DOCUMENT RESUME

ED 289 909 TM 870 738

AUTHOR Kippel, Gary M.; Forehand, Garlie A.
TITLE SMRT-STEPS: School Mastery of Reading Test System to Enhance Progress of Schools. Fall, 1986 Progress Report.
INSTITUTION New York City Board of Education, Brooklyn. Office of Educational Assessment.
PUB DATE 86
NOTE 111p.
PUB TYPE Reports - Research/Technical (143) -- Statistical Data (110) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Administrator Attitudes; Attitude Measures; Comparative Testing; Field Tests; Grade 4; Instructional Improvement; Intermediate Grades; Local Norms; *Mastery Tests; Minimum Competency Testing; Reading Diagnosis; *Reading Tests; School Statistics; Scoring; Surveys; *Teacher Attitudes; *Test Construction
IDENTIFIERS Curriculum Related Testing; New York City Board of Education; *School Mastery of Reading Test

ABSTRACT

This progress report describes the development and field testing of the School Mastery of Reading Test (SMRT). The SMRT was administered in May 1986 to 889 fourth graders in nine schools in Brooklyn, New York. Counts of students tested in each school and school profile data are provided. Machine scoring procedures are detailed, and results tabulated, showing that the SMRT was relatively easy (mean about 80% correct) with a negatively skewed score distribution. Subtest percentage scores are given by school and by district, as well as norms in percentiles and stanines. A professional panel of teachers, supervisors, and two Educational Testing Service personnel rated the usefulness of SMRT and made suggestions for relevant school improvement plans. To develop mastery scores and minimum standards, the panel judged item difficulty for three reading competency levels (satisfactory, minimum, below minimum); these judgments were coordinated with actual proportion correct on each item for three groups defined by Degrees of Reading Power test scores. The report considers the validity, reliability, and predictive value of SMRT, and its relation to National Assessment of Educational Progress norms. Attachments include questionnaires administered to the expert panel and the test-site teachers. (LPG)

SMRT - STEPS

SCHOOL MASTERY OF READING TEST
SYSTEM TO ENHANCE PROGRESS OF SCHOOLS

Fall, 1986 Progress Report

Report Prepared by:

Gary M. Kippel
Project Director
Office of Educational Assessment
Board of Education of New York City

Garlie A. Forehand
Director of Research
Educational Testing Service
Princeton, New Jersey

New York City Public Schools
Office of Educational Assessment
Richard Guttenberg, Director

It is the policy of the Board of Education not to discriminate on the basis of race, creed, national origin, age, handicapping condition, sexual orientation, or sex, in its educational programs, activities, and employment policies, as required by law. Any person who believes he or she has been discriminated against should contact: Carole Guerra, Local Equal Opportunity Coordinator, Office of Educational Assessment, 110 Livingston Street, Room 743, Brooklyn, New York 11201. Inquiries regarding compliance with appropriate laws may also be directed to: Mercedes A. Nesfield, Director, Office of Equal Opportunity, 110 Livingston Street, Room 601, Brooklyn, New York; or the Director, Office of Civil Rights, U.S. Department of Education, 26 Federal Plaza, Room 33-130, New York, New York 10278.

NEW YORK CITY BOARD OF EDUCATION

Robert F. Wagner, Jr., President

Irene Impellizzeri, Vice President

Gwendolyn C. Baker

Richard I. Beattie

Stephen R. Franse

James F. Regan

Edward L. Sadowsky

Nathan Quinones, Chancellor

Charles I. Schonhaut, Deputy Chancellor

Louise Latty, Chief Executive for Instruction

EXECUTIVE SUMMARY OF ACCOMPLISHMENTS OF THE SMRT-STEPS PROJECT

The following provides a brief listing of accomplishments of the SMRT-STEPS Project:

-- Field tested the fourth-grade School Mastery of Reading Test (SMRT) in a total of nine Comprehensive Assessment Report (CAR) elementary schools in three Community School Districts.

-- Prepared a test administration manual and procedures which were used successfully by regular fourth grade teachers, with no advanced test administration training.

-- Demonstrated that SMRT could be administered cost-effectively by developing re-usable test booklets and machine scannable answer sheets.

-- With the Educational Testing Service (ETS) in Princeton, New Jersey, began the development of a high-quality funding proposal for potential submission to federal, state government and private foundations.

-- Conferred with ETS regarding future possibilities of using state-of-the-art computer technology in the administration, scoring, reporting and storing of SMRT.

-- Established a Professional Panel comprised of both teachers and supervisors to review SMRT for potential bias and to provide judgments related to mastery criteria. In addition, obtained panel member opinions regarding the usefulness of standardized tests and the types of test scores found most useful. The overall objective of the Professional Panel is to increase the meaningfulness and instructional usefulness of all aspects of SMRT-STEPS.

-- Field practitioners including both teachers who administered SMRT and Professional Panel members provided ratings and opinions reflecting very favorably on the potential usefulness of SMRT.

-- Demonstrated that SMRT results from a mid-May administration are relatively high and result in the type of test score distribution which would be expected from a mastery test which is related to curriculum and is administered at the end of the academic year.

-- Established a prototype of SMRT New York City norms by generating SMRT raw score to percentile and stanine norms.

-- Reviewed all SMRT items and found them related to New York City reading curriculum.

-- Provided the basis for comparing SMRT to the National Assessment of Education Progress (NAEP) by embedding NAEP items within SMRT and administering both at the same time to the same students. This may result in the capability of interpreting SMRT results using NAEP national norms and, possibly, of using NAEP as a cost-effective source of new test items for SMRT.

-- Established the basis for comparing SMRT results to both Degrees of Reading Power (DRP) test and Metropolitan Achievement Test (MAT) performance of grade four students.

-- Conferred with School Improvement Program (SIP) staff regarding the relationship between SMRT-STEPS and current New York City school improvement efforts.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd)

LIST OF TABLES

LIST OF FIGURES

I. BRIEF PROGRAM DESCRIPTION

The primary objective of this project is to develop a system to provide school administrators and teachers with reading performance scores and information useful for improving the school instructional program. Furthermore, it is our intention to review various current New York City School improvement plans and to assess the potential linkage between school level diagnosis and prescription in order to enhance the progress of schools. This system is expected to be a particularly useful adjunct to the New York State Comprehensive Assessment Report (CAR) by diagnosing school needs for particular improvement plans developed by the New York City Board of Education.

Consequently, when SMRT-STEPS (pronounced "SMART STEPS") is completely validated it may be considered a school level diagnostic-prescriptive system. In effect, weaknesses requiring remediation will be identified. Subsequently, results from testing may "elicit" or assist in the selection of school improvement plans designed to improve the effectiveness of the instructional program in those grades for which it is available.

To expedite communication, the acronym "SMRT-STEPS" will be used to refer to the entire School Mastery of Reading Test System to Enhance Progress of Schools. The acronym "SMRT" will be used to refer primarily to the assessment component School Mastery of Reading Test.

To enhance its relevance and usefulness for improving instruction, SMRT is being developed as an objective test of mastery of reading, rather than as a norm-referenced test. As such, SMRT is being designed to indicate the extent to which specific reading skills have been mastered, rather than to differentiate or discriminate between children. Consequently, resulting subtest scores will reflect mastery or competence. This is in contrast to norm-referenced scores such as grade equivalents, normal curve equivalents (NCE's) and percentiles which can be misleading and are susceptible to misinterpretation. It is proposed that the SMRT mastery scores identify separately reported and potentially diagnostic dimensions including word attack, word meaning, literal comprehension, and reasoning comprehension.

The currently available partially validated research version of this test is designed to identify reading subtest areas in which either small instructional groups, intact classes or the entire grade four in particular schools are not achieving mastery. The short-term objective is to develop an instructionally useful grade four reading test. After its usefulness in grade four has been established, the development of either a grade three or grade seven school mastery of reading test will be considered. It is anticipated that fall-administered SMRT tests would be most useful to schools for instructional purposes. It would be possible, also, to administer SMRT at various subsequent times throughout the school year to assess progress.

Need for SMRT-STEPS is particularly timely in light of requirements of Part 100 of Commissioner of Education Regulations (New York State Education Department, 1984, 1985). These regulations initiate an innovative Comprehensive Assessment Report (CAR) which summarizes state testing program results, in addition to other school data (e.g. enrollment numbers, graduation results, attendance and dropout rates). Based upon the CAR, 393 New York City schools (237 elementary, 102 junior high/intermediate and 54 high schools have been identified by the New York State Education Department as in need of improvement.

II. UNIQUE AND INNOVATIVE ASPECTS OF THE SCHOOL MASTERY OF READING TEST SYSTEM TO ENHANCE THE PROGRESS OF SCHOOLS

This school improvement system will be characterized by the following unique and innovative aspects:

1) It is being developed by a consortium comprised of the New York City Board of Education and the Educational Testing Service of Princeton, New Jersey. In addition to providing a technically sound and useful system, the public schools will not have to pay royalties to a test publisher for the diagnostic part of the system.

