Blaschke, "Contracting for Educational
First, the cost may be excessive in terms of what the schools can, or the taxpayer will, afford. If the per-pupil-cost of the program is greatly in excess of district averages, then the cost-effectiveness in terms of educational output per dollar input may be low. Consideration must be given to all factors. Some per-pupil cost increase can be justified if there is marked change in performance output. If children who normally are poor achievers achieve better, then the increase may be justifiable. How much improvement justifies how much additional cost is a matter for the judgment of the board, and its constituency.

The second factor is, can and will the regular staff assume the program. Does the present staff have any special skills needed? Can these skills be feasibly attained in in-service programs? Will the present staff accept the program and its challenge? These questions all need to be considered, and made part of the contract planning. If teacher resentment is built because of the contract, it may be difficult to effect a smooth transition. (Supra, pp. 34-35)

Any plan for performance contracting should include careful consideration of the financial and personnel ramification of the turnkey phase.

Recommendation XV

Before any performance contract is entered, the implications of the turnkey phase should be considered. If it is believed impossible or impractical due to either fiscal or staff limitations,
to continue a program should it be found successful, then alternate programs with less limited possibilities should be explored.

The essence of performance contracts is in the word performance. Unless the contractor can give evidence that he has fulfilled his agreed end of the bargain there can be no payment, or the payment is scaled down in accordance with the contract terms.

The question of evaluation becomes paramount. Who is to determine if the objectives have been met, or how much is to be paid? Obviously the contractor should not be the evaluator also. He has a vested interest in the outcome, and could not therefore be objective. The school system, too, may not have the capability of an effective evaluation, and also may be subject to bias. The obvious answer, as evidenced in the literature, is to contract the evaluative function to an independent evaluator.24

For the evaluation to be effective, the terms of the contract must be drawn so that the objectives are clear and measurable. This is a further emphasis upon the statement of all objective being in measurable behavior terms. The contract should include the specifications for the evaluation, and, as noted in earlier general recommendations, the instruments used for evaluation should match the objectives in terms of units of measure.

24Wildausky, op. cit.
Recommendation XVI

A performance contract should also include specifications for the evaluation of performance. These specifications should provide for an independent audit of performance and for the instrument or instruments to be used.

Since there are legal implications in the unspecified or unlimited contracting for educational services, there should be included in the contract the duration of services. In the original Texarkana project, there was the provision for a re-evaluation of the contract at specific times. This type of provision enhances the board's position of control, and lessens the possibility of a claim that the board is relinquishing its policy prerogatives. Embodied in the contract should be a termination phase. This "turnkey" provision should enable the board to assume the ongoing programs with the minimum of disruption.

Recommendation XVII

If the contract is for an extended period of time, for example, more than a single year, the time should be specified and a termination or turnkey date set.

It would be well to also include review points whereat the contract could be reviewed, renewed, terminated, or renegotiated.

---

25 Martin and Blaschke, op. cit.
The terms of the contract may contain any number of variations in the formula used to determine payment. The literature indicated instances where it would be virtually impossible for the contractor to lose. (Supra, pp. 67-68) The original Texarkana planning called for a penalty should students' registered gains not remain after six months. The finalized version of the contract omitted this clause.

Some contract payments are based upon gains by individual students. Others make compensation dependent upon group progress. There are contracts where the contractor profits even though the recorded performance gains are no greater than those registered by other similar students in the usual school setting.

The intent of performance contracting is to develop new improved methods of meeting student needs, so that student learning may be improved. The contractor should not benefit if the students show no better progress toward meeting the stated objectives under the "innovative" program than under the traditional approach.

In all performance contracts, payments involve measurement of gain or student growth. This is the key issue in the contract. The use of tests to measure gain should be contrasted with the use of tests to measure present status. The latter involves the determination of the level at which the student or group of students is performing. Traditionally the most typical use of standardized achievement tests has been for this purpose. Most of the theories about constructing and interpreting tests, as well as most of the statistical data about tests,
has assumed that the tests will be used to identify present status.

Now the demand is for standardized achievement tests to measure growth. When tests are used for other purposes they are accompanied by all the problems attendant to the measurement of status. The test must be selected which has high content validity, it must be carefully administered with attention given to instructions and timing, and care must be taken in scoring and converting raw scores to the desired derived score.

When tests are used to measure gain, additional problems arise which are not encountered in the measurement of status, and ones which are seldom recognized except by those cognizant of the statistical implication. Special consideration must be given to the meaning of gain or growth.

Almost universally, the literature and the survey noted, achievement was expressed in terms of grade equivalent. Normal or average gain is defined as one month of increase in grade equivalent scores for each month of instruction. For example, if a program is in effect for five months, the students are expected to show five months (0.5) of growth in CE units.

The normative process has established a growth rate in grade equivalent scores based upon the average gain for all students in the

---

26Hogan, op. cit.


28Ibid., p. 16.
sample. Although approximately one half of the students scored above, and one half below this average, there is the implicit assumption that each child should register this average rate of gain.

Another measure used by many systems to report achievement, and suggested by Hogan\textsuperscript{29} is the percentile rank. This type of score shows the students' relative positions compared to the normative population. If the students were to retain the relative position, year to year, they may be considered to be showing normal growth. An example of this would be a student who scores at the 40th percentile in reading at the beginning of grade five and again scores at the 40th percentile at the beginning of grade six. If, on the other hand, a student scored at the 40th percentile at the beginning of grade five, but scores only at the 30th percentile at the beginning of grade six, he would be considered as having shown less than the expected amount of learning, therefore less than normal growth.

The following illustration presents a comparison between grade equivalent scores and projected growth based upon percentile rank on a commonly used reading test.

In the illustration it can be seen that students scoring at the 10th percentile at the beginning of grade five, and who maintained this same relative position at the beginning of grade six, would have shown five months of gain on the grade equivalent scale. The students

\textsuperscript{29}Hogan, \textit{op. cit.}
Figure 1.—Illustration of differential rates of growth in terms of grade equivalent units for different percentile positions on the California Test of Basic Skills reading test.*

*Examiners' Manual, CTBS, Levels 1, 2, and 3, (Monterey, California: California Test Bureau), 1968.

on the 50th percentile curve at the beginning of both grade levels would have progressed one full grade equivalent. The consistent 90th percentile students, on the other hand, would have scored grade equivalent gains of fifteen months, or 1.5 GE units.

Recommendation XVII

Any performance contract should carefully consider the group to be served by the program and what level of performance is made the basis for payment. The
expectancy level for those students should be viewed in relation to the "promised" achievement so that the cost-effectiveness of the gains may be reviewed. The past achievement of students may be projected, the past gains of similar students may be established as the base, or a control group could be established, against which to judge the contract group's gains.

Some payments are specified in the contracts as being based on group gains while some are based on individual gains. There are several statistical phenomenon which should be given consideration in the establishment of contractual terms. No attempt will be made here to present these in detail, but merely to give recognition to the concepts.

All test publishers provide, in their technical manuals, the standard error inherent in their instruments. These error components may well represent, as noted in Chapter II, the equal of over one grade equivalent at the extremes of the test, and several months of grade equivalency even at the middle ranges. Since many of the contract programs are aimed at populations which would normally score at the lower extreme, the full year grade equivalent error becomes a matter of considerable concern.

If the contractor is paid for individual measured gains, but not penalized for measured losses, the error concept in itself could result in payment, even with no intervention on the part of the contractor.
If the payment is based upon group gains, then the errors of measurement tend to cancel each other, so are less beneficial to the contractor.

Another statistical concept which comes into play is regression toward the mean. This concept holds that scores at either extreme tend, upon repeated measure, to move toward the mean, or middle. Again, if a total group of students is involved, the two extremes, each moving toward the middle would tend to cancel each other. Significantly, however, most contracts deal with specific groups within a school, or with specific populations within a system, and further, these populations typically represent the low achievers at the lower extreme. With group measures, any overall regression would be expected to move upward, or favor the contractor.30

These statistical concepts have intentionally been oversimplified, with the aim only to make a point that whether dealing with individual or group measures of achievement, it is possible for certain factors other than innovative and effective instruction to work to the benefit of a contractor.

Recommendation XIX

In reviewing the implications in any proposed performance contract, careful consideration should be given to the statistical factors which may unduly favor the contractor. Payment for "chance" successes should be

offset by adjustment for measured loss. There are statistical adjustments for error and regression, which, though they can never eliminate inequities, can reduce the probability of unfair bias, either for or against the school system or contractor.

**Criterion-Referenced Tests**

Eighty three per cent of the directors surveyed felt there was definite need for criterion-referenced measures of school achievement. The literature contained many references to such instruments, including the fact that many OEO-sponsored performance contracts were to base payment upon criterion-referenced measures. (Supra pp. 61-62)

Major test publishers are giving serious considerations to criterion-referenced tests, although designing such measures which would have wide applicability poses severe problems. The item bank approach seems to be one method of utilizing the item building expertise of the publisher, yet maintaining the flexibility to serve local needs.

The Federal and other specially funded programs are providing the impetus in most systems now using criterion measures, so, as in the case of accountability itself, from these beginnings may come the foundation for criterion-referenced measures. Whether the system is entering into performance contracting, or is merely seeking to better evaluate the success of its own program improvement efforts, criterion-referenced testing may be a partial answer.
Recommendation XX

Those responsible for instruction and testing within systems should carefully monitor the results of experiments now being conducted by OEO involving criterion-referenced tests. Test publishers will vie for "first" in the criterion-referenced field, but such tests should be carefully screened for "fit" with system objectives. If the school district has its own in-house staff with the expertise to develop test items based upon stated objectives, consideration should be given to criterion-referenced instrument construction in a critical areas, at least on a pilot basis.

The recommendations stated above were based upon data from several sources. The experiences of several systems involved in some facet of accountability were related in the literature. A number of authorities in the fields of testing, school administration, and educational research stated their positions, also in the literature. The experiences and recommendations of the thirty-eight large city test directors were solicited, and the composite of the best thinkings from these various sources formed the basis of the recommendations.

The statements made in the recommendation were intentionally broad. Each system has its own unique characteristics. Plans developed within one system would need careful study, and probably extensive
retailoring before it could be adapted to another. The intent here was to provide broad generalities which would serve to call attention to potential problem areas, alert those responsible for testing to some important ramifications of accountability, and to stimulate the thinking of school officials faced now or in the future with pressures for accountability.

CRITIQUE OF RECOMMENDATIONS

The next step in the study was to present these recommendations to a panel of individuals each with specific and unique qualifications in the areas of testing and educational research. The complete listing of these individuals and their positions is in Appendix A, page 184. Four are university professors, although one is on leave of absence to direct the National Assessment Program. Another was formerly director of research in a major city's public schools, while still another was formerly a state supervisor of measurement and evaluation.

Of the remaining four individuals, two represent major text and test publishers, one a private research and testing company, and the other is the director of research and evaluation in a major urban metropolitan area not included in the basic survey. Two of the eight evaluators were presented the recommendations by mail along with a request for their comments. The remaining six were queried as to their opinion in personal interviews.
General Recommendations

All of the evaluators were in agreement with the intent of the first recommendation, which concerned community involvement in planning school goals and objectives. Mazur expressed concern lest this involvement be given lip service, and felt it might not be possible in a meaningful way. Wysong agreed to the necessity of community involvement, but felt that school systems have a long way to go to achieve this goal. He questioned, "are we really working toward this goal?" Lennon noted that there existed some constraints upon freedom of choice. State requirements of courses of study, approved texts, and accreditation agencies such as North Central, all fix parameters upon freedom. But, he stated, there was need to stimulate thought about goals and priorities in elective and affective areas, and within the constraints posed by the institutions.

There was a consensus, also, on the second recommendation, regarding the establishment of behavioral objectives. Thiagarajan admits that defining objectives, much less stating them behaviorally, is most difficult, though an absolute necessity. Walker cited the difficulty that the Behavioral Objective Exchange in California has in finding worthy items. Even though teachers and systems submit their best efforts, over 90 per cent of the items are found to be useless by virtue of faulty statement or lack of validity. This same difficulty which occurred in simply constructing test items would, he proposed, be found in defining broader objectives in appropriate form. Smallenburg referred to the state selection of tests as being an effort to
establish objectives in a somewhat reverse order, because many smaller districts were without well stated guidelines. Lennon said that the more precise thinking needed to formalize objectives would not only aid program planning, but would enable system personnel to select texts and test instruments better suited to the programs.

Recommendation three related to the need for an evaluation component. There was no disagreement with this concept, although both Mazur and Anderson expressed agreement on the need for an external audit or verification, particularly of experimental or specially funded programs. This audit agency could perform the actual evaluation, or merely verify the evaluation. Wysong agreed that such a verification would increase the credibility of the evaluation. Lennon expressed concern that much program evaluation was an afterthought, and that an inhouse evaluation component should be better able to provide the expertise to design the evaluation concurrently with the design of the program. He also noted that there was substantial feeling that whatever a school system might do in its own self-evaluation endeavor, independent auditors are desirable.

The selection of instruments was the subject of Recommendation four. There was complete consensus as to the need for careful instrumentation. Walker and Thiagarajan emphasized the need to carefully compare properly stated objectives with the projected tests to insure that content validity was high. Womer cited the experience of National Assessment evaluators reaching agreement as to the appropriateness of
items versus objectives. Mazur noted the difficulty of measuring in the affective areas, but stated that scale measures, which might not have ratios or intervals, could be employed. Though not norm oriented, they could reveal change.

The fifth recommendation concerned the inclusion of demographic data, along with achievement scores, in the accountability releases. There was no disagreement as to the general philosophy, however, there were several notes of caution about specific data. Wysong feared the inclusion of academic aptitude data would be publically received as meaning innate capacity—a fear shared by Walker, who stated that the public was no longer accepting the belief that low IQ was a valid excuse for schools not teaching children. Smallenburg echoed this fear, and noted the California legal hassel over even individual psychological examinations and their suspected lack of validity for certain individuals. Lennon referred to the Barro and Dyer models (See Supra, pp.21-22) which take into account "other elements" which affect achievement. Any of these which need to be understood or are amenable to change, either by the school or the community should be brought into the light of day. Other school and community items were given no special notice by the evaluators, except the consensus that each district should provide whichever data could be made meaningful to its publics.

All agreed with the sixth recommendation, that accountability should be at more than a single point in time to be most meaningful. Womer felt that at least three points should be included, at elementary,
junior, and senior high levels. National Assessment has this, in essence, with age criteria rather than strict grade level criteria. Smallenburg pointed to the five levels mandated in California, and Wysong noted with interest House Bill No. 105 being introduced in the Ohio legislature, which would establish reading check points at grades three, seven and ten. (See Appendix C, p. 207.)

There was no disagreement or notable comment other than concurrence on the seventh recommendation, regarding the exclusion of special category students from testing or accountability.

Community Involvement

There was general agreement on numbers eight, nine, and ten, regarding the involvement of the community, news media, and school staff, respectively, in the evaluation and release phases. Anderson commented on the media's finding innovation or progressiveness in education so newsworthy. Lennon's remarks were in agreement. He stated that accountability is newsworthy because it "is an idea whose time has come."

Wysong felt that meaningful community involvement was a long way off, and that it would be a slow, developmental process. Smallenburg noted that there would, for some time, be conflict of interests in communities with divergent cultural groups, which would be difficult to resolve. Walker stated that community involvement meant taking risks which many school people are not eager to take.

Mazur indicated that teachers and teacher organizations would accept the challenge of involvement in accountability— if they were
given support in the means by which to accomplish objectives. Wysong agreed that many teachers will accept a viable role in accountability, but not all.

The format of accountability releases was the subject of the eleventh recommendation, and elicited no special comment other than agreement with the philosophy. Walker simply reinforced the note that complexity must be kept to the level of public understanding and acceptance.

Decentralization

Decentralization as a means of meaningful community involvement was covered in Recommendation twelve. Womer doubted the ability of many top school officials to live with the ambiguity posed by decentralization of any real authority. Wysong reiterated his concern for the local communities' being ready for effective involvement. Smallenburg felt the need for increasing the responsiveness of schools through placing controls "closer home" to the community. Mazur and Lennon both spoke strongly to the belief that most decentralization today is not effective attunement to the local community, but for administrative convenience.

The next recommendation cautioned against not retaining some measure of central control, especially over evaluation. There was complete agreement that such control was necessary if sound accountability is to be maintained. Lennon, particularly, noted that the difficulty of effective evaluation was enough with a trained evaluation
staff, but if the efforts were fragmented, the evaluations would lose a substantial measure of validity and comparability.

**Performance Contracting**

Performance contracts aroused special responses from several of the evaluators. Thiagarajan expressed concern over the legal implications, having some particular interest in the Gary, Indiana experiment. He particularly decried the failure to develop specific objectives for the contractor. Walker noted that publishers of texts and other materials were being asked to accept what are in essence "performance contracts" for these materials. Simply place the materials on consignment, they are being told, and if the students perform better, the district will pay for the materials. Lennon made a similar observation, and continued that it is inappropriate for schools to seek guarantees of performance for the materials, since the surveyors have little control over the way in which materials and services are used.

Lennon further warned against the "low-risk approach" appeal of the performance contract. Low financial risk because of guarantee or penalty clauses, and low political risk because failure could be imputed to the contractor.

Anderson picturesquely referred to performance contracts as the "hot pants of education." As her specific responsibility of late has been as an auditor of performance contract evaluations, and as she was the one to "audit the audit" of the controversial Texarkana project,
Anderson spoke strongly in support of the need for external audit as recommended in item sixteen, though there was agreement by all. Her further statement was that the audit became more external in direct ratio to the distance of the source of funds. The farther away from the school system the source of funds, the more external the audit was likely to be.

Lennon's final note on performance contracting was that he hoped that future contracts would be of longer duration than some part or single year contracts now under consideration—though a turnkey provision was essential. Greater attention, he continued, needed to be paid to the reliability of initial and final measures, and the selection of evaluative instruments must receive far more searching attention prior to writing the contract that he thought had been true heretofore.

With respect to the statistical implications for measurement of gain, there was support for the notes of caution to prospective signatories of performance contracts. All commented upon the probability that the grade equivalent standard would remain, although both Wysong and Lennon cautioned specifically against its use for individual data interpretation. Lennon continued, noting one contract which called for "125 per cent of normal gain" in terms of grade equivalent. This, he stated, is meaningless, as grade equivalent scales are notoriously "un-equal unit scales" and are not on any ratio. GE gain measures, he concluded, are most appropriate as group measures on the elementary level. Even group GE measures are less valid at secondary levels on basic skills, as they are largely extrapolations of learning curves.
Wysong sounded the strongest note of disagreement in the total response, and this was against recommendation fifteen. While others agreed that it would be largely unwise for a school system to enter into a contract involving programs with unrealistic chances of being continued, Wysong spoke to the contrary. Such contracts could help specific children during the time of the program, and such limitations could retard new developments in the school.

Recommendation eighteen, concerning a base of expectancy in performance contracting, received general accord, but Lennon, Walker, and Mazur questioned the limitation to the performance contract. Not only are performance contracts and Federal programs important for precise measurement, but local program evaluations also. All programs, they independently stated, need to consider what is expected of children before judgements are made. Walker expressed concern lest the tendency to expect uniform learning from all be the basis of teacher rejection of the whole concept of accountability. Lennon commented that we have no common "ergs of education" by which to measure gain. Unlike the manufacturer, education's raw materials are infinitely varied, and the product is not of unvarying sameness as in the manufacturing operation, he concluded, but one with its initial richness and variety enhanced and multiplied.

The nineteenth recommendation concerned statistical phenomenen, and received the same trio's warning against the limitation to performance contracts. There was the further caution by Walker that statistical
adjustments, however accurate, if applied to public accountability could be viewed as "statistical gerrymandering." Keep accounting simple, yet defensible, he concluded.

The final recommendation concerned the use of criterion-referenced tests. Here too, there was general concurrence. Smallenburg felt that there is a need for test publishers to enter the field, perhaps via the item bank route. Thiagarajan cites the use of programmed learning techniques based almost exclusively on criterion-referenced measures. Walker felt that schools had almost gone full circle back to teacher made criterion tests, but stated that publishers were moving rapidly toward tests to accompany texts, thus criterion-referenced instruments could be devised for use by any system adopting the text. He further stated that any publisher soon to come out with "universal" criterion-referenced tests was "naive, didn't care, or (was) an outright fraud." Wysong favored criterion measures, but felt that school staffs could make better use of their time, so that systems should look to outside sources.

Mazur also expressed the opinion that publishers may have the answer in gearing tests to texts. School staffs could do the job, but as in any specialized field, would require appropriate in-service training. Lennon stated that there are strong arguments to support criterion-referenced measures as more valid measures of certain outcomes, yet how they can be translated into units that would yield measures of gain or growth is not clear. He saw no either/or to the
criterion versus norm referenced question; both are useful and needed in appropriate situations.

**SUMMARY**

Chapter I posed the problems to which this dissertation was addressed, stated the purpose of the study, and presented a definition of terms. Assumptions made were stated, and the limitations to the study were outlined. The significance of the study was posited, the design for the study set forth, and finally, the organization of the dissertation presented.

The second chapter detailed the findings of the review of literature and research. Several definitive areas were explored which held implications for testing and evaluation. First, there were reviewed the general aspects of the rise of educational accountability; some causal factors, some limitations, and some challenges.

Next were explored some of the emerging innovations in organization or program development which have come about largely as a product of accountability. These all hold direct implications for testing and evaluation. Some of these are decentralization for community responsiveness, performance contracting, and voucher systems.

Finally, there was a review of the literature related to testing and evaluation as it pertains to accountability and the manifestations noted above.

Chapter III concerned the development of the questionnaire which was submitted to thirty-eight directors of testing in the largest
school systems of the country. The rationale behind the questions was established. These questions covered the areas noted in the literature as they pertained both to cities which had already released publically school-by-school test information, and those which had not as yet done so.

The fourth chapter summarized the replies to the questionnaire and noted the added comments many directors made concerning various aspects of the questions. Such added comments were encouraged, to elicit from the directors special matters of concern pertaining to their school systems.

The fifth chapter brought together the findings in the literature and the experiences or considered judgements of the test directors, to generate some recommendations to be used as guidelines by any system entering into the era of accountability. These recommendations were of necessity broad, so as to have general applicability or adaptability. While it might seem that some recommendations are statements of the obvious, the literature reviewed in Chapter two, and the responses summarized in Chapter four revealed instances of policy or practice not in line with those recommended.

The recommendations were then presented for the evaluation of eight authorities in the fields of educational testing, research or programming. Their judgement as to the validity, applicability and feasibility of the recommendations was sought. In light of their critiques, several amendments were made, or limitations either imposed
or removed. The recommendations as revised are herewith concisely presented.

**AMENDED RECOMMENDATIONS**

Recommendation I

Viable communications must be established to gain input from all segments of the educational community. This should include the establishing of realistic educational goals and objectives. This must involve a developmental program of community education. Goals and objectives may of necessity be varied or modified by subdistrict or by school. Communication may be accomplished, as locally feasible, by central committees, by local advisory groups, by complete decentralization, or by some combination of these approaches.

Recommendation II

After the educational goals of the district, sub-district, or school have been determined, the entire educational program pertaining to both the academic and socio-personal development of students must be reduced to a series of specific, measurable, behaviorally-stated objectives.