2) A professional panel of New York City school administrators, teachers, reading experts and curriculum specialists has reviewed SMRT for appropriateness, usefulness and potential bias. This panel will continue to be involved in the program in order to review and establish the relationship between assessment and school improvement materials, plans and programs.

3) Common scaling between SMRT and National Assessment of Educational Progress (NAEP) is being established. It is anticipated that this will provide both cost-effective access to NAEP items and interpretation of SMRT results with respect to NAEP national norms.

4) The diagnostic component provides an objective test of mastery rather than a norm-referenced test. As such, it is designed to assess reading proficiency and provides a relatively sensitive measure of instruction.

5) The diagnostic component provides instructionally useful subscale scores to identify specific reading skills for diagnostic-prescriptive school improvement purposes.

6) It is our intention to assess the feasibility of employing advanced computer technology and state-of-the-art psychometric techniques in the development and delivery of the diagnostic component and, also, in the linkage between assessment and school improvement materials, plans and programs.

7) It is our eventual intention to design meaningful and useful reports of test results. Furthermore, the feasibility of relating subtest score profiles to prescriptive choices or menus of school improvement materials, plans and programs will be explored.

III. THE SCHOOL MASTERY OF READING TEST (SMRT)

The primary short term objective is to develop SMRT as a standardized measure of reading performance which can be readily administered and scored on a large scale and which accurately reflects multiple skills involved in reading. During May, 1986 SMRT was administered to 889 students.

SMRT is divided into two 50 item parts administered with a brief intervening intermission. Approximately 34 minutes are required for administration of each of the two parts for a total testing time of approximately 68 minutes. In general, there was sufficient time for most students to finish. In the future, however, serious consideration may be given to administering SMRT as a "power" rather than as a "speed" test.

The current 100 item research version of SMRT consists of four subtests including: Word Attack (18 items), Word Meaning (21 items), Literal Comprehension (31 items) and Reasoning Comprehension (27 items). When scored, SMRT provides four subtest scores and one total test score. Three additional Word Recognition items appear at the beginning of the test. These low difficulty items are used to orient students to test directions, format and the separate answer sheet. In addition, they begin students on a positive note in that they are relatively easy items. It is noted that there are a small number of additional SMRT items, including some cloze comprehension items, which were eliminated from the current test in order to limit the amount of time required for test administration. These additional items remain part of the available item bank. Descriptions of the different categories of items are provided in Attachment # 1.

In the current 100 item test, three items are used as "examples" to illustrate directions. These include two Word Recognition and one Word Attack item. In effect, students are told the correct answer after they attempt to respond.

Directions for the most part were read to students. Incorporated within the remaining 97 items are 16 items obtained from the National Assessment of Education Progress (NAEP). Of these 16, ten are Literal Comprehension and the remaining six are Reasoning Comprehension items. The reasons for imbedding NAEP items within SMRT are discussed later in the section entitled: "National Assessment of Education Progress and The School Mastery of Reading Test."

The SMRT booklet is not comprised of clearly defined subtests. Rather, items from the various subtests appear in both parts one and two. Furthermore, the actual tasks required of students change frequently.

0294U/13

IV. TEST ADMINISTRATION

Three Schools in each of three Brooklyn Community School Districts participated in the testing program. Each school previously had been identified by the New York State Education Department's Comprehensive Assessment Report as in need of improvement based upon their grade three Degrees of Reading Power test results. All participating schools were selected from within the borough of Brooklyn for logistical, control and test security reasons.

Profiles of each of the nine participating schools are presented in Table 2. Included in these profiles are summaries of student and school characteristics for the following three school years: 1982-1983; 1983-1984; 1984-1985

During the second and third weeks of May 1986, a total of 889 fourth grade students were tested. As depicted in Table 1, these students represented approximately 90% of those for whom answer sheets were provided. In Table 1, "Number of Absentees" refers to those students for whom identification information was provided, but item responses did not appear on their respective answer sheets. In other words, those students appear to have been absent for the test administration.

Schools were requested to provide answer sheets for every student who was eligible for the annual citywide reading test, with the exception of those Special Education students for whom some testing variance (e.g. large print, extended time limits, etc.) was required. Limited English Proficient students exempted from the annual citywide reading test also were exempted from SMRT. In order to minimize disruption of instruction, provision was not made for "make-up" testing of absentees.

Test booklets and administration manuals were delivered and retrieved from all nine participating schools by Functional Marketing Service, the same company that transports citywide test materials. The following schedule was followed:

1) Monday, May 12 - Cartons were obtained from 110 Livingston Street, Room 714 and delivered the same day to all nine schools.

2) Friday, May 16 - Cartons were retrieved from three schools in district #17 and another three schools in district #19 and subsequently delivered to 110 Livingston Street, Room 714.

3) Friday, May 23 - Cartons were retrieved from three schools in district #21 and delivered to 110 Livingston Street, Room 714.

To ensure test security, each test administration manual and test booklet was stamped with a unique identification number (See Attachment # 2). Careful track was kept of the range of numbers on both administration manuals and test booklets delivered to, and retrieved from, every school.

All test materials were delivered in strong cartons carefully sealed with white tape with the following message in red letters: "SECURE TEST MATERIALS - DO NOT OPEN." Cartons were delivered directly to the Principal's office and receipts were signed. The sealed cartons were then placed in secure storage closets, usually in the Principal's office.

On the day of testing, program staff visited each school. These visitors retrieved the sealed cartons from locked closets and distributed test materials to participating classes. A careful accounting was maintained of the quantity and identification numbers of both test administrative manuals and test booklets delivered to, and subsequently retrieved from, each class.

Program staff monitored the test administration in each school. All tests were administered by fourth grade teachers using the test administration manuals prepared for that purpose. Appropriate signs (See Attachment # 3 for sample) were placed on the door of each class indicating that "Testing" was being conducted. Students read the test questions from their test booklet, then responded on the separate answer sheet (See Attachment # 4) provided for that purpose. After testing, each teacher was asked to complete a one-page survey (See Attachment # 5) designed to assess their opinions of the test and testing procedures.

0294U/8 & 9

Table 1

NUMBER OF FOURTH GRADE STUDENTS TESTED

COMMUNITY SCHOOL DISTRICT	SCHOOL NUMBER	DATE TESTED	NUMBER OF ANSWER SHEETS	NUMBER OF ABSENTEES	PERCENTAGE OF STUDENTS TESTED	PERCENTAGE OF ABSENTEES
17	191	5/13/86	116	6	95%	5%
	289	5/14/86	131	12	92%	8%
	398	5/13/86	176	11	94%	6%
District 17 Subtotals:			423	29	94%	6%
19	213	5/14/86	111	14	89%	11%
	290	5/15/86	114	11	91%	9%
	328	5/15/86	70	8	90%	10%
District 19 Subtotals:			295	33	90%	10%
21	90	5/21/86	53	14	79%	21%
	212	5/22/86	70	12	85%	15%
	329	5/21/86	48	9	84%	16%
District 21 Subtotals:			171	35	83%	17%
Totals:			889	97	90%	10%

0294U/6

SCHOOL PROFILES

Definitions for Table 2

Listed below are definitions of the sub-headings used in the tables on the following pages (source: School Profiles, New York City Board of Education).[1]

Admissions - The number of pupils admitted into a school during the 1984-1985 school year divided by the average daily register for this year.

Discharges (Departures) - The number of pupils leaving a school during the 1984-1985 school year divided by the average daily register for that year.

Limited English Proficiency - This number is based on the Language Assessment Battery; if this is not available it is based on school entitlement. It is the number of Limited English Proficient students divided by the register, multiplied by 100.

Special Education - The number of Special Education students divided by the register, multiplied by 100.

Utilization - A measure of the usage of a school building in relation to it's rated capacity.

Percent Attendance (Percent of Average Daily Register) - The mean pupil attendance rate for the 1984-1985 school year.

Degrees of Reading Power Test (DRP) - The official citywide reading test for grades Three through Nine beginning April 1986.

California Achievement Test (CAT) - The official citywide reading test from 1978 to 1985.

---

[1] See, for example, New York City Board of Education School Profiles; 1982-1983. Brooklyn, New York, 1984.

Table 2

SCHOOL PROFILES

1982 - 1983 SCHOOL YEAR - PERCENT

Community School District	School	Admissions	Discharges	Limited English Proficiency	Special Education	Utilization	Attendance
17	P.S. 191	59.0	38.7	0.0	9.9	100	81.9
17	P.S. 289	44.3	28.1	0.1	12.7	87.0	86.0
17	P.S. 398	40.5	21.8	2.7	4.6	95.0	88.4
19	P.S. 213	48.7	33.1	11.5	5.9	97.0	88.4
19	P.S. 290	43.7	28.1	21.2	0.0	114.0	87.0
19	P.S. 328	37.6	24.7	22.1	17.9	60.0	87.2
21	P.S. 90	42.7	26.8	3.5	11.7	63.0	86.9
21	P.S. 212	24.8	24.8	5.3	7.7	78.0	88.3
21	P.S. 329	44.7	29.8	14.4	13.5	57.0	86.6

Table 2 continued...