Recommendation III

An effective evaluation component must be established, either within or without the system, whether integral, contracted or shared. This component can then certify to the school and the community as to the effectiveness
of the programs in attaining the stated objectives. The evaluators should be sufficiently divorced from the programs themselves so that a reasonable degree of objectivity may be obtained.

Recommendation IV

Selection of evaluative instruments or tools must be made with the utmost care. Program content and the stated behavioral objectives must be carefully examined, and equated with the available instruments in order to obtain ones with the best possible fit. Not only must the academic areas be so treated, but all possible efforts must be made to evaluate objectives within the affective domain.

Recommendation V

Each factor which is thought to have a bearing on student achievement, whether school or community related, should be considered judgement of those preparing the data, that the release of the specific data can be turned to a positive end to bring school and/or community attention to bear with the view of improving or correcting the condition, it could be an item worthy of inclusion.

Recommendation VI

Accountability is not just for visibility but must serve a purpose. The purpose is to bring attention to any need for
change. To this end the school should be accountable for performance at each major juncture in the total school program, and at as many intermediate points as it is felt the data could be meaningfully employed.

Recommendation VII

EMR and other special students may or may not be included in the testing program, according to building or program conditions. If the content of their program is not congruent to the test content, then the data may be validly excluded from inclusion in school-by-school or system summary data.

Recommendation VIII

All segments of the community should be included in the formulation of school objectives, and should be involved in relating school and community factors to school performance. Therefore, these same segments should be represented in the evaluation stages so that they may help interpret the data to their local community. This is an important component of a continuing program of community education.

Recommendation IX

The representatives of the various news media should be included in the planning for accountability. If they are
fully aware of the meaning and intent of all data included, the "power of the press" may be an invaluable aid in working for the equitable solution of school and community problems related to education. The press is an important element in community education.

Recommendation X

Teachers, principals, and counselors should be given a voice in formulating general and local objectives, both cognitive and affective. Their knowledge of the local school community and familiarity with local problems can be invaluable in helping to state local needs. They must further be involved in the evaluation and release stages, for understanding will both reduce threat, and make them more effective ambassadors to the local community.

Recommendation XI

The release format and reporting units should be planned at the onset of the program of accountability. The format should be kept as simple as is consistent with the data to be presented, and should report in the measurement terms used to state the behavioral objective. Involvement of the various publics at all stages of planning and release will maximize the understanding and therefore acceptance of the release as a sincere and credible report of accountability.
Recommendation XII

Decentralization should be considered as a means of meeting local schools needs. If the plan does not provide for locally elected boards, there must be provision for broader flexibility in local policy and program, and the local community must be involved in a meaningful way in planning and decision making.

Recommendation XIII

Though the system may decentralize, certain elements of central control must remain. One of these is the component of evaluation. Accountability dictates that the public be kept aware of the results of program change or innovation, and the effectiveness of any investment in educational reform. The data base, for comparative purposes, must maintain consistency through the system, though special or different instruments may be needed for selective intra-subdistrict or intra-school evaluation.

Recommendation XIV

If a system is considering entry into a performance contract with any organization, public or private, the board of education and its hired administrative staff should first establish, in specific terms, the parameters of the services sought, and the program objectives. These objectives should be stated in measurable behavioral terms. Appropriate legal
steps should then be taken to obtain bids for these services
from qualified contractors, and to assure the retention of
control by the board.

Recommendation XV

Before any performance contract is entered, the implications
of the turnkey phase should be considered. If it is believed
impossible or impractical, due to either fiscal or staff
limitations, to continue a program should it be found
successful, then the value of the program to the specific
students participating should be weighed against the
possible value of alternate programs.

Recommendation XVI

A performance contract should also include specifications
for the evaluation of performance. These specifications
should name the evaluator, provide for an audit of the
evaluation, and specify the instruments or type of
instruments to be used.

Recommendation XVII

If the contract is for an extended period of time, for
example, more than a single year, the time should be
specified and a termination or turnkey date set. It
would be well to include also review points whereat
the contract could be reviewed, renewed, terminated,
or renegotiated.
Recommendation XVIII

Any program evaluation, including a performance contract, should carefully consider the group to be served by the program and what level of performance is made on the basis for program success or contractor payment. The expectancy level for the students should be viewed in relation to the anticipated achievement so that the cost-effectiveness of the gains may be reviewed. The past achievement of students may be projected, the past gains of similar students may be established as the base, or a control group could be established against which to judge the contract or program group's gains.

Recommendation XIX

In reviewing the implications in any proposed performance contract or program evaluation, careful consideration should be given to the statistical factors which may unduly influence the judgements. "Change" successes should be offset by adjustment for measured loss. There are statistical adjustments for error and regression which, though they can never eliminate inequities, can reduce the probability of unfair bias, either for or against a school system or a performance contractor.
Recommendation XX

Those responsible for instruction and testing within systems should carefully monitor the results of experiments now being conducted by OEO involving criterion-referenced tests. Test publishers will vie for "first" in the criterion-referenced field, but such tests should be carefully screened for "fit" with system objectives. If the school district has its own inhouse staff with the expertise to develop test items based upon stated objectives, consideration might be given to criterion-referenced instrument construction in critical areas, at least on a pilot basis.

CONCLUSIONS

The original hypothesis was that, due to the newness of the emerging concept of educational accountability, there were few guidelines for school districts coming under public pressure to give an accounting for their educational effectiveness. The research of literature largely confirmed this belief. There were many articles dealing with one or another facet of the major theme, and there were varying views with respect to many of these. There was lacking any concerted effort to pull all the fragmentary aspects of accountability together into a unified whole which could give one a full perspective.
A major component not found in the literature, with only one or two minor exceptions, was the reaction to accountability and school-by-school data releases as seen by those who were in the forefront of the activity itself—the test directors in those cities already committed to public accountability. Nor did the literature present the thinking of those test directors who would soon be faced with a part in the determination of policy in systems as yet uncommitted to accountability. This study added that important element.

The attempt has been to bring together the thinking of the theorists and the practices of the practitioners, and so to develop guidelines which will be usable by the educators in those systems which have as yet to enter into the arena of public accountability.

The newness of the present concept of educational accountability prevents making any firm conclusions at this time, only recommendations in light of experiences to date. Research is being conducted in the areas of performance contracting, program development, testing, and the like as was noted in Chapter two. The findings may well have implications for future accountability, so the educator must keep a close watch on changing developments. Education today is facing a grim challenge—prove its effectiveness, or make way for change. Testing and evaluation will play a vital role—either in furnishing the proof, or evaluating the change.
APPENDIX A

ROSTER OF LARGE CITY TEST DIRECTORS

PARTICIPATING IN THE SURVEY

SURVEY QUESTIONNAIRE

ROSTER OF EVALUATORS OF

THE RECOMMENDATIONS

175
<table>
<thead>
<tr>
<th>CITY</th>
<th>TEST DIRECTOR</th>
<th>TITLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlanta</td>
<td>Dr. Hugh F. Moss</td>
<td>Director of Guidance and Testing</td>
</tr>
<tr>
<td>Baltimore</td>
<td>Dr. Paul Yaffe</td>
<td>Director, Bureau of Educational Testing</td>
</tr>
<tr>
<td>Chicago</td>
<td>Dr. Elmer Casey</td>
<td>Director, Evaluation and Pupil Services</td>
</tr>
<tr>
<td>Cincinnati</td>
<td>Miss Joan Bollenbacher</td>
<td>Director, Evaluation Services</td>
</tr>
<tr>
<td>Clearwater</td>
<td>Mr. Don Lambert</td>
<td>Supervisor of Evaluation Services</td>
</tr>
<tr>
<td>Cleveland</td>
<td>Dr. Margaret Fleming</td>
<td>Director, Division of Research and Development</td>
</tr>
<tr>
<td>Columbus</td>
<td>Mr. William D. LeSage</td>
<td>Supervisor of Testing and Evaluation</td>
</tr>
<tr>
<td>Dallas</td>
<td>Dr. Ruby Morris</td>
<td>Director, Research and Psychological Services</td>
</tr>
<tr>
<td>Denver</td>
<td>Mr. Gerald P. Cavanaugh</td>
<td>Supervisor, Research and Testing</td>
</tr>
<tr>
<td>Detroit</td>
<td>Dr. Robert S. Lankton</td>
<td>Divisional Director, Department of Research and Development</td>
</tr>
<tr>
<td>Honolulu</td>
<td>Miss Beatrice Loui</td>
<td>Staff Specialist, Testing</td>
</tr>
<tr>
<td>Houston</td>
<td>Mr. Tommy G. Hall</td>
<td>Coordinator, Group Testing</td>
</tr>
<tr>
<td>Indianapolis</td>
<td>Dr. Paul F. Brown</td>
<td>Supervisor, Testing and Evaluation</td>
</tr>
<tr>
<td>Jacksonville</td>
<td>Mr. Coke L. Barr</td>
<td>Supervisor, Secondary Tests of Measurements</td>
</tr>
<tr>
<td>Kansas City</td>
<td>Mrs. Elizabeth Orahood</td>
<td>Supervisor of Educational Testing</td>
</tr>
<tr>
<td>CITY</td>
<td>TEST DIRECTOR</td>
<td>TITLE</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------------------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>Las Vegas</td>
<td>Mr. Emory Robertson</td>
<td>Research Director</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>Dr. Howard A. Bowman</td>
<td>Director, Measurement and Evaluation</td>
</tr>
<tr>
<td>Memphis</td>
<td>Mr. Wallace A. Wilson</td>
<td>Director of Guidance and Pupil Adjustment</td>
</tr>
<tr>
<td>Miami</td>
<td>Mr. George Loiselle</td>
<td>Supervisor of Testing</td>
</tr>
<tr>
<td>Milwaukee</td>
<td>Mr. Elfred Bloedel</td>
<td>Supervisor of Testing Services</td>
</tr>
<tr>
<td>Minneapolis</td>
<td>Dr. Ralph H. Johnson</td>
<td>Director of Guidance Services</td>
</tr>
<tr>
<td>Mobile</td>
<td>Mr. Floyd Replogle</td>
<td>School Psychologist</td>
</tr>
<tr>
<td>Nashville</td>
<td>Dr. Ed Binkley</td>
<td>Director of Educational Research</td>
</tr>
<tr>
<td>New Orleans</td>
<td>Miss Anna B. Klein</td>
<td>Director of Guidance and Testing</td>
</tr>
<tr>
<td>New York</td>
<td>Dr. Samuel H. McClelland</td>
<td>Acting Director, Bureau of Research</td>
</tr>
<tr>
<td>Oakland</td>
<td>Dr. Alden W. Badal</td>
<td>Director of Research</td>
</tr>
<tr>
<td>Oklahoma City</td>
<td>Dr. Frances Peters</td>
<td>Coordinator of Assessment and Placement</td>
</tr>
<tr>
<td>Orlando</td>
<td>Mr. Royce B. Walden</td>
<td>Director of Testing and Research</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>Dr. Jules Grosswald</td>
<td>Director, Division of Testing</td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>Mr. Melvin N. Vesely</td>
<td>Associate Director, Testing and Statistics</td>
</tr>
<tr>
<td>Portland</td>
<td>Dr. George Ingebo</td>
<td>Evaluation Administrator</td>
</tr>
<tr>
<td>St. Louis</td>
<td>Dr. Clem A. Powers</td>
<td>Assistant Superintendent</td>
</tr>
<tr>
<td>San Diego</td>
<td>Mr. William H. Vogler</td>
<td>Director of Testing Services</td>
</tr>
<tr>
<td>San Francisco</td>
<td>Mr. Lawrence L. Lew</td>
<td>Research Assistant</td>
</tr>
<tr>
<td>CITY</td>
<td>TEST DIRECTOR</td>
<td>TITLE</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------------</td>
<td>---------------------------------------------------------</td>
</tr>
<tr>
<td>Seattle</td>
<td>Mr. William J. Lagrajd</td>
<td>Coordinator, Group Testing</td>
</tr>
<tr>
<td>Tulsa</td>
<td>Dr. Mary Jo Keatley</td>
<td>Director, Psychological and Testing Department</td>
</tr>
<tr>
<td>Washington, D.C.</td>
<td>Dr. Wilbur A. Millard</td>
<td>Assistant Superintendent, Department of Pupil Personnel Services</td>
</tr>
<tr>
<td>Wichita</td>
<td>Mr. Corwin Bare</td>
<td>Director, Pupil Services Division</td>
</tr>
</tbody>
</table>
SURVEY - LARGE CITY TEST DIRECTIONS

CITY _______________________

I 1. Has your system established a practice of releasing standardized test results to the public on a school-by-school basis?
   YES ____  NO ____

II If the answer to the above question was YES, please respond to the following questions in Section II. If the answer was NO, please turn to Section III on page 3. All please respond to Section IV on pages 3 and 4.

2. Achievement test results are released for grade(s):
   1 ____  4 ____  7 ____  10 ____
   2 ____  5 ____  8 ____  11 ____
   3 ____  6 ____  9 ____  12 ____

3. Achievement is reported in terms of:
   Grade Equivalents ____  Stanines ____
   Percentiles ____  Standard Scores ____
   Other ____________________________

4. Achievement test scores are based upon:
   National Norms ____  State Norms ____
   Local Norms ____  Other ____

5. Summary intelligence or academic aptitude test results are released concurrently.
   YES ____  NO ____

6. Intelligence or academic aptitude test results, if released, are reported in terms of:
   I.Q. ____  Stanine ____
   Percentile ____  Standard Scores ____
7. What other demographic data, if any, is also presented concerning the school or school community?

   Mobility rate  ____  Student/teacher ratio  ____
   Attendance data  ____  Economic data (ADC, etc.)  ____
   Teacher preparation  ____  Parents educational level  ____
   Teacher experience  ____  Racial composition  ____

   Other ____________________________________________

8. Which, if any, of these data which you do not currently report do you feel might lead to a better public understanding of the students educational setting?

   Please list ____________________________________________

9. Prior to your system's first release of test data, was in-service and/or detailed briefings held for:

   News Media  YES ____  NO ____
   Administrators  YES ____  NO ____
   Teachers  YES ____  NO ____
   Significant community representatives  YES ____  NO ____

   Others ____________________________________________

10. Do you feel that these programs were, (or could have been) of value?

    Little or no value ____  Some Value ____  Great Value ____

11. Was first year treatment of the release by news media:

    1. Factual and unbiased ____
    2. Factual but sensationalized ____
    3. Factual but supportive of schools ____
    4. Distorted or out of context ____
    5. Other ____________________________________________
12. Were subsequent year's releases treated by the news media:

More positively ____ The same ____ More negatively ____

13. Do you feel that individual school staffs feel threatened by their implied evaluation in the release of school-by-school test data?

NO ____ Some ____ Strongly ____ Don't Know ____

14. Have the communities of those schools reflecting the poorer performance been:

Completely negative ____ Apathetic ____
Critical but passive ____ Critical but supportive ____
Other ___________________________________________

15. Do you feel there has been an increased pressure on smaller adjacent school systems for similar release of school-by-school test data?

16. Have any adjacent small systems followed your lead in the release of school-by-school data?

Yes ____ No ____

17. What, if any, categories of students are excluded from participation in city-wide testing and/or inclusion in summary test data, such as EMR? ____________________________________________________________

III

This section applies if your system has not as yet released school-by-school summary test data.

18. Is there pressure within your system's school community for the release of school-by-school standardized test data?

Little or none ____ Some ____ Strong ____

19. Is your system planning to release such summary test data?

Soon: ____ Probably later ____ No ____
20. What data do you feel you will release (if any)?

<table>
<thead>
<tr>
<th>Achievement data</th>
<th>Academic aptitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attendance data</td>
<td>Teacher preparation</td>
</tr>
<tr>
<td>Teacher mobility</td>
<td>Student mobility</td>
</tr>
<tr>
<td>Economic data (ADC, etc.)</td>
<td>Parental education</td>
</tr>
<tr>
<td>Racial composition</td>
<td>Student/teacher ratio</td>
</tr>
</tbody>
</table>

Other

21. Has there been any expressed opposition by staff to the implied evaluation of their performance should summary test data be released?

Little or none __ Some ___ Strong ____ Don't know __

IV

All Test Directors please complete this section:

22. Has your system adopted any plan of decentralization?

Yes ____ No ____

23. Have there been any pressures upon your system for decentralization as a means of making schools more responsive to local community needs?

None ____ Some ___ Strong ____

24. Has your system entered into any form of educational performance contracting?

Yes ____ No ____
25. Have there been any pressures within your system or community for entry into some form of educational performance contracting?

Little or none ____ Some ____ Strong ____

26. Has there been any perceptable shift of students toward parochial or private schools in your area, which you feel is attributable to lack of confidence in public school performance?

Little or none ____ Some ____ Considerable ____

27. Do you feel there is a need for criterion referenced testing as a better measure of achievement, particularly of small increments of performance?

Yes ____ No ____

28. Does your system use system wide criterion referenced tests?

None ____ Some ____ Extensively ____

29. Do you feel that some standardized, norm referenced testing will be needed and retained in your system for:

Public Accountability Yes ____ No ____
System Planning Yes ____ No ____
Student Counseling Yes ____ No ____
Staff Evaluation Yes ____ No ____
Other ________________________________

30. Does your local teachers organization (union or association) formally recognize the need for educational accountability?

Yes ____ No ____ Don't Know ____
### Reviewers of Recommendations

<table>
<thead>
<tr>
<th>Name</th>
<th>Position and Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr. Scarvia Anderson</td>
<td>Executive Director for Special Development, Educational Testing Service</td>
</tr>
<tr>
<td>Dr. Roger T. Lennon</td>
<td>Senior Vice-President, Harcourt Brace Jovanovich, President, Psychological Corporation</td>
</tr>
<tr>
<td>Dr. Joseph Mazur</td>
<td>Professor of Education, University of Southern Florida, Formerly Director of Research,</td>
</tr>
<tr>
<td></td>
<td>Cleveland Public Schools, Cleveland, Ohio</td>
</tr>
<tr>
<td>Dr. Harry Smallenburg</td>
<td>Director of Research and Pupil Services, Los Angeles County Schools, Los Angeles, California</td>
</tr>
<tr>
<td>Dr. Sivasailam Thisgarajan</td>
<td>Professor of Education, Indiana University, Bloomington, Indiana, Director of Program Development Laboratories</td>
</tr>
<tr>
<td>Dr. Robert N. Walker</td>
<td>Director of Test Division, Ginn and Company, Boston, Massachusetts, Formerly Director of Guidance and Testing, Akron, Ohio</td>
</tr>
<tr>
<td>Dr. Frank B. Womer</td>
<td>Professor of Education, University of Michigan (on leave), Director of National Assessment Program</td>
</tr>
<tr>
<td>Dr. Eugene Wysong</td>
<td>Professor of Education, University of Toledo, Formerly Supervisor of Measurement, Division of Guidance and Testing, State of Ohio, President-elect, Association for Measurement and Evaluation in Guidance</td>
</tr>
</tbody>
</table>
APPENDIX B

SAMPLES OF RELEASE FORMATS
OF LARGE CITIES RELEASING
SCHOOL-BY-SCHOOL TEST DATA
### RESULTS OF CITY-WIDE TESTING PROGRAM 1968-1969
CHICAGO PUBLIC SCHOOLS

#### School Entering First Year Readiness

<table>
<thead>
<tr>
<th>Grade</th>
<th>National</th>
<th>City</th>
<th>School</th>
<th>National</th>
<th>City</th>
<th>School</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>69</td>
<td>52</td>
<td></td>
<td>69</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Probably Not Ready</td>
<td>26</td>
<td>35</td>
<td></td>
<td>26</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>Not Ready</td>
<td>7</td>
<td>13</td>
<td></td>
<td>7</td>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>

#### Mid First Year School Learning Ability

<table>
<thead>
<tr>
<th>Grade</th>
<th>National</th>
<th>City</th>
<th>School</th>
<th>National</th>
<th>City</th>
<th>School</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>70</td>
<td>57</td>
<td></td>
<td>70</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>Probably Not Ready</td>
<td>26</td>
<td>35</td>
<td></td>
<td>26</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>Not Ready</td>
<td>7</td>
<td>13</td>
<td></td>
<td>7</td>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>

#### Explanations

For explanation and interpretation in addition to the following definitions of the symbols used, see the introduction.

- $N$ = number of pupils
- $X$ = percent
- $Q_1$ = point below which 25% of scores fall
- $Q_2$ = median or middle score; 50% of scores are higher; 50% are lower
- $Q_3$ = point below which 75% of scores fall
- $PR$ = percentile rank
- $GE$ = grade equivalent
- $IQ$ = school learning ability derived from intelligence tests

#### Grade Level, Number Tested, and Quantiles

<table>
<thead>
<tr>
<th>Grade</th>
<th>National</th>
<th>City</th>
<th>School</th>
<th>National</th>
<th>City</th>
<th>School</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Fourth Grade (PICT) April 1969</td>
<td>111</td>
<td>75</td>
<td>106</td>
<td>63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td>4.6</td>
<td>75</td>
<td>3.9</td>
<td>63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>3.5</td>
<td>50</td>
<td>3.2</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>2.9</td>
<td>25</td>
<td>2.7</td>
<td>18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade</th>
<th>National</th>
<th>City</th>
<th>School</th>
<th>National</th>
<th>City</th>
<th>School</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>106</td>
<td>63</td>
<td>106</td>
<td>63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>102</td>
<td>50</td>
<td>103</td>
<td>54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>95</td>
<td>25</td>
<td>95</td>
<td>26</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade</th>
<th>National</th>
<th>City</th>
<th>School</th>
<th>National</th>
<th>City</th>
<th>School</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Fourth Grade (PICT) March 1969</td>
<td>111</td>
<td>75</td>
<td>106</td>
<td>63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td>9.0</td>
<td>75</td>
<td>7.1</td>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>6.7</td>
<td>50</td>
<td>5.3</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>6.2</td>
<td>25</td>
<td>6.3</td>
<td>18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade</th>
<th>National</th>
<th>City</th>
<th>School</th>
<th>National</th>
<th>City</th>
<th>School</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>106</td>
<td>63</td>
<td>106</td>
<td>63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>102</td>
<td>50</td>
<td>103</td>
<td>54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>95</td>
<td>25</td>
<td>95</td>
<td>26</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade</th>
<th>National</th>
<th>City</th>
<th>School</th>
<th>National</th>
<th>City</th>
<th>School</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Fourth Grade (PICT) October 1969</td>
<td>111</td>
<td>75</td>
<td>106</td>
<td>63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td>10.2</td>
<td>75</td>
<td>7.1</td>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>6.2</td>
<td>50</td>
<td>6.0</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>5.8</td>
<td>25</td>
<td>5.0</td>
<td>12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade</th>
<th>National</th>
<th>City</th>
<th>School</th>
<th>National</th>
<th>City</th>
<th>School</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>106</td>
<td>63</td>
<td>106</td>
<td>63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>102</td>
<td>50</td>
<td>103</td>
<td>54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>95</td>
<td>25</td>
<td>95</td>
<td>26</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade</th>
<th>National</th>
<th>City</th>
<th>School</th>
<th>National</th>
<th>City</th>
<th>School</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>106</td>
<td>63</td>
<td>106</td>
<td>63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>102</td>
<td>50</td>
<td>103</td>
<td>54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>95</td>
<td>25</td>
<td>95</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate (%) of Attendance</td>
<td>CHICAGO</td>
<td>Rate (%) of Attendance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>---------</td>
<td>-----------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>as of January, 1969</td>
<td></td>
<td>as of January, 1969</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Ready for School</td>
<td></td>
<td>% Ready for School</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median IQ Score</td>
<td></td>
<td>Median IQ Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th Grade:</td>
<td></td>
<td>5th Grade:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>California Short-Form Test of Mental Maturity</td>
<td></td>
<td>California Short-Form Test of Mental Maturity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDG = Pre-4th, 6th and 8th Grades: Reading subtest of Metropolitan Achievement Test - Elementary, Intermediate and Advanced batteries, respectively</td>
<td></td>
<td>RDG = Pre-4th, 6th and 8th Grades: Reading subtest of Metropolitan Achievement Test - Elementary, Intermediate and Advanced batteries, respectively</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median IQ Score</td>
<td></td>
<td>Median IQ Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median IQ Score</td>
<td></td>
<td>Median IQ Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median IQ Score</td>
<td></td>
<td>Median IQ Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% with less than 1 yr. exp.</td>
<td></td>
<td>% with less than 1 yr. exp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% with 1-5 yrs. exp.</td>
<td></td>
<td>% with 1-5 yrs. exp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% with 6-14 yrs. exp.</td>
<td></td>
<td>% with 6-14 yrs. exp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% with 15 or more yrs. exp.</td>
<td></td>
<td>% with 15 or more yrs. exp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% B.A.</td>
<td></td>
<td>% B.A.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% M.A.</td>
<td></td>
<td>% M.A.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% M.A.</td>
<td></td>
<td>% M.A.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36 Hrs.</td>
<td></td>
<td>36 Hrs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Staff Experience and Education