SCHOOL PROFILES

1982 - 1983 SCHOOL YEAR - PERCENT AT & ABOVE GRADE LEVEL

Community School District	School	DEGREES OF READING POWER TEST (Grades) Three	Four	Six	CALIFORNIA ACHIEVEMENT TEST (Grades) Two	Three	Four	Five	Six	All Grades
17	P.S. 191	22.9	23.4	23.8	17.6	20.9	25.0	27.2	30.2	23.0
17	P.S. 289	48.0	38.5	--	21.2	42.9	28.6	60.3	--	37.4
17	P.S. 398	30.1	40.6	43.6	48.6	42.1	39.3	36.1	44.7	42.4
19	P.S. 213	53.8	53.5	53.8	32.1	54.0	40.8	56.6	55.9	47.4
19	P.S. 290	43.8	52.0	--	28.1	50.0	41.3	54.1	--	42.3
19	P.S. 328	26.3	42.6	30.3	29.6	23.4	20.3	24.2	30.6	25.8
21	P.S. 90	53.6	61.9	61.7	32.3	43.4	56.9	51.7	61.5	48.8
21	P.S. 212	39.6	52.4	74.2	40.5	44.1	49.1	65.4	70.7	53.0
21	P.S. 329	53.4	60.0	25.4	31.6	62.5	45.8	56.9	37.7	47.9

Table 2 continued...

SCHOOL PROFILES

1983 - 1984 SCHOOL YEAR - PERCENT

Community School District	School	Admissions	Discharges	Limited English Proficiency	Special Education	Utilization	Attendance
17	P.S. 191	46.1	34.0	2.1	11.6	92.0	81.9
17	P.S. 289	56.4	28.8	4.1	12.8	125.0	86.0
17	P.S. 398	35.9	26.2	3.0	4.7	109.0	88.4
19	P.S. 213	41.1	26.1	14.0	6.8	98.0	88.4
19	P.S. 290	48.5	27.1	31.9	0.0	120.0	86.1
19	P.S. 328	37.5	24.5	27.5	19.2	58.0	86.9
21	P.S. 90	42.5	25.7	9.3	9.3	70.0	86.9
21	P.S. 212	35.9	20.0	6.1	7.9	78.0	88.2
21	P.S. 329	48.9	29.5	14.9	11.1	77.0	86.6

Table 2 continued...

SCHOOL PROFILES

1983 - 1984 SCHOOL YEAR - PERCENT AT & ABOVE GRADE LEVEL

Community School District	School	DEGREES OF READING POWER TEST (Grades) Three	Four	Six	CALIFORNIA ACHIEVEMENT TEST (Grades) Two	Three	Four	Five	Six	All Grades
17	P.S. 191	23.3	42.1	24.4	9.4	26.5	35.7	32.5	--	23.4
17	P.S. 289	28.4	54.3	--	52.8	35.1	42.8	73.6	--	50.0
17	P.S. 398	39.8	43.4	43.6	30.5	25.9	42.7	36.4	40.9	34.9
19	P.S. 213	48.8	44.1	52.0	31.6	34.1	45.3	51.2	40.6	40.2
19	P.S. 290	46.8	60.7	--	36.1	43.3	53.8	45.9	--	44.6
19	P.S. 328	43.2	55.7	48.2	48.3	26.1	51.4	29.8	36.2	38.7
21	P.S. 90	46.3	58.5	74.2	37.3	42.6	55.9	63.0	59.4	51.1
21	P.S. 212	46.7	54.9	78.7	53.2	53.3	46.6	56.9	68.1	54.9
21	P.S. 329	44.4	69.6	51.7	42.5	39.7	39.2	41.4	51.7	42.7

Table 2 continued...

SCHOOL PROFILES

1984 - 1985 SCHOOL YEAR - PERCENT

Community School District	School	Admissions	Discharges	Limited English Proficiency	Special Education	Utilization	Attendance
17	P.S. 191	44.2	25.0	5.0	11.8	97.0	85.0
17	P.S. 289	49.2	24.0	4.8	12.1	131.0	88.2
17	P.S. 398	33.7	25.0	4.5	4.2	106.0	89.5
19	P.S. 213	42.0	25.6	18.2	5.8	97.0	87.7
19	P.S. 290	44.8	32.4	38.7	1.3	131.0	87.0
19	P.S. 328	29.9	21.0	28.1	19.6	54.0	87.8
21	P.S. 90	34.4	22.6	19.1	9.1	65.0	87.4
21	P.S. 212	32.8	18.2	10.1	5.9	74.0	89.4
21	P.S. 329	48.4	26.5	24.3	11.5	78.0	87.1

Table 2 continued...

SCHOOL PROFILES

1984 - 1985 SCHOOL YEAR - PERCENT AT & ABOVE GRADE LEVEL

Community School District	School	DEGREES OF READING POWER TEST (Grades) Three	Four	Six	CALIFORNIA ACHIEVEMENT TEST (Grades) Two	Three	Four	Five	Six	All Grades
17	P.S. 191	49.3	43.6	49.1	31.8	36.7	43.9	35.0	56.0	39.7
17	P.S. 289	47.7	53.2	--	59.1	58.7	50.8	80.2	--	61.8
17	P.S. 398	42.5	55.9	46.7	33.3	45.2	44.1	47.1	52.0	43.5
19	P.S. 213	47.1	59.6	30.0	30.9	38.6	53.6	58.7	56.1	46.7
19	P.S. 290	41.4	58.3	--	23.0	44.4	48.7	60.2	--	43.3
19	P.S. 328	48.1	49.3	37.7	33.8	47.0	34.8	55.8	22.2	38.9
21	P.S. 90	56.7	50.9	60.7	37.3	33.9	56.5	60.0	67.3	50.3
21	P.S. 212	63.0	43.5	72.0	45.9	40.5	46.6	60.0	71.8	53.8
21	P.S. 329	57.1	42.1	42.9	25.8	38.6	38.1	57.4	42.6	39.7

V. MACHINE SCORING OF STUDENT ANSWER SHEETS

Student answer sheets were scanned and scored by the Office of Educational Assessment Scan Center at 49 Flatbush Avenue, Brooklyn, New York. The machine scannable general purpose NCS answer sheets (i.e., NCS Trans-Optic EB08-4521:223222) were scanned on an NCS 7018 Optical Mark Reader with NCS Scanpak "Test Scoring Package" software.

To ensure the accuracy of results, the following manual quality control procedures were instituted :

Preparation of Machine Readable Answer Sheets

1. Inserted a school code for each answer sheet
2. Separated the answer sheets of the absentees
3. Checked the Name, NYC Identification Number, Date of Birth, and Class Number for completeness. In the instances where the nine digit NYC student Identification Numbers on the answer sheet was either incomplete, missing or was not the same as that on the Office of Student Information Services (OSIS) Student Bio-File Roster, the school was telephoned. In every instance, the identification number accepted as final was that provided by the school. However, for a small number of students who were either in Promotional Gates classes or whose name did not appear on the OSIS Student Bio-File roster, it was not possible to cross-check the accuracy of the identification number corresponding to the names.
4. For all the items mentioned in the previous step, verified that the circles corresponded with the characters
5. Tidied answer sheets -
   a. Erased unnecessary marks
   b. Shaped circles that extended to other areas
   c. Darkened light circles
   d. Identified and rectified circles which were only partially filled
   e. When a light circle (e.g.,erased circle) was adjacent to a dark circle (intended circle) the circles were lightened and darkened respectively
   f. Any torn answer sheet was replaced by a fresh answer sheet, appropriately completed

Scanning

1. To evaluate the performance of the scanner under varied conditions, two dummy answer sheets were prepared and scanned. The features of the dummy sheets were as follows:

SHEET 1	SHEET 2
Clean sheet	Untidy sheet
Included answer choice "e"	Included answer choice "e"
Included missing values	Included missing values
Included multiple answers	Included multiple answers
	Simulated erased circles

2. The results of the dummy answer sheets and one additional answer sheet were hand-scored and compared with the score obtained by scanning.

Scan Tape and Printout

A printout was produced from the magnetic tape resulting from the scanning procedure. This printout was reviewed to be certain that :

a. Both schools and classes were grouped together, as required.
b. The name, identification number appeared to be complete
c. The date of birth, identification number and class number appeared in the required field

Computer Scoring

1. The answers that were scanned and stored on magnetic tape as a data file, were read and scored using a SPSS-X program which generated results in the form of Five Subtest Scores and a Total Test Score. These results were compared with the scores generated by the scanning program.

2. In running the SPSS-X program, a few errors were detected in the form of incomplete names, identification numbers, dates of birth and class numbers. This provided a final check and improved the completeness of the biographical data fields.