- as indicated by step and lane on salary schedule
- as of October 24, 1969

### Lunch Program

<table>
<thead>
<tr>
<th>Yes or No - indicates full kitchen in the school</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Lunch</td>
<td>Lunch</td>
</tr>
<tr>
<td>Hot lunch prepared locally or in another location</td>
<td>Hot Lunches</td>
</tr>
<tr>
<td>Cold lunch prepared at another location</td>
<td>Cold Lunches</td>
</tr>
<tr>
<td>average number served daily during the month of October</td>
<td>$Free Lunches</td>
</tr>
<tr>
<td>lunches brought from home on a typical day, December, 1969</td>
<td>$Bag Lunches</td>
</tr>
</tbody>
</table>

### Special Programs which affect staffing:

- CCTV - Closed Circuit TV Studio: Govt. & Bd. of Ed. Funded
- CO-Plus* - Model Cities: Govt. Funded
- ESEA: Govt. Funded
- Follow Through: Govt. Funded
- Gifted Center: Govt. Funded
- Kenwood Experimental: Bd. of Ed. Funded
- Magnet School: Bd. of Ed. Funded
- Maximum Class Size: Bd. of Ed. Funded
- Mini-Magnet School: Bd. of Ed. Funded
- READ* - Reading Environment and Development: Bd. of Ed. Funded
- UTC - Urban Teacher Corps: Govt. Funded
- Woodlawn Experimental Project: Govt. Funded

* READ & CO-Plus staffing not completed in October, 1969

---

**Note:** The data presented includes various test results, student characteristics, and school program information.
A special service teacher is assigned to a school when the number of students in grades 1-8 warrants another classroom position, and there is no space available for the class—the function of the special service teacher is to provide instructional assistance to the 1-8 classroom teachers.

<table>
<thead>
<tr>
<th>Date of Construction</th>
<th>School Organization</th>
<th>School Category</th>
<th>Grades 1-8</th>
<th>Auxiliary Staff</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/24/69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student Membership as of October 24, 1969</th>
</tr>
</thead>
<tbody>
<tr>
<td>as of October 24, 1969</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Teacher aides and school-community representatives as of October 24, 1969</th>
</tr>
</thead>
<tbody>
<tr>
<td>as of October 24, 1969</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student membership, in primary 1 through 8 classes, including blind students, as of October 24, 1969</th>
</tr>
</thead>
<tbody>
<tr>
<td>as of October 24, 1969</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>number of students in primary 1 through 8 divided by number of primary 1 through 8 classroom teachers as of October 24, 1969</th>
</tr>
</thead>
<tbody>
<tr>
<td>as of October 24, 1969</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>number of students in primary 1 through 8 divided by the sum of the number of primary 1 through 8 classroom and special service teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>as of October 24, 1969</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Library, adjustment, physical education, master and other non-classroom teachers, excluding special service and freed assistant principals</th>
</tr>
</thead>
<tbody>
<tr>
<td>as of September 26, 1969</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C=Caucasian; N=Negroid; Pr=Puerto Rican (principal excluded)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Freed Asst. Prin.</th>
</tr>
</thead>
<tbody>
<tr>
<td>as of September 26, 1969</td>
</tr>
</tbody>
</table>
### COMMUNITY FACTORS

- **Percentages**:
  - 0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100

  - **Per cent of low-income families**
  - **Per cent of substandard housing**
  - **Rate of population mobility**
  - **Rate of in-migration**
  - **Per cent of parents with high school diplomas**
  - **Per cent of intact families**
  - **Incidence of ABC cases in enrollment**

### SCHOOL FACTORS

- **Percentages**:
  - 0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100

  - **Absence rate**
  - **Pupil mobility rate**
  - **Per cent of pupils new to Columbus public schools**
  - **Per cent of pupils above age in grade level**
  - **Per cent of white pupils**
  - **Per cent of classroom teachers to total professional staff**
  - **Per cent of professional staff in General Fund special programs**
  - **Per cent of professional staff in state federal funded programs**
  - **Staff turnover rate**
  - **Per cent of teachers with Bachelor's degree**
  - **Per cent of teachers with 150 hours of study, including M.A.**
  - **Per cent of teachers with Master's degree**
  - **Per cent of teachers with M.A. plus 20 or more hours of study**
  - **Per cent of teachers with 1 year or less experience**
  - **Per cent of teachers with 1 1/2 to 5 years experience**
  - **Per cent of teachers with 6 to 10 years experience**
  - **Per cent of teachers with 11 or more years experience**

**DATA WILL BE PORTRAYED FOR THESE FACTORS AS SOON AS REPORTS ON THE 1970 U.S. CENSUS BECOME AVAILABLE.**
Table VII
School Means (Averages) and Ranges of Elementary and Junior High School Pupils’
Composite Scores on the Iowa Tests of Basic Skills, October, 1968

<table>
<thead>
<tr>
<th>School</th>
<th>Grade 4 (National Mean, 1.1)</th>
<th>Grade 6B (National Mean, 6.1)</th>
<th>Grade 8B (National Mean, 8.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lowest Score</td>
<td>Middle 2/3</td>
<td>Highest Score</td>
</tr>
<tr>
<td>Addams</td>
<td>2.4</td>
<td>2.6</td>
<td>3.2</td>
</tr>
<tr>
<td>Alger</td>
<td>2.1</td>
<td>2.5</td>
<td>3.1</td>
</tr>
<tr>
<td>Amsa</td>
<td>2.3</td>
<td>2.6</td>
<td>3.1</td>
</tr>
<tr>
<td>Angell</td>
<td>2.1</td>
<td>2.6</td>
<td>3.2</td>
</tr>
<tr>
<td>Ann Arbor Trail</td>
<td>2.5</td>
<td>3.1</td>
<td>4.0</td>
</tr>
<tr>
<td>Arthur</td>
<td>2.1</td>
<td>2.5</td>
<td>3.1</td>
</tr>
<tr>
<td>Atkinson</td>
<td>2.1</td>
<td>2.5</td>
<td>3.1</td>
</tr>
<tr>
<td>Foyle</td>
<td>2.1</td>
<td>2.9</td>
<td>3.9</td>
</tr>
<tr>
<td>E. I. H.</td>
<td>2.4</td>
<td>2.5</td>
<td>3.1</td>
</tr>
<tr>
<td>Barbour</td>
<td>2.3</td>
<td>2.6</td>
<td>3.3</td>
</tr>
<tr>
<td>Barton</td>
<td>2.5</td>
<td>2.7</td>
<td>3.5</td>
</tr>
<tr>
<td>Beard</td>
<td>2.5</td>
<td>2.7</td>
<td>3.5</td>
</tr>
<tr>
<td>Beaubien</td>
<td>2.2</td>
<td>2.4</td>
<td>3.0</td>
</tr>
<tr>
<td>Bellevue</td>
<td>2.4</td>
<td>2.7</td>
<td>3.4</td>
</tr>
<tr>
<td>Bennett</td>
<td>2.4</td>
<td>2.7</td>
<td>3.4</td>
</tr>
<tr>
<td>Berry</td>
<td>2.3</td>
<td>2.5</td>
<td>3.1</td>
</tr>
<tr>
<td>Biddle</td>
<td>2.3</td>
<td>2.5</td>
<td>3.2</td>
</tr>
<tr>
<td>Burley Annex</td>
<td>2.7</td>
<td>2.9</td>
<td>3.4</td>
</tr>
<tr>
<td>Bow</td>
<td>2.5</td>
<td>3.2</td>
<td>3.8</td>
</tr>
<tr>
<td>Boynton</td>
<td>2.3</td>
<td>2.6</td>
<td>3.0</td>
</tr>
<tr>
<td>Brady</td>
<td>2.3</td>
<td>2.5</td>
<td>3.1</td>
</tr>
<tr>
<td>Breitmayer</td>
<td>2.4</td>
<td>2.9</td>
<td>3.5</td>
</tr>
<tr>
<td>Brooks</td>
<td>2.3</td>
<td>2.4</td>
<td>2.9</td>
</tr>
<tr>
<td>Bunche</td>
<td>2.3</td>
<td>2.4</td>
<td>2.9</td>
</tr>
<tr>
<td>Burbank</td>
<td>2.7</td>
<td>3.1</td>
<td>3.9</td>
</tr>
</tbody>
</table>

*The range of the middle 2/3 of the scores is determined from the mean and standard deviation, M-SD to M+SD.
One sixth of the pupils obtained scores above this range, one sixth below.
Table II

Per cents of Senior High School Students Whose Achievement Test Scores Are At or Above the Means of the Publisher's "National" Norms Groups on the Sequential Tests of Educational Progress, June, 1960

<table>
<thead>
<tr>
<th>Grade and Test</th>
<th>&quot;National&quot; Norms Group</th>
<th>Detroit</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 10B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>50%</td>
<td>37%</td>
<td>13%</td>
</tr>
<tr>
<td>Writing</td>
<td>50</td>
<td>42</td>
<td>3</td>
</tr>
<tr>
<td>Mathematics</td>
<td>50</td>
<td>41</td>
<td>9</td>
</tr>
<tr>
<td>Science</td>
<td>50</td>
<td>23</td>
<td>27</td>
</tr>
<tr>
<td>Social Studies</td>
<td>50</td>
<td>37</td>
<td>13</td>
</tr>
<tr>
<td>Grade 12B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>50%</td>
<td>44%</td>
<td>6%</td>
</tr>
<tr>
<td>Writing</td>
<td>50</td>
<td>45</td>
<td>5</td>
</tr>
<tr>
<td>Mathematics</td>
<td>50</td>
<td>44</td>
<td>6</td>
</tr>
<tr>
<td>Science</td>
<td>50</td>
<td>36</td>
<td>14</td>
</tr>
<tr>
<td>Social Studies</td>
<td>50</td>
<td>37</td>
<td>13</td>
</tr>
</tbody>
</table>

This table is read: In Reading in grade 10B, 50% of the students in the "national" norms group had scores that equalled or exceeded the mean of the scores of their group. In Detroit, 37% of the students achieved Reading test scores that equalled or exceeded the mean of the scores of the "national" norms group of grade 10B students; this is 13% less than the corresponding percent of students in the norms group.
### RESULTS OF STATE TESTING PROGRAM 1967-68

<table>
<thead>
<tr>
<th>Distribution Points</th>
<th>GRADE 1</th>
<th>GRADE 2</th>
<th>GRADE 3</th>
<th>GRADE 4</th>
<th>GRADE 5</th>
<th>Long-Thurstone Intelligence Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SAT Reading Test</td>
<td>SAT Reading Test</td>
<td>SAT Reading Test</td>
<td>SAT Reading Test</td>
<td>SAT Reading Test</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Calif. O.S. School</td>
<td>Calif. O.S. School</td>
<td>Calif. O.S. School</td>
<td>Calif. O.S. School</td>
<td>Calif. O.S. School</td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td>1.9</td>
<td>1.9</td>
<td>1.8</td>
<td>2.1</td>
<td>2.9</td>
<td>3.5</td>
</tr>
<tr>
<td>Md.</td>
<td>1.6</td>
<td>1.6</td>
<td>1.6</td>
<td>2.5</td>
<td>2.3</td>
<td>1.9</td>
</tr>
<tr>
<td>Q1</td>
<td>1.5</td>
<td>1.5</td>
<td>1.4</td>
<td>1.9</td>
<td>1.9</td>
<td>1.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SAT PRALINE TEST</th>
<th>1 SAT</th>
<th>2 SAT</th>
<th>3 SAT</th>
<th>4 SAT</th>
<th>5 SAT</th>
<th>6 SAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calif. O.S. School</td>
<td>65</td>
<td>65</td>
<td>65</td>
<td>65</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>Calif. O.S. School</td>
<td>65</td>
<td>65</td>
<td>65</td>
<td>65</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>Calif. O.S. School</td>
<td>65</td>
<td>65</td>
<td>65</td>
<td>65</td>
<td>65</td>
<td>65</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AVERAGE UNEXCUSED ABSENCES (Per Pupil)</th>
<th>PERCENT PUPIL MOBILITY</th>
<th>INDIAN FAMILY LUMBER (1960 Census)</th>
<th>RACIAL COMPOSITION OF STUDENT BODY March 1968</th>
</tr>
</thead>
<tbody>
<tr>
<td>City School</td>
<td>City School</td>
<td>City School</td>
<td>City School</td>
</tr>
<tr>
<td>3.0</td>
<td>4.6</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>64.1</td>
<td>90.4</td>
<td>95.0</td>
<td>95.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GRADES K-6</th>
<th>GRADES 1-6</th>
</tr>
</thead>
<tbody>
<tr>
<td>PUPIL-STAFF RATIO</td>
<td>MEAN CLASS SIZE</td>
</tr>
<tr>
<td>City School</td>
<td>City School</td>
</tr>
<tr>
<td>24.4</td>
<td>19.2</td>
</tr>
</tbody>
</table>

* State Results for 1966-67
### Vocabulary

<table>
<thead>
<tr>
<th>District</th>
<th>Grade Equivalent Score</th>
<th>National Grade Equivalent Score</th>
<th>District Average</th>
<th>National Rank of District Average</th>
<th>1969</th>
<th>1968</th>
<th>Change in National Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.8</td>
<td>5.6</td>
<td>5.7</td>
<td>21</td>
<td>20</td>
<td>1</td>
<td>+1</td>
</tr>
<tr>
<td>2</td>
<td>6.8</td>
<td>5.6</td>
<td>5.3</td>
<td>23</td>
<td>20</td>
<td>3</td>
<td>+3</td>
</tr>
<tr>
<td>3</td>
<td>6.8</td>
<td>5.6</td>
<td>5.3</td>
<td>23</td>
<td>19</td>
<td>4</td>
<td>+4</td>
</tr>
<tr>
<td>4</td>
<td>6.8</td>
<td>5.6</td>
<td>5.1</td>
<td>20</td>
<td>20</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>6.8</td>
<td>5.6</td>
<td>5.0</td>
<td>19</td>
<td>16</td>
<td>3</td>
<td>+3</td>
</tr>
<tr>
<td>6</td>
<td>6.8</td>
<td>5.6</td>
<td>5.9</td>
<td>33</td>
<td>35</td>
<td>2</td>
<td>-2</td>
</tr>
<tr>
<td>7</td>
<td>6.8</td>
<td>5.6</td>
<td>6.0</td>
<td>34</td>
<td>37</td>
<td>-3</td>
<td>-3</td>
</tr>
<tr>
<td>8</td>
<td>6.8</td>
<td>5.6</td>
<td>6.7</td>
<td>49</td>
<td>49</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### Reading

<table>
<thead>
<tr>
<th>District</th>
<th>Grade Equivalent Score</th>
<th>National Grade Equivalent Score</th>
<th>District Average</th>
<th>National Rank of District Average</th>
<th>1969</th>
<th>1968</th>
<th>Change in National Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.8</td>
<td>5.9</td>
<td>5.1</td>
<td>17</td>
<td>19</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>2</td>
<td>6.8</td>
<td>5.5</td>
<td>5.1</td>
<td>17</td>
<td>17</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>6.8</td>
<td>5.5</td>
<td>5.1</td>
<td>17</td>
<td>17</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>6.8</td>
<td>5.5</td>
<td>5.0</td>
<td>15</td>
<td>17</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>5</td>
<td>6.8</td>
<td>5.5</td>
<td>5.0</td>
<td>15</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>6.8</td>
<td>5.5</td>
<td>5.8</td>
<td>30</td>
<td>34</td>
<td>-4</td>
<td>-4</td>
</tr>
<tr>
<td>7</td>
<td>6.8</td>
<td>5.5</td>
<td>5.9</td>
<td>32</td>
<td>37</td>
<td>-5</td>
<td>-5</td>
</tr>
<tr>
<td>8</td>
<td>6.8</td>
<td>5.5</td>
<td>6.6</td>
<td>47</td>
<td>50</td>
<td>-3</td>
<td>-3</td>
</tr>
</tbody>
</table>

### Total Language

<table>
<thead>
<tr>
<th>District</th>
<th>Grade Equivalent Score</th>
<th>National Grade Equivalent Score</th>
<th>District Average</th>
<th>National Rank of District Average</th>
<th>1969</th>
<th>1968</th>
<th>Change in National Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.8</td>
<td>5.5</td>
<td>5.1</td>
<td>21</td>
<td>21</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>6.8</td>
<td>5.5</td>
<td>5.1</td>
<td>21</td>
<td>21</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>6.8</td>
<td>5.5</td>
<td>5.0</td>
<td>19</td>
<td>19</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>6.8</td>
<td>5.5</td>
<td>4.9</td>
<td>16</td>
<td>15</td>
<td>+1</td>
<td>+1</td>
</tr>
<tr>
<td>5</td>
<td>6.8</td>
<td>5.5</td>
<td>5.9</td>
<td>35</td>
<td>37</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>6</td>
<td>6.8</td>
<td>5.5</td>
<td>6.0</td>
<td>37</td>
<td>38</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>7</td>
<td>6.8</td>
<td>5.5</td>
<td>6.7</td>
<td>50</td>
<td>50</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>6.8</td>
<td>5.5</td>
<td>6.7</td>
<td>50</td>
<td>50</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
# THE SCHOOL DISTRICT OF PHILADELPHIA
Office of Research and Evaluation
Division of Testing

PHILADELPHIA CITY-WIDE TESTING PROGRAM - SPRING 1969
IOWA TESTS OF BASIC SKILLS - GRADE 6

BUILDING AVERAGES - DISTRICT NO. 4

<table>
<thead>
<tr>
<th>Reading (Reading Comprehension)</th>
<th>Composite (3)</th>
<th>Total Arithmetic (Concepts &amp; Problem Solving)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G.E. Score(1)</td>
<td>Nat'l File Score(1)</td>
</tr>
<tr>
<td>National Norm</td>
<td>6.8</td>
<td>52</td>
</tr>
<tr>
<td>City Average</td>
<td>5.9</td>
<td>24</td>
</tr>
<tr>
<td>District Average</td>
<td>5.0</td>
<td>24</td>
</tr>
<tr>
<td>Allison</td>
<td>3.9</td>
<td>4</td>
</tr>
<tr>
<td>Blaine</td>
<td>4.7</td>
<td>11</td>
</tr>
<tr>
<td>Blankenburg</td>
<td>4.9</td>
<td>11</td>
</tr>
<tr>
<td>Cassidy</td>
<td>4.9</td>
<td>11</td>
</tr>
<tr>
<td>Cleveland</td>
<td>5.0</td>
<td>15</td>
</tr>
<tr>
<td>Dick</td>
<td>4.6</td>
<td>10</td>
</tr>
<tr>
<td>Duckrey</td>
<td>4.0</td>
<td>4</td>
</tr>
<tr>
<td>Hampers</td>
<td>6.6</td>
<td>47</td>
</tr>
<tr>
<td>Neaia</td>
<td>5.0</td>
<td>15</td>
</tr>
<tr>
<td>Newton</td>
<td>5.2</td>
<td>19</td>
</tr>
<tr>
<td>Hill, L. F.</td>
<td>5.5</td>
<td>24</td>
</tr>
<tr>
<td>Konderton</td>
<td>4.4</td>
<td>8</td>
</tr>
<tr>
<td>Lambert</td>
<td>6.7</td>
<td>50</td>
</tr>
<tr>
<td>Lehigh</td>
<td>4.7</td>
<td>12</td>
</tr>
<tr>
<td>Ledy</td>
<td>4.7</td>
<td>12</td>
</tr>
<tr>
<td>Mann</td>
<td>5.6</td>
<td>26</td>
</tr>
<tr>
<td>Melody</td>
<td>4.6</td>
<td>10</td>
</tr>
<tr>
<td>Overbrook</td>
<td>5.8</td>
<td>30</td>
</tr>
<tr>
<td>Peirce, T. K.</td>
<td>4.5</td>
<td>9</td>
</tr>
<tr>
<td>Pratt-Arnold</td>
<td>4.8</td>
<td>12</td>
</tr>
<tr>
<td>Stanton, M. H.</td>
<td>4.8</td>
<td>12</td>
</tr>
<tr>
<td>Stokley</td>
<td>4.8</td>
<td>12</td>
</tr>
<tr>
<td>Walton</td>
<td>4.7</td>
<td>12</td>
</tr>
<tr>
<td>Whittier</td>
<td>4.4</td>
<td>8</td>
</tr>
</tbody>
</table>

Notes: 1) and 2) - See explanation on page 57
3) The Composite Score is the average of scores for Vocabulary, Reading, Total Language, Total Work-Study and Total Arithmetic.
RESULTS OF 1969-70 TESTING PROGRAM

Changes in Required Tests

The 1969-70 school year saw the substitution of new State-mandated tests for old ones at grades one and six, the elimination of required testing at grades eight and ten, and the introduction of required testing at grade twelve.

The State-mandated Cooperative Primary Tests, used in the first grade during 1969-70, will be extended to the second grade in 1970-71 and to the third grade in 1971-72.

Districtwide Reading Achievement Test Results

San Francisco Unified School District elementary students achieved higher median reading scores during 1969-70 than in 1968-69.