VI. RESULTS FROM THE SPRING 1986 SCHOOL MASTERY OF READING TEST

Preliminary SMRT descriptive statistics from the May 1986 test administration are presented in Table 3. Total test score summaries are presented for the nine participating schools in three school districts. These results are depicted also in Figures 1 and 2. Specifically, Figure 1 depicts percentages for all nine schools. Figure 2 depicts percentages for the three school districts. Results for each of the subtests, in addition to the total test, are presented in Table 4. In addition to the summary statistics, it is noted that additional individual student listings including subtest and total test scores were generated by class and grade.

The overall mean total test score was 80.54 with a standard deviation of 13.49. The overall median total test score was 84 with an inter-quartile range of 16. In effect, these statistics reflect a negatively skewed distribution. In other words, this is a relatively easy test with a "piling up of scores" at the high end of the score distribution.

This distribution was expected for a curriculum-based test, such as SMRT, administered at the end of the school year. It is likely that the relatively high overall scores reflect mastery, at least to some extent, of fourth grade curriculum taught during the school year.

Results for both boys and girls were examined. Of the 889 students tested, 447 girls achieved a mean score of 81.08 (with a standard deviation of 12.78) and 442 boys achieved a mean score of 80.00 (with a standard deviation of 14.16)

0294U/11

Table 3

SMRT SUMMARY STATISTICS

DISTRICT SCHOOL	NUMBER OF STUDENTS	MEAN	STANDARD DEVIATION	MEDIAN	INTER QUARTILE RANGE
CSD 17	423	80.63	13.53	83	15
PS 191	116	80.96	12.98	84	13
PS 289	131	82.17	13.75	85	14
PS 398	176	79.27	13.67	81	15
CSD 19	295	79.05	14.37	83	18
PS 213	111	77.41	15.70	81	22
PS 290	114	78.63	13.70	81	17
PS 328	70	82.33	12.80	86	14
CSD 21	171	82.91	11.38	86	13
PS 90	53	85.11	9.81	87	11
PS 212	70	79.87	13.57	81	19
PS 329	48	84.90	8.26	87	11
TOTAL	889	80.54	13.49	84	16

Figure 1

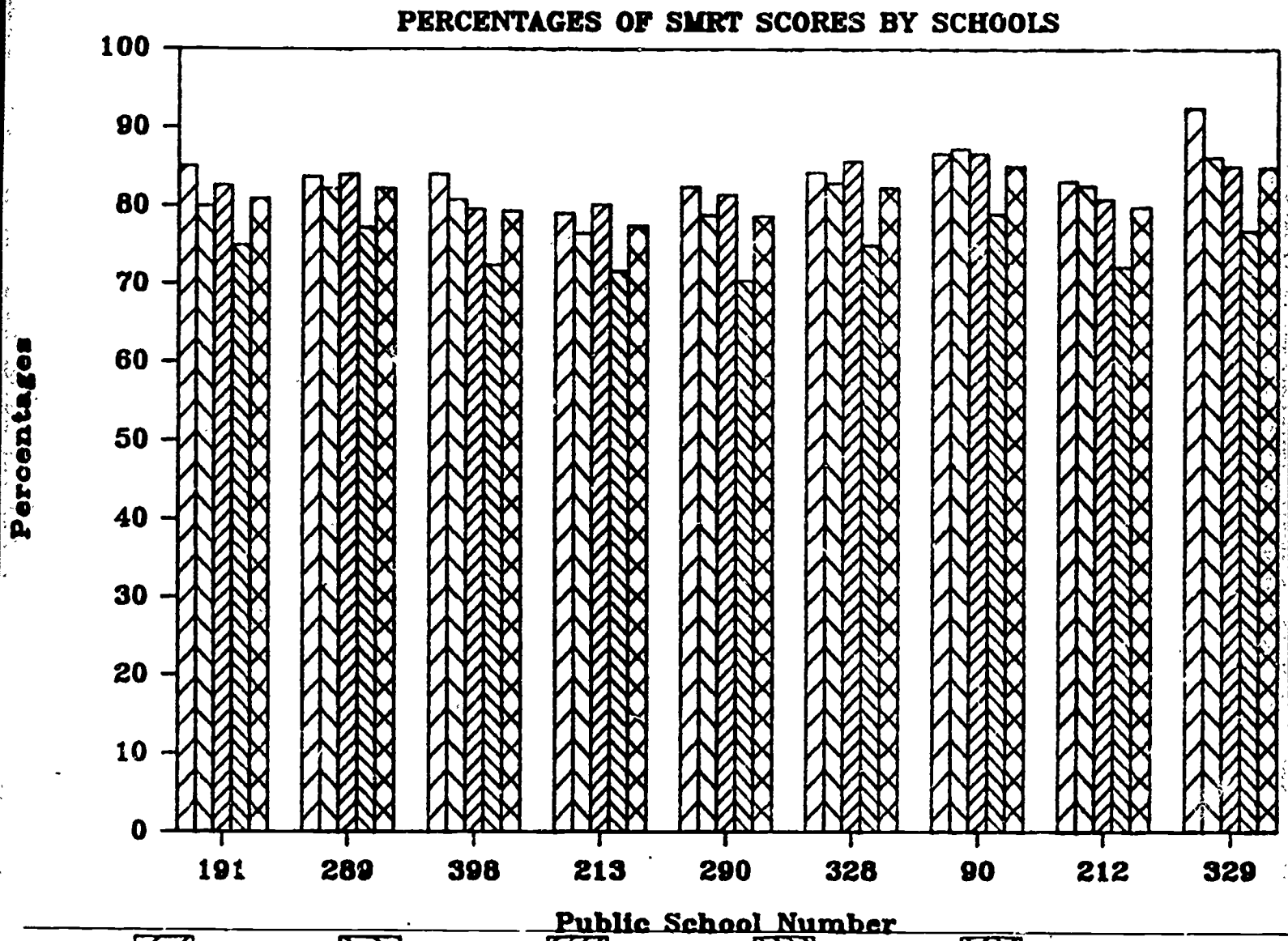

PERCENTAGES OF SMRT SCORES BY SCHOOLS
Percentages
100
90
80
70
60
50
40
30
20
10
0
191
289
398
213
290
328
90
212
329
Public School Number
Key:
WORD ATTACK
WORD MEANING
LITERAL COMPREHENSION
REASONING COMPREHENSION
TOTAL