<table>
<thead>
<tr>
<th>Grade</th>
<th>1960-68</th>
<th>1968-69</th>
<th>1969-70</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.8</td>
<td>1.6</td>
<td>1.8*</td>
</tr>
<tr>
<td>2</td>
<td>2.8</td>
<td>2.2</td>
<td>2.4</td>
</tr>
<tr>
<td>3</td>
<td>3.8</td>
<td>3.0</td>
<td>3.1</td>
</tr>
</tbody>
</table>

Districtwide Mental Ability Test Results

<table>
<thead>
<tr>
<th>Grade</th>
<th>Verbal</th>
<th>Nonverbal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>96</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>12</td>
<td>99</td>
<td>106</td>
<td>103</td>
</tr>
</tbody>
</table>

CONTENTS

Page 2  Describes the 1969-70 testing program
Pages 3-9 Gives the results for grades 1, 2, 3, and 6 by school
Pages 10-11 Gives the results for grades 8, 11, and 12 by school
<table>
<thead>
<tr>
<th>School</th>
<th>Grade 1</th>
<th>Grade 2</th>
<th>Grade 3</th>
<th>Grade 4</th>
<th>Grade 5</th>
<th>Grade 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Columbus</td>
<td>1.9</td>
<td>1.6</td>
<td>1.5</td>
<td>1.2</td>
<td>1.0</td>
<td>1.2</td>
</tr>
<tr>
<td>Comer, Shute</td>
<td>2.1</td>
<td>2.0</td>
<td>1.9</td>
<td>1.8</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Comer, Stockton</td>
<td>1.9</td>
<td>2.1</td>
<td>1.7</td>
<td>1.6</td>
<td>1.4</td>
<td>1.4</td>
</tr>
<tr>
<td>Cortis</td>
<td>1.7</td>
<td>1.6</td>
<td>1.5</td>
<td>1.3</td>
<td>1.1</td>
<td>1.2</td>
</tr>
<tr>
<td>Dan. Webster</td>
<td>1.6</td>
<td>1.6</td>
<td>1.5</td>
<td>1.4</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>Diamond Hts.</td>
<td>2.0</td>
<td>1.6</td>
<td>1.5</td>
<td>1.4</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>Douglas</td>
<td>2.3</td>
<td>2.3</td>
<td>2.2</td>
<td>2.1</td>
<td>1.9</td>
<td>1.9</td>
</tr>
<tr>
<td>Dudley Stone</td>
<td>1.6</td>
<td>1.6</td>
<td>1.5</td>
<td>1.4</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>Edinburg</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.4</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>E. R. Taylor</td>
<td>1.8</td>
<td>1.5</td>
<td>1.6</td>
<td>1.5</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>El Cerrado</td>
<td>1.4</td>
<td>1.5</td>
<td>1.4</td>
<td>1.3</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>Emerson</td>
<td>1.6</td>
<td>1.6</td>
<td>1.5</td>
<td>1.4</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>Fairfax</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.4</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>Fairmont</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.4</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>Forsyatt</td>
<td>1.4</td>
<td>1.4</td>
<td>1.4</td>
<td>1.3</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>F. J. Ky.</td>
<td>1.9</td>
<td>1.9</td>
<td>1.8</td>
<td>1.7</td>
<td>1.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>
APPENDIX C

EXAMPLES OF NEWS COVERAGE OF
LARGE CITY ACCOUNTABILITY
RELEASES
Highlights:

Here are the highlights of 1968-69 test scores for the Chicago public schools, presented to the board of education yesterday:

1. On a city-wide basis, the city's school children fell further behind national test norms than they had been before.

2. On a school-by-school basis, white schools had markedly higher test scores than black schools.

3. School officials cited lack of money, inexperienced teachers, and socio-economic factors as reasons for the discouraging test results.

4. Supt. James F. Redmond indicated that school principals might be judged on how well their pupils learn.
The test score issue was even more sticky. There was the argument that schools, principals, and even teachers would be blamed in specific instances for low test scores instead of recognizing that pupils' home environment is a key factor. There was the supposition that results would be used to compare schools and label many as "bad." This was not usually accompanied by the admission that just about everyone knew which schools were "bad" anyway. There was the claim that test scores must be a private matter between school officials and parents of pupils. There was the cry of the educational establishment that people just don't understand the concept of test scores and would make invalid use of the results.

When test scores were made public last January, nothing new was revealed. Predictions that doom would accompany their release have proved to be unfounded. On the question of interpreting the scores, one leading administrator emphasizes that publication of the scores should be a challenge to the professional staff to use them as a vehicle for educating the public thru interpretation and explanation.

When School Supt. James F. Redmond assumed his duties here in 1963, he listed the phone number flap and low test scores hang-up as priority items. This precipitated considerable indignant grumbling among the educators. Predicted incessantly by reporters about when he would take those two small steps forward, the superintendent repeatedly, and slyly, answered that phone numbers and test scores would come in due time. Being the politician he is, Mr. Redmond accurately gauged the value of waiting until the natives were less resistive. So it took two years to get the phone numbers in the book and nearly four to get the test scores out in the public domain.

Now that these two myths have been shattered, why not start attacking two others? One is the view held in some quarters that black children cannot be expected to perform at grade level because of home backgrounds. Another is the contention that vast increases of money are needed to operate an effective school system capable of turning out thousands of young people who can fulfill worthwhile roles in society and achieve their life goals.
"City schools' inequities told in new report"

By Hope Justice

School Supt. James F. Redmond Wednesday released a research report showing that schools with the most problems tend to be staffed with the least-experienced and least-educated faculties.

The report indicates that it is the city's least-experienced teachers who, for the most part, are coping with such major education problems as:

- Overcrowded classrooms and related problems in racially changing neighborhoods where enrollment is increasing rapidly.
- Student bodies with the highest numbers of poor academic achievers.
- The bulk of the 30,000 students who speak little or no English.

The per-pupil staffing costs given in the report were derived by dividing the total salaries paid to classroom teachers in a school by the number of pupils.

The report's per-pupil staffing costs, however, do not include such factors as textbooks, materials and other costs usually included in per pupil cost figures given out by school systems.

The figures released Wednesday document for the first time what has long been generally accepted — the uneven distribution of teachers by years of service and education also.

- The concentration of better-educated and more experienced teachers can be seen in these figures:
  - Of 423 elementary schools, 159 have faculties composed of 50 per cent or more teachers with at least 15 years experience. Of the 63, only 4 have heavily Negro student bodies.
  - Most of the 63 are in the outlying areas of the city on the North West and Southwest sides.

The school with the highest per-pupil staffing cost in the city was Spalding School for Handicapped Children, 1611 West 58th Place, with a per-pupil staffing cost of $11.33.

Libby Elementary, 2300 South Damen, which has an enrollment 47 per cent white, 46 per cent Negro and 6 per cent Puerto Rican, and Goldblatt Elementary, 4237 West Adams, with an enrollment 33 per cent Negro, had the lowest. Both had per-pupil staffing costs of $2.22.

The report shows that the racial headcounts of the last few years have shown that school faculties are racially segregated along the same general geographic lines with the majority of black teachers working in the black schools of the South and West sides.

The figures released Wednesday document for the first time what has long been generally accepted — the uneven distribution of teachers by years of service and education also.

- The concentration of better-educated and more experienced teachers can be seen in these figures:
  - Of 423 elementary schools, 159 have faculties composed of 50 per cent or more teachers with at least 15 years experience. Of the 63, only 4 have heavily Negro student bodies.
  - Most of the 63 are in the outlying areas of the city on the North West and Southwest sides.

The figures released Wednesday document for the first time what has long been generally accepted — the uneven distribution of teachers by years of service and education also.

- The concentration of better-educated and more experienced teachers can be seen in these figures:
  - Of 423 elementary schools, 159 have faculties composed of 50 per cent or more teachers with at least 15 years experience. Of the 63, only 4 have heavily Negro student bodies.
  - Most of the 63 are in the outlying areas of the city on the North West and Southwest sides.

AMONG other highlights of the report were figures showing:

- The system has 139 elementary schools and 36 high schools where an average of more than 10 per cent of the pupils are absent daily.
- The high school with the highest dropout rate is Farragut on the South Side where between September, 1965, and June, 1970, 23.2 per cent of the students either left school on their own or were placed in the category identified as "dropped from the school roll for lack of ability to adjust to school environment." During the same period on the North Side militias had the lowest dropout rate — half of 1 per cent.
- Forty schools have some pupils for whom English is a second language.

- Of 185 schools where 20 per cent or more of the faculty have master's degrees or further education, 145 have heavily white student bodies, 38 have predominantly Negro student bodies, and one school has a Puerto Rican majority.

IN MANY of the black schools where better-educated and more-experienced teachers are found, they appear to have been attracted there by special program or some other "favored" condition.

An example is the Pershing Elementary School, 3113 South Rhodes, where 61 per cent of the faculty has 15 years or more experience and 45 per cent have master's degrees. The school is located near the Prairie Shores-Lake Meadow apartments complex where children tend to come from middle-class backgrounds and teachers can find housing nearby also.

- The system has 139 elementary schools and 36 high schools where an average of more than 10 per cent of the pupils are absent daily.
- The high school with the highest dropout rate is Farragut on the South Side where between September, 1965, and June, 1970, 23.2 per cent of the students either left school on their own or were placed in the category identified as "dropped from the school roll for lack of ability to adjust to school environment." During the same period on the North Side militias had the lowest dropout rate — half of 1 per cent.
- Forty schools have some pupils for whom English is a second language.
Children in the Chicago public schools are lagging further behind national norms than in the past, according to city-wide test scores from 1968-69 which were presented to the board of education yesterday.

The scores indicated that children attending white or predominantly white schools did markedly better on the tests than did children attending black or predominantly black schools.

The first release ever of elementary and high school scores on a school-by-school basis, a step long resisted by school administrators, threatens to start a long controversy over whether the children are no as intelligent as they were in the past, or whether the schools have fallen down on the job of educating them.

Behind in Reading
The tests were given at various times during the 1968-69 school year at the beginning and middle of 1st grade, and in 3rd, 6th, 8th, 9th, and 11th grades. They covered numerous areas of knowledge and learning ability, but focused most heavily on reading and other verbal skills, and on arithmetic.

In reading, stressed by educators as the most important skill a child must acquire in school, the test results show children in Chicago continue to rank behind the national median of 50.

In fact, they fell even farther behind where they ranked in 1966-67 in 3rd, 6th, and 8th grade reading tests.

Lack H. S. Comparisons
High school pupils also continued to perform below the national test norms. Although comparison scores were available, school officials said there was no significant improvement over previous years.

Here's how they came out, as expressed in percentile rank related to the median:

<table>
<thead>
<tr>
<th>Grade</th>
<th>1966-67</th>
<th>1968-69</th>
</tr>
</thead>
<tbody>
<tr>
<td>3rd grade</td>
<td>38</td>
<td>34</td>
</tr>
<tr>
<td>6th grade</td>
<td>31</td>
<td>29</td>
</tr>
<tr>
<td>8th grade</td>
<td>44</td>
<td>38</td>
</tr>
</tbody>
</table>

The same trend was evident in tests on arithmetic:

<table>
<thead>
<tr>
<th>Grade</th>
<th>1966-67</th>
<th>1968-69</th>
</tr>
</thead>
<tbody>
<tr>
<td>3rd grade</td>
<td>45</td>
<td>43</td>
</tr>
<tr>
<td>6th grade</td>
<td>39</td>
<td>35</td>
</tr>
<tr>
<td>8th grade</td>
<td>37</td>
<td>28</td>
</tr>
</tbody>
</table>

And in tests on school learning ability:

<table>
<thead>
<tr>
<th>Grade</th>
<th>1966-67</th>
<th>1968-69</th>
</tr>
</thead>
<tbody>
<tr>
<td>3rd grade</td>
<td>34</td>
<td>37</td>
</tr>
<tr>
<td>6th grade</td>
<td>40</td>
<td>38</td>
</tr>
<tr>
<td>8th grade</td>
<td>43</td>
<td>38</td>
</tr>
</tbody>
</table>

Court Orders Tests of IQ
In Pupils’ Home Languages

By a WALL STREET JOURNAL Staff Reporter

SAN FRANCISCO — California school children whose language at home is other than English must be tested in both their primary language and in English before they can be placed in classes for the mentally retarded, according to a settlement reached in Federal district court here.

The settlement between state education officials and California Rural Legal Assistance, an anti-poverty agency, was made in response to a suit filed by the agency last Jan. 7. The suit was brought on behalf of nine Mexican-American children who, the agency argued, were put in classes for the mentally retarded after they were given Q. or intelligence, tests in English.

The settlement also provides that Mexican-American and Chinese children already in classes for mentally retarded must be retested in their primary language, unless they were previously tested in it.
**Test Indicates School Performance Here**

The following are the profiles of Baltimore's elementary and junior high schools as depicted by the Iowa test of basic skills, administered in the city in May, 1969, and covering 98,000 students.

In addition to each school median for a given grade, median totals are given for "large cities." Philadelphia and the whole of Baltimore.

Philadelphia's scores are compared to those for a given grade, median total for the spring of 1968, and totals are given for "large cities," Philadelphia, and the Spring of 1968, and totals are given for "large cities," Philadelphia, and the whole of Baltimore.

Junior High Schools

<table>
<thead>
<tr>
<th>School Number</th>
<th>Grade 7 Median</th>
<th>Grade 8 Median</th>
<th>Grade 9 Median</th>
<th>Per Pupil Enrollment</th>
<th>Per Pupil White</th>
<th>Building Capacity</th>
<th>Enrollment</th>
<th>Pupil-Staff Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baltimore City</td>
<td>Philadelphia</td>
<td>Baltimore City</td>
<td></td>
<td></td>
<td>Building Capacity</td>
<td>Enrollment</td>
<td>Pupil-Staff Ratio</td>
</tr>
</tbody>
</table>

Elementary Schools

<table>
<thead>
<tr>
<th>School Number</th>
<th>Grade 1 Median</th>
<th>Grade 2 Median</th>
<th>Grade 3 Median</th>
<th>Per Pupil Enrollment</th>
<th>Per Pupil White</th>
<th>Building Capacity</th>
<th>Enrollment</th>
<th>Pupil-Staff Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baltimore City</td>
<td>Philadelphia</td>
<td>Building Capacity</td>
<td></td>
<td></td>
<td>Building Capacity</td>
<td>Enrollment</td>
<td>Pupil-Staff Ratio</td>
</tr>
</tbody>
</table>
CITY PUPILS RATE BELOW THE AVERAGE

Do Not Equal Scores Made In Other Big Cities, Tests Show

BY ANTERO PIETILA

Results of the first uniform city-wide achievement test indicate that an average Baltimore elementary school student scored noticeably lower than students in Philadelphia and other large cities.

Baltimore junior high school students, on the other hand, while still performing worse than the average student in "large cities," do generally better than their classmates in Philadelphia.

These were among the findings of a test costing $350,000 that was given to 93,000 Baltimore elementary and junior high school students last May. They were made public yesterday by the city's Department of Education.

A Marked Difference

Other than that, the test which gives only hard number scores and no interpretation, indicates sharp inequalities in performance between "white" schools and schools whose majority student population the test defines as "non-white."

Out of the 154 elementary schools tested, all but 21 of the 55 "black" schools showed a performance level lower than the average in Baltimore. And of the 15 "black" junior high schools, only two, the Pimlico Junior High School and the Woodbourne Junior High School, ranked better than average in overall scores.

More important than this, however, seems to be the fact that while performance in many "white" schools varied between below-and-above-average, Negro schools seldom broke any given trend during the six years of education, according to the test.

The testing method used in Baltimore was the Iowa Testing of Basic Skills, whose composite score is an average of vocabulary, reading, language skills, work-study skills and arithmetic skills.

The median scores for the "large cities," which included such places as Columbus, Ga.; Rockford, Ill.; Des Moines; Detroit; Omaha; New York; Cleveland; Allentown, Pa.; Providence, Rhode Island, and San Antonio, were developed by random sampling, not by an absolute compilation of all test scores of all students tested in those cities.

Therefore, because the Philadelphia scores and the Baltimore scores were the absolute median scores of all the thousands of students who took the Iowa test in both cities, these two sets of scores are thought to be a better basis for comparison with the results of individual Baltimore schools.

School officials pointed out yesterday, however, that test scores are only one indicator of a student's achievement or lack of it. Nor do the test scores indicate the causes underlying the student's level of performance.
Those Low Scores

School officials had been reluctant to tell the results of the Baltimore-wide achievement tests on a school-by-school basis, fearing "misinterpretation" of the data. Over the weekend they downed their reluctance and released the information for publication. There are a lot of figures; doubtless there will be some misunderstanding, and doubtless a lot of parents are going to be asking why Baltimore's school children have done worse, on average, than their counterparts in Des Moines, Detroit, Omaha, New York and other large cities. The city Department of Education would do well to be ready with some answers.

One thing the test results reveal quite clearly is that schools with predominately black student populations did worse than those schools predominately white or all white. The black schools dragged down the general level for the whole city. This bit of information is one which school authorities feared would be most misinterpreted. They were worried that a lot of people would not understand the reasons why black students scored below whites on the tests.

Among the causes is what school officials call "out of school factors." Among these are poverty, lack of motivation, unstable home life, parents who are illiterate or themselves unmotivated. There is more of this in the black ghettos, and all of it retards education.

It is difficult for a child to be oriented toward academic achievements while living in a ghetto. It is fairly obvious from the averages that a vast remedial effort is needed in the city's black schools. The better teachers and administrative talent should be directed there. That's where the problem is.

Model (?) Schools

With one exception, Baltimore's six model elementary schools do not make a good showing in the citywide test data recently released by the school administration. Despite the city policy of keeping class sizes in the model schools at less than 20 pupils per teacher and providing a full panoply of supportive services, the test results among model-school children are not as high as those in many less-favored inner-city schools.

In three of the six model schools the sixth-grade scores were a full year below the citywide median, which in turn was a year below the national norm. In other words, most of the sixth-graders performed on the tests at a fourth-grade level. In another of the model schools the third-graders were exactly on a par with the citywide average but the fourth and fifth grades were somewhat below the city norm and the sixth grades were six months behind. At a fifth school the third-graders were four months ahead of the Baltimore median, while the sixth graders were two months behind.
Two out of every five Philadelphia public school pupils perform "below minimum functioning levels," according to latest standardized tests.

The tests were administered here last spring to 140,496 pupils in grades two through eight. Nationally, 1.4 million pupils took the tests.

"Minimum functioning levels" are defined by the tests' publishers as levels sufficient to cope with basic instructional materials, such as basal readers and text books.

Nationally, 18 percent of all pupils tested fell below the minimum functioning level compared to 40 percent here.

Two Basic Areas

The tests measure skills in two basic areas: Reading and arithmetic.

Reading skills include language, vocabulary, paragraph meaning, spelling, punctuation, and use of maps and graphs. Arithmetic includes both computational and conceptual abilities.

This was the fourth city-wide testing since the program began in 1966. Comparisons between each year show little change except in the eighth grade, where average performance fell four percentage points below 1966 and two percentage points below 1968.

Separate tests were administered to pupils in grades two and to those in grades three through eight.

Principal Findings

In grade two 21,357 children took the Stanford Achievement Test. Some of the principal findings include the following:

—Average scores ranged from a high of two months below national averages in languages to a low of six months below.

—Scores were higher in spelling and language than in other areas tested, including arithmetic.

—In three areas related to reading (word meaning, paragraph meaning, and word study skills), one out of every two pupils was below minimum functioning levels.

—In the other five areas tested (science and social studies, spelling, language, arithmetic computation and arithmetic concepts), two out of every five pupils performed below minimum functioning levels.

The tests also measured the average performance of Philadelphia pupils against the average performance of pupils in the nation.

2d Grade Averages

In grade two the average Philadelphia pupil performed better than 34 percent of all pupils tested in language and better than 18 percent of all pupils in paragraph meaning.

Pupils surpassing nationwide averages ranged from 30 percent in spelling to 16 percent in word study skills.

Some school officials question the validity of national averages since they do not take into account factors affecting performance, such as social and economic deprivation.

"The norms suggest levels of performance for urban children which, probably, cannot be expected when all factors are considered."

More meaningful for Philadelphia, they said, would be a set of "great city" norms by which performance here could be compared with New York, Chicago, Detroit and other large cities.

In grades three through eight, 118,943 pupils took the Iowa Tests of Basic Skills. Findings include the following:

—Average scores ranged from a minimum of four months below national averages in year three (language) to a maximum of a year and nine months below national norms in grade eight (arithmetic).
To enact section 3301.18 of the Revised Code to require the state board of education to establish a statewide testing program to measure reading proficiency among pupils in elementary and secondary schools.

Be it enacted by the General Assembly of the State of Ohio:

SECTION 1. That section 3301.18 of the Revised Code be enacted to read as follows:

Sec. 3301.18. (A) THE STATE BOARD OF EDUCATION SHALL ESTABLISH STANDARDS FOR READING PROFICIENCY AMONG STUDENTS AT THE THIRD, SEVENTH, AND TENTH GRADE LEVELS WHO ATTEND SCHOOLS FOR WHICH THE STATE BOARD OF EDUCATION SETS STANDARDS PURSUANT TO SECTION 3301.07 OF THE REVISED CODE.
(B) THE STATE BOARD OF EDUCATION SHALL ESTABLISH A TESTING PROGRAM FOR MEASURING THE DEVELOPMENT OF READING SKILLS AMONG PUPILS ATTENDING SCHOOLS FOR WHICH THE STATE BOARD OF EDUCATION SETS STANDARDS. ALL SUCH PUPILS SHALL BE TESTED UPON THEIR ENTRY INTO THE THIRD, SEVENTH, AND TENTH GRADE LEVELS. THE PROGRAM SHALL BE ADMINISTERED THROUGH THE STATE DEPARTMENT OF EDUCATION AND THE SUPERINTENDENT OF PUBLIC INSTRUCTION, WHO SHALL ANNUALLY COMPILE THE DATA RESULTING FROM THE TESTING PROGRAM AND SUBMIT IT TO THE STATE BOARD OF EDUCATION ALONG WITH HIS RECOMMENDATIONS FOR IMPROVING THE QUALITY OF READING INSTRUCTION THROUGHOUT THE STATE AND FOR MEETING THE STANDARDS SET PURSUANT TO DIVISION (A) OF THIS SECTION.
BIBLIOGRAPHY
BIBLIOGRAPHY

BOOKS


Jarvis, Oscar T., Gentry, Harold W., and Stephens, Lester D. Public School Administration and Finance. West Wynak, New Jersey


PERIODICALS


"No Simple Solution Seen to Reading Problems." Education U. S. A., October 5, 1970, p. 27.


"Alternative Schools: Challenge to Traditional Education."


UNPUBLISHED MATERIALS

Blanchard, Robert W. Statement at Special Board Meeting, Portland, Oregon, January 21, 1970, pp. 1-4 (Mimeographed.)

Bulletin of Technical Data, CTBS. Monterey, California: California Test Bureau, 1969. (Mimeographed.)

Cavanaugh, Gerald. Address at the Large City Test Directors' Meeting, Las Vegas, Nevada, April, 1969.


Examiners Manual, CTBS. Monterey, California, California Test Bureau, 1968. (Mimeographed.)


Large City Norms. Monterey, California: California Test Bureau, 1969. (Mimeographed.)


McClelland, Samuel D. Personal interview held during Invitational Measurement Conference, Philadelphia, April, 1969.