Figure 2

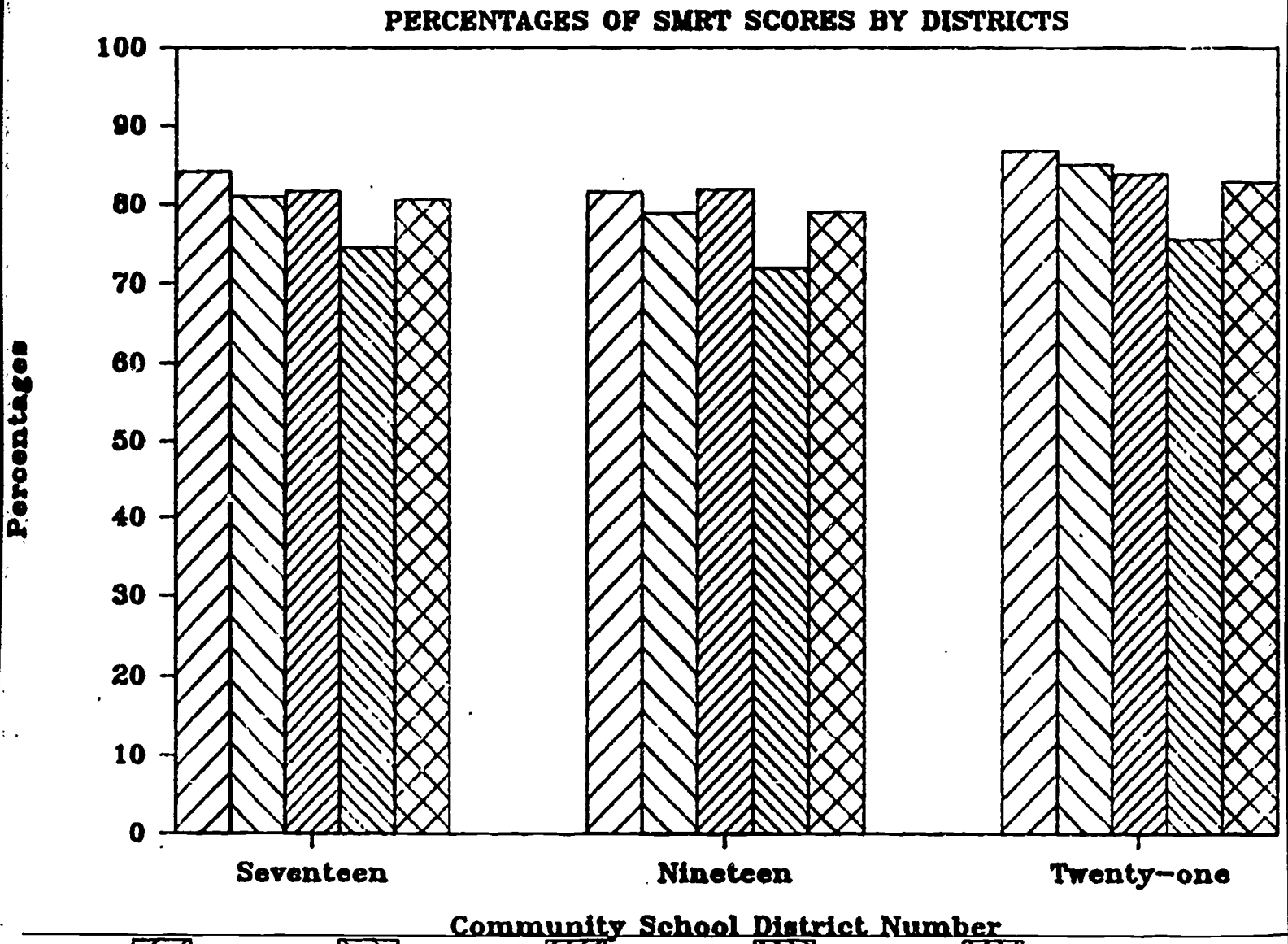

PERCENTAGES OF SMRT SCORES BY DISTRICTS
100
90
80
70
60
50
40
30
20
10
0
Percentages
Seventeen
Nineteen
Twenty-one
Community School District Number
Key:
WORD ATTACK
WORD MEANING
LITERAL COMPREHENSION
REASONING COMPREHENSION
TOTAL

Table 4

SPRING 1986 SMRT RAW SCORE RESULTS

(Number of Students = 889)

	WORD RECOGNITION (i=3)	WORD ATTACK (i=18)	WORD MEANING (i=21)	LITERAL COMPREHENSION (i=31)	REASONING COMPREHENSION (i=27)	TOTAL SCORE (i=100)
MEAN	2.97	15.09	17.02	25.49	19.97	80.54
STANDARD DEVIATION	0.20	2.90	3.55	4.58	4.25	13.49
MEDIAN	3	16	18	27	21	84
INTER-QUARTILE RANGE	0	3	5	6	5	16
RANGE	3	16	21	31	27	94
MAXIMUM	3	18	21	31	27	99
MINIMUM	0	2	0	0	0	5

0294U

VII. NORMS TABLE FOR THE SPRING 1986 SCHOOL MASTERY OF READING TEST

Table 5 provides an example of "local" SMRT norms which eventually could be developed for New York City. This table was based upon student scores from all participating nine schools and was developed for illustrative purposes. It is important to note that percentile scores are most appropriate for norm-referenced interpretations of test results. The primary intention is to develop SMRT as a test which reflects mastery or competence rather than as a norm-referenced test. Primary effort, therefore, is being placed on developing mastery scores. Percentiles may be used, subsequently, in a secondary or ancillary capacity to mastery scores.

Table 5 provides the basis for any subtest or total test raw score (i.e., number of correct test items) to be interpreted relative to overall group performance. In other words, any SMRT subtest or total test raw score can be converted into either a percentile or stanine by using this "raw score to percentile rank and stanine norms" table.

Percentiles allow us to compare any student's performance at a given time with that of the overall group on a scale ranging from 1 to 99. On this scale, the fiftieth percentile, or median, might be considered "average." Use of this table may be illustrated by locating a raw score of 90. This raw score corresponds to the seventy-sixth percentile. This seventy-sixth percentile means that the student's score exceeded those scores of approximately 76 percent of the norm group for that time of the year. In this case, the norm group is comprised of 889 students who were administered SMRT during May 1986 in nine New York City public schools located in three Community School Districts.

For total test scores, percentiles may be instructionally useful and assist in the interpretation of raw scores. For reporting subtest results, however, percentiles are less useful. In this report, percentiles have been provided for each of the subtests for illustrative purposes only. The use of percentiles for reporting subtest results is not recommended because they may be misleading and subject to misinterpretation. Due to the relatively small number of test items within any given subtest, the change of percentiles from one raw score to another may be relatively large. Consequently, these percentiles for subtests may erroneously appear to reflect significant and meaningful differences in test performance from one raw score to an adjacent raw score. In reality, the difference in actual performance from one raw score to another may be relatively modest. For example, examination of the norms table for the "Word Attack" subtest reveals that a change in raw score from 17 to 18 correct corresponds to a change from the 77th to the 99th

percentile. In effect, a change of one raw score corresponds to a 22 point change in percentiles.

This table presents, also, stanine scores. Stanine scores (i.e., a contraction of "standard nine") are normally distributed standard scores which range from one to nine. They have the convenience of being a single digit and are relatively easy to interpret. The stanine scale has a mean of five and a standard deviation of 1.96.

Table 5

SMRT RAW SCORES, PERCENTILES AND STANINES

STANINES	PERCENTILES	RAW SCORES					
		W.R.	W.A.	W.M.	L.C.	R.C.	TOTAL
	99	3	18	21	31	27	98
9	98					26	
	97						97
	96						
	95					25	
	94						96
8	93				30		
	92						95
	91						
	90						94
	89						
	88			20			
	87						93
	86					24	
	85						
7	84						92
	83						
	82				29		
	81						91
	80						
	79					23	
	78						
	77		17				
	76						90
	75						
	74						
	73			19			
	72						89
	71				28		
	70					22	
6	69						
	68						
	67						88
	66						
	65						
	64						
	63						87
	62						
	61						
	60						86

W.R. = Word Recognition
W.A. = Word Attack
W.M. = Word Meaning
L.C. = Literal Comprehension
R.C. = Reasoning Comprehension

Table 5 continued

SMRT RAW SCORES, PERCENTILES AND STANINES

STANINES	PERCENTILES	RAW SCORES W.R.	W.A.	W.M.	L.C.	R.C.	TOTAL
	59		16		27		
	58					21	
	57						
	56			18			85
	55						
	54						
	53						84
	52						
	51						
5	50						
	49						83
	48				26		
	47					20	
	46						82
	45		15				
	44			17			
	43						81
	42						
	41						
	40						
	39				25		80
	38						
	37					19	
	36						79
	35						
	34						78
	33		14	16			
4	32						
	31						
	30				24	18	77
	29						
	28						76
	27						
	26						75
	25			15			74
	24				23	17	
	23		13				
	22						73
3	21						
	20				22		72

Table 5 continued

SMRT RAW SCORES, PERCENTILES AND STANINES

STANINES	PERCENTILES	RAW SCORES					TOTAL
		W.R.	W.A.	W.M.	L.C.	R.C.	
	19			14		16	71
	18						70
	17				21		
	16		12				69
3	15			13		15	68
	14						67
	13				20		66
	12		11				65
	11						63
	10			12	19	14	62
	9						60
2	8		10		18	13	58
	7			11	17		57
	6						55
	5		9	10	16	12	
	4		8	8	15	11	50
1	3			7		10	47
	2	2			12	9	43
	1				11	8	35

VIII. TYPES OF SCORES USED TO REPORT TEST RESULTS

In anticipation of the eventual need to develop a meaningful and instructionally useful method of reporting test results, the Preliminary Professional Panel was asked to rate their opinion of the usefulness of the five types of test scores (see Attachment # 6). Definitions of the five different types of test scores are provided in Attachment # 7. A summary of these ratings is provided in Table 6. It is apparent that the panel of professional educators considered percentile rankings as the most instructionally useful type of test score.

In general, the types of scores referred to above reflect relative rather than absolute performance. In effect, they provide norm-referenced interpretations of test results. In contrast, it is anticipated that the most instructionally useful and appropriate type of score will reflect mastery of reading. In other words, criteria, standards or "passing scores" should be developed for item clusters, subtests and, perhaps, for the total test. Subsequently, these "benchmarks" will be used to determine if results reflect mastery. Such indices reflect absolute performance and are most appropriate for criterion-referenced or mastery tests such as SMRT.

It should be noted that alternate methods of reporting test results to enhance school progress may be developed. These methods might, for example, be based upon the diagnostic-prescriptive relationship between SMRT assessment and school improvement plans, programs and materials.

0294U/10

Table 6

PROFESSIONAL PANEL MEMBER OPINIONS OF THE USEFULNESS OF VARIOUS TYPES OF TEST SCORES

	NUMBER OF RESPONDENTS	MEAN	S.D	PERCENTAGES OF RATINGS												
				LOW 1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	HIGH 7
PERCENTILE RANKINGS	17	5.65	0.98	0	0	0	5.9	0	0	0	0	23.5	5.9	52.9	0	11.8
GRADE EQUIVALENTS	17	5.35	1.46	5.9	0	0	0	0	0	17.6	0	17.6	5.9	29.4	5.9	17.6
RAW SCORES	17	4.65	1.45	0	5.9	0	0	23.5	0	5.9	5.9	17.6	11.8	23.5	0	5.9
STANINES	17	4.12	1.59	0	0	23.5	0	11.8	5.9	5.9	11.8	23.5	5.9	0	0	11.8
NORMAL CURVE-EQUIVALENTS	16	2.41	1.28	25	6.3	25	12.5	12.5	0	6.3	0	12.5	0	0	0	0

IX. DEVELOPING MASTERY SCORES AND ESTABLISHING MINIMUM STANDARDS

In order to establish mastery criteria for each test item, both empirical data and judgments of experts are being reviewed. The eventual primary objective is to identify SMRT performance standards. It is likely that such standards will be based either upon item clusters or subtests, rather than upon individual items or total test scores.

Empirical data which have been obtained include:

1) SMRT scores of fourth graders in the nine schools tested during the second and third weeks in May 1986,

2) Degrees of Reading Power (DRP) test scores of the same students tested on May 7, 1986, and

3) Metropolitan Achievement Test (MAT) reading scores of the same students tested on both April 21 and 22, 1986.

Judgments of experts were provided by the Professional Panel. This group was comprised of professional educators including teachers, assistant principals, principals, district reading coordinators and curriculum supervisors. Panel members were asked to estimate the difficulty level of each item in the Group Test for each of the three hypothetical groups of students described below:

Group 1: Satisfactory readers. Students in this group
- a) read well enough to learn from fourth grade text material in reading and other subject areas,
- b) read well enough to follow instructions in workbooks, arithmetic problems, and other school work, and
- c) can be expected to continue to learn in the fifth grade.

Group 2: Minimally competent readers. Students in this group
- a) have developed sufficient reading skills that they can continue to learn to read, perhaps with special help,
- b) can be expected to have some difficulty with fourth grade text material, but can learn at a minimal level from such material, and
- c) can be expected to need continuing special help with basic reading skills in the fifth grade.

Group 3: Readers below minimum competence. Students in this group
- a) have not achieved some or all of the basic reading skills appropriate to fourth grade,
- b) cannot learn by reading fourth grade text material in reading and other subject areas,
- c) cannot read sufficiently well to follow directions in workbooks and arithmetic problems.

Both empirical and judgmental information will be reviewed carefully and alternative strategies for setting performance standards (see, for discussion, Livingston & Zieky, 1982) will be considered.

Table 7 presents Professional/Panel judgments or expectations of performance on all SMRT items for the three hypothetical groups of students described earlier. In addition, actual performance on all SMRT items for relatively low, moderate and relatively high achieving groups of students, as determined by their DRP test scores, is presented in Table 8. The three DRP score categories were established based upon the grade four DRP Promotional Gates criterion. Students achieving DRP scores within one standard error either below or above the fourth grade criterion were assigned to Group 2. Students achieving DRP scores lower than one standard error below the Promotional Gates criterion were assigned to Group 3. Students achieving DRP scores higher than one standard error above the Promotional Gates criterion were assigned to Group 1. Similar groupings can be established based upon Metropolitan Achievement Test scores.

Once reasonable performance expectations of items are established, standards for test objectives, item clusters and/or subtests can be set by aggregating items.

Additional information, such as frequency distributions of test scores, can be useful in setting standards. For example, SMRT subtest and total test scores can be grouped and depicted in various ways. Figure 3, for example, depicts SMRT subtest and total test raw scores in ten groups based upon Degrees of Reading Power (DRP) test performance. The distribution of DRP scores was divided into ten groups ranging from the lowest to the highest test scores. Subsequently, SMRT subtest and total test performance was depicted as histograms which illustrate the profile of SMRT scores for each DRP category. Using this information, performance standards could be established by picking the point between two specific groups of SMRT histograms. It is noted, in addition, that it is an interesting research question to determine if different school improvement or corrective action plans can be related to profiles of scores obtained by students of different achievement levels.

Table 7

PROFESSIONAL PANEL JUDGEMENTS OF TEST ITEM DIFFICULTY
FOR THREE HYPOTHETICAL GROUPS OF STUDENTS

What proportion of students do you think will get this item correct ?

QUESTION	GROUP	All or almost all (91% or more)	About three-quarters (61-90%)	About half (36-60%)	About one quarter (11-35%)	None or almost none (10% or less)
1	Group 1	---	---	(SAMPLE QUESTION)		---
	Group 2	---	---	---	---	---
	Group 3	---	---	---	---	---
2	Group 1	---	---	(SAMPLE QUESTION)		---
	Group 2	---	---	---	---	---
	Group 3	---	---	---	---	---
3	Group 1	84.2	15.8	0	0	0
	Group 2	21	63.2	15.8	0	0
	Group 3	5.3	5.3	52.6	26.3	10.5
4	Group 1	100	0	0	0	0
	Group 2	55	40	5	0	0
	Group 3	5	45	25	15	10
5	Group 1	85	15	0	0	0
	Group 2	20	55	25	0	0
	Group 3	5	5	40	30	20
6	Group 1	95	5	0	0	0
	Group 2	25	65	10	0	0
	Group 3	5	20	35	30	10
7	Group 1	75	25	0	0	0
	Group 2	20	45	35	0	0
	Group 3	5	20	10	45	20
8	Group 1	95	5	0	0	0
	Group 2	40	55	5	0	0
	Group 3	10	30	30	20	10
9	Group 1	---	---	(SAMPLE QUESTION)		---
	Group 2	---	---	---	---	---
	Group 3	---	---	---	---	---
10	Group 1	50	50	0	0	0
	Group 2	10	30	45	15	0
	Group 3	0	5	15	40	40

(continued in next page)

Table 7 continued...

PROFESSIONAL PANEL JUDGEMENTS OF TEST ITEM DIFFICULTY
FOR THREE HYPOTHETICAL GROUPS OF STUDENTS

What proportion of students do you think will get this item correct ?

QUESTION	GROUP	All or almost all (91% or more)	About three-quarters (61-90%)	About half (36-60%)	About one quarter (11-35%)	None or almost none (10% or less)
11	Group 1	70	25	5	0	0
	Group 2	10	35	50	5	0
	Group 3	0	5	15	45	35
12	Group 1	80	20	0	0	0
	Group 2	5	50	45	0	0
	Group 3	0	0	15	60	25
13	Group 1	85	10	5	0	0
	Group 2	25	50	20	5	0
	Group 3	0	5	35	40	20
14	Group 1	90	10	0	0	0
	Group 2	50	45	5	0	0
	Group 3	0	30	40	25	5
15	Group 1	95	5	0	0	0
	Group 2	15	70	15	0	0
	Group 3	0	10	20	50	20
16	Group 1	90	10	0	0	0
	Group 2	25	45	30	0	0
	Group 3	0	20	25	20	35
17	Group 1	100	0	0	0	0
	Group 2	50	35	15	0	0
	Group 3	0	40	20	30	10
18	Group 1	80	20	0	0	0
	Group 2	20	60	15	5	0
	Group 3	5	10	35	30	20
19	Group 1	95	5	0	0	0
	Group 2	25	70	5	0	0
	Group 3	0	5	50	35	10
20	Group 1	55	45	0	0	0
	Group 2	5	30	50	15	0
	Group 3	0	5	15	25	55

(continued in next page)

Table 7 continued...

PROFESSIONAL PANEL JUDGEMENTS OF TEST ITEM DIFFICULTY
FOR THREE HYPOTHETICAL GROUPS OF STUDENTS

What proportion of students do you think will get this item correct ?

QUESTION	GROUP	All or almost all (91% or more)	About three-quarters (61-90%)	About half (36-60%)	About one quarter (11-35%)	None or almost none (10% or less)
21	Group 1	90	10	0	0	0
	Group 2	25	55	20	0	0
	Group 3	5	10	15	55	15
22	Group 1	95	5	0	0	0
	Group 2	45	40	15	0	0
	Group 3	5	15	25	40	15
23	Group 1	40	60	0	0	0
	Group 2	5	30	55	10	0
	Group 3	0	0	10	20	70
24	Group 1	55	45	0	0	0
	Group 2	5	30	55	10	0
	Group 3	0	5	10	35	50
25	Group 1	85	15	0	0	0
	Group 2	25	60	10	5	0
	Group 3	5	5	35	45	10
26	Group 1	80	20	0	0	0
	Group 2	5	55	30	10	0
	Group 3	0	15	10	40	35
27	Group 1	70	30	0	0	0
	Group 2	15	40	45	0	0
	Group 3	5	10	20	25	40
28	Group 1	85	15	0	0	0
	Group 2	20	35	45	0	0
	Group 3	5	5	30	40	20
29	Group 1	80	20	0	0	0
	Group 2	20	40	40	0	0
	Group 3	0	10	35	30	25
30	Group 1	95	5	0	0	0
	Group 2	50	40	5	5	0
	Group 3	5	30	30	20	15

(continued i. next page)

Table 7 continued...

PROFESSIONAL PANEL JUDGEMENTS OF TEST ITEM DIFFICULTY
FOR THREE HYPOTHETICAL GROUPS OF STUDENTS

What proportion of students do you think will get this item correct ?

QUESTION	GROUP	All or almost all (91% or more)	About three-quarters (61-90%)	About half (36-60%)	About one quarter (11-35%)	None or almost none (10% or less)
31	Group 1	95	5	0	0	0
	Group 2	50	40	10	0	0
	Group 3	0	35	30	25	10
32	Group 1	60	40	0	0	0
	Group 2	0	35	55	10	0
	Group 3	0	0	15	30	55
33	Group 1	75	20	5	0	0
	Group 2	25	40	25	10	0
	Group 3	5	10	20	35	30
34	Group 1	65	35	0	0	0
	Group 2	5	45	25	25	0
	Group 3	0	0	10	20	70
35	Group 1	36.8	63.2	0	0	0
	Group 2	0	15.8	68.4	15.8	0
	Group 3	0	0	0	26.3	73.7
36	Group 1	55	45	0	0	0
	Group 2	0	35	50	15	0
	Group 3	0	0	5	35	60
37	Group 1	55	40	5	0	0
	Group 2	5	35	45	15	0
	Group 3	0	10	5	40	45
38	Group 1	30	60	10	0	0
	Group 2	5	10	55	30	0
	Group 3	0	0	5	25	70
39	Group 1	65	35	0	0	0
	Group 2	20	15	55	10	0
	Group 3	0	5	5	40	50
40	Group 1	50	45	5	0	0
	Group 2	0	35	55	10	0
	Group 3	0	0	5	40	55

(continued in next page)

Table 7 continued...

PROFESSIONAL PANEL JUDGEMENTS OF TEST ITEM DIFFICULTY
FOR THREE HYPOTHETICAL GROUPS OF STUDENTS

What proportion of students do you think will get this item correct ?

QUESTION	GROUP	All or almost all (91% or more)	About three-quarters (61-90%)	About half (36-60%)	About one quarter (11-35%)	None or almost none (10% or less)
41	Group 1	65	30	5	0	0
	Group 2	20	35	20	25	0
	Group 3	5	10	10	20	55
42	Group 1	80	20	0	0	0
	Group 2	35	35	30	0	0
	Group 3	5	10	30	30	25
43	Group 1	85	15	0	0	0
	Group 2	45	30	25	0	0
	Group 3	5	20	20	35	20
44	Group 1	60	40	0	0	0
	Group 2	15	45	25	15	0
	Group 3	0	15	15	30	40
45	Group 1	70	30	0	0	0
	Group 2	30	25	35	10	0
	Group 3	0	15	15	35	35
46	Group 1	65	35	0	0	0
	Group 2	25	35	35	5	0
	Group 3	0	10	20	30	40
47	Group 1	30	65	5	0	0
	Group 2	0	25	55	20	0
	Group 3	0	0	0	20	80
48	Group 1	30	70	0	0	0
	Group 2	0	30	50	20	0
	Group 3	0	0	0	25	75
49	Group 1	52.6	42.1	5.3	0	0
	Group 2	0	31.5	63.2	5.3	0
	Group 3	0	0	0	21.1	78.9
50	Group 1	35	65	0	0	0
	Group 2	0	20	65	15	0
	Group 3	0	0	0	20	80

(continued in next page)

Table 7 continued...

PROFESSIONAL PANEL JUDGEMENTS OF TEST ITEM DIFFICULTY
FOR THREE HYPOTHETICAL GROUPS OF STUDENTS

What proportion of students do you think will get this item correct ?

QUESTION	GROUP	All or almost all (91% or more)	About three-quarters (61-90%)	About half (36-60%)	About one quarter (11-35%)	None or almost none (10% or less)
51	Group 1	75	25	0	0	0
	Group 2	30	40	25	5	0
	Group 3	5	15	15	35	30
52	Group 1	80	20	0	0	0
	Group 2	35	20	35	10	0
	Group 3	5	15	20	45	15
53	Group 1	60	35	5	0	0
	Group 2	15	40	20	25	0
	Group 3	0	10	15	25	50
54	Group 1	45	50	5	0	0
	Group 2	0	40	45	15	0
	Group 3	0	0	5	40	55
55	Group 1	65	30	5	0	0
	Group 2	20	35	40	5	0
	Group 3	0	0	10	45	45
56	Group 1	85	15	0	0	0
	Group 2	35	25	40	0	0
	Group 3	5	15	15	40	25
57	Group 1	65	35	0	0	0
	Group 2	15	40	35	10	0
	Group 3	0	0	30	35	35
58	Group 1	65	35	0	0	0
	Group 2	0	55	25	20	0
	Group 3	0	0	20	25	55
59	Group 1	85	15	0	0	0
	Group 2	15	60	20	5	0
	Group 3	5	10	10	45	30
60	Group 1	85	15	0	0	0
	Group 2	35	35	30	0	0
	Group 3	0	20	5	50	25

(continued in next page)

Table 7 continued...

PROFESSIONAL PANEL JUDGEMENTS OF TEST ITEM DIFFICULTY
FOR THREE HYPOTHETICAL GROUPS OF STUDENTS

What proportion of students do you think will get this item correct ?

QUESTION	GROUP	All or almost all (91% or more)	About three-quarters (61-90%)	About half (36-60%)	About one quarter (11-35%)	None or almost none (10% or less)
61	Group 1	70	30	0	0	0
	Group 2	40	20	35	5	0
	Group 3	0	10	20	40	30
62	Group 1	90	10	0	0	0
	Group 2	20	50	30	0	0
	Group 3	0	5	25	50	20
63	Group 1	50	50	0	0	0
	Group 2	10	30	40	20	0
	Group 3	0	10	15	20	55
64	Group 1	30	60	10	0	0
	Group 2	0	15	50	35	0
	Group 3	0	0	0	35	65
65	Group 1	75	20	5	0	0
	Group 2	30	25	35	10	0
	Group 3	0	5	15	55	25
66	Group 1	65	30	5	0	0
	Group 2	15	25	40	20	0
	Group 3	0	10	5	35	50
67	Group 1	70	25	5	0	0
	Group 2	20	50	5	25	0
	Group 3	0	10	15	40	35
68	Group 1	50	45	5	0	0
	Group 2	15	30	30	25	0
	Group 3	0	0	15	30	55
69	Group 1	80	20	0	0	0
	Group 2	35	30	25	10	0
	Group 3	0	5	25	50	20
70	Group 1	55	35	10	0	0
	Group 2	10	30	45	15	0
	Group 3	0	10	10	25	55

(continued in next page)

Table 7 continued...

PROFESSIONAL PANEL JUDGEMENTS OF TEST ITEM DIFFICULTY
FOR THREE HYPOTHETICAL GROUPS OF STUDENTS

What proportion of students do you think will get this item correct ?

QUESTION	GROUP	All or almost all (91% or more)	About three-quarters (61-90%)	About half (36-60%)	About one quarter (11-35%)	None or almost none (10% or less)
71	Group 1	70	25	5	0	0
	Group 2	30	20	40	10	0
	Group 3	0	5	25	40	30
72	Group 1	60	30	10	0	0
	Group 2	10	25	45	20	0
	Group 3	0	0	10	35	55
73	Group 1	50	45	5	0	0
	Group 2	10	20	55	15	0
	Group 3	0	0	0	50	50
74	Group 1	60	35	5	0	0
	Group 2	10	45	35	10	0
	Group 3	0	0	10	45	45
75	Group 1	75	25	0	0	0
	Group 2	10	35	55	0	0
	Group 3	0	0	15	40	45
76	Group 1	70	30	0	0	0
	Group 2	20	35	35	10	0
	Group 3	0	0	20	55	25
77	Group 1	80	20	0	0	0
	Group 2	30	30	40	0	0
	Group 3	0	20	20	40	20
78	Group 1	55	45	0	0	0
	Group 2	20	35	15	30	0
	Group 3	0	0	35	25	40
79	Group 1	55	45	0	0	0
	Group 2	20	35	25	20	0
	Group 3	0	10	15	30	45
80	Group 1	70	30	0	0	0
	Group 2	20	30	40	10	0
	Group 3	0	10	20	35	35

(continued in next page)

Table 7 continued...

PROFESSIONAL PANEL JUDGEMENTS OF TEST ITEM DIFFICULTY
FOR THREE HYPOTHETICAL GROUPS OF STUDENTS

What proportion of students do you think will get this item correct ?

QUESTION	GROUP	All or almost all (91% or more)	About three-quarters (61-90%)	About half (36-60%)	About one quarter (11-35%)	None or almost none (10% or less)
81	Group 1	30	60	10	0	0
	Group 2	0	10.5	52.6	31.6	5.3
	Group 3	0	5	5	10	80
82	Group 1	20	65	10	5	0
	Group 2	5	5	55	25	10
	Group 3	0	5	0	20	75
83	Group 1	60	40	0	0	0
	Group 2	5	40	40	15	0
	Group 3	0	5	20	35	40
84	Group 1	60	35	5	0	0
	Group 2	15	35	35	15	0
	Group 3	0	10	0	45	45
85	Group 1	65	30	5	0	0
	Group 2	15	40	25	20	0
	Group 3	0	5	20	25	50
86	Group 1	55	40	5	0	0
	Group 2	10	30	50	10	0
	Group 3	0	5	15	35	45
87	Group 1	60	35	5	0	0
	Group 2	5	35	50	10	0
	Group 3	5	5	15	30	45
88	Group 1	100	0	0	0	0
	Group 2	50	30	20	0	0
	Group 3	15	5	35	45	0
89	Group 1	100	0	0	0	0
	Group 2	45	30	25	0	0
	Group 3	10	10	35	30	15
90	Group 1	100	0	0	0	0
	Group 2	40	40	20	0	0
	Group 3	5	20	30	35	10

(continued in next page)

Table 7 continued...

PROFESSIONAL PANEL JUDGEMENTS OF TEST ITEM DIFFICULTY
FOR THREE HYPOTHETICAL GROUPS OF STUDENTS

What proportion of students do you think will get this item correct ?

QUESTION	GROUP	All or almost all (91% or more)	About three-quarters (61-90%)	About half (36-60%)	About one quarter (11-35%)	None or almost none (10% or less)
91	Group 1	60	40	0	0	0
	Group 2	20	30	45	5	0
	Group 3	5	15	20	20	40
92	Group 1	75	25	0	0	0
	Group 2	15	55	25	5	0
	Group 3	0	15	20	25	40
93	Group 1	45	40	15	0	0
	Group 2	5	15	50	30	0
	Group 3	0	10	5	25	60
94	Group 1	85	15	0	0	0
	Group 2	35	40	25	0	0
	Group 3	5	15	30	30	20
95	Group 1	45	45	10	0	0
	Group 2	0	30	60	10	0
	Group 3	0	0	20	35	45
96	Group 1	65	30	5	0	0
	Group 2	25	30	40	5	0
	Group 3	0	5	30	30	35
97	Group 1	55	25	15	5	0
	Group 2	10	40	25	20	5
	Group 3	0	0	20	25	55
98	Group 1	50	40	10	0	0
	Group 2	10	35	40	15	0
	Group 3	0	5	5	30	60
99	Group 1	35	55	10	0	0
	Group 2	5	25	50	20	0
	Group 3	0	0	10	20	70
100	Group 1	26.3	42.1	31.6	0	0
	Group 2	0	21	42.1	31.6	5.3
	Group 3	0	0	0	33.3	66.7

Table 8

NUMBER AND PERCENT OF CORRECT RESPONSES TO EACH SCHOOL MASTERY OF READING TEST (SMRT) ITEM FOR THREE GROUPS OF STUDENTS DEFINED BY THEIR DEGREES OF READING POWER (DRP) TEST SCORES

(n = 744)

SMRT ITEM	Group 3 (low DRP scores) (n = 98)		Group 2 (marginal DRP scores) (n = 123)		Group 1 (high DRP scores) (n = 523)	
	Number	Percent	Number	Percent	Number	Percent
1	97	99.0	123	100.0	522	99.8
2	98	100.0	123	100.0	521	99.6
3	90	91.8	120	97.6	517	98.9
4	91	92.9	114	92.7	510	97.5
5	90	91.8	117	95.1	518	99.0
6	78	79.6	114	92.7	504	96.4
7	60	61.2	94	76.4	457	87.4
8	76	77.6	111	90.2	495	94.6
9	95	96.9	122	99.2	517	98.9
10	50	51.0	91	74.0	464	88.7
11	48	49.0	90	73.2	446	85.3
12	48	49.0	76	61.8	426	81.5
13	53	54.1	83	67.5	456	87.2
14	74	75.5	117	95.1	513	98.1
15	77	78.6	112	91.1	507	96.9
16	69	70.4	111	90.2	513	98.1
17	74	75.5	101	82.1	493	94.3
18	67	68.4	104	84.6	471	90.1
19	58	59.2	100	81.3	484	92.5
20	40	40.8	78	63.4	396	75.7
21	86	87.8	122	99.2	514	98.3
22	89	90.8	118	95.9	505	96.6
23	58	59.2	78	63.4	422	80.7
24	73	74.5	106	86.2	471	90.1
25	94	95.9	118	95.9	517	98.9
26	77	78.6	118	95.9	512	97.9
27	80	81.6	120	97.6	519	99.2
28	68	69.4	119	96.7	506	96.7
29	35	35.7	59	48.0	383	73.2
30	82	83.7	119	96.7	519	99.2
31	81	82.7	119	96.7	516	98.7
32	76	77.6	106	86.2	508	97.1
33	72	73.5	113	91.9	512	97.9
34	82	83.7	112	91.1	512	97.9
35	20	20.4	33	26.8	244	46.7
36	51	52.0	97	78.9	487	93.1
37	50	51.0	97	78.9	476	91.0

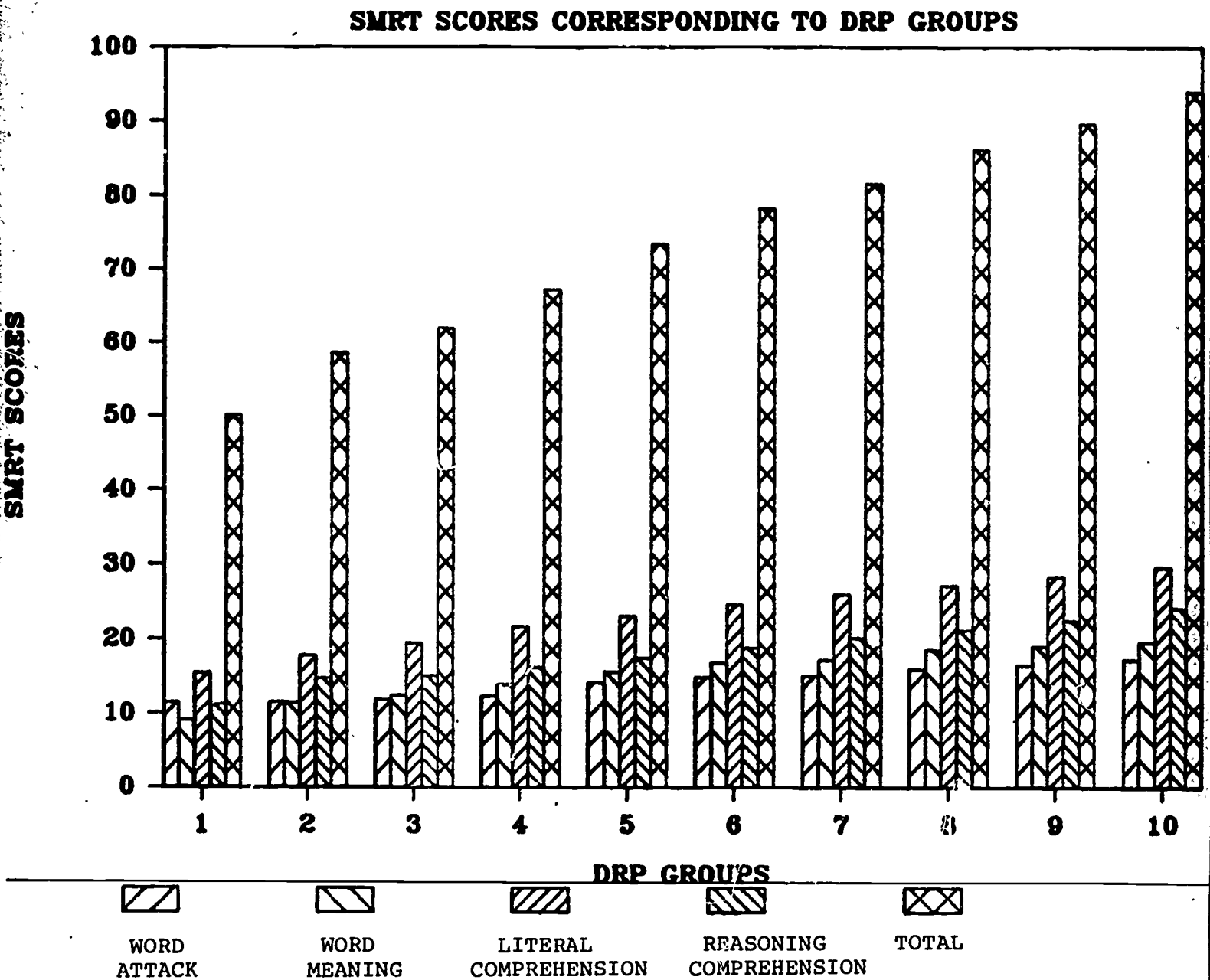
Table 8 continued...

SMRT ITEM	Group 3 (low DRP scores) (n = 98) Number	Percent	Group 2 (marginal DRP scores) (n = 123) Number	Percent	Group 1 (high DRP scores) (n = 523) Number	Percent
38	34	34.7	73	59.3	418	49.9
39	32	32.7	80	65.0	446	85.3
40	28	28.6	66	53.7	428	81.8
41	58	59.2	90	73.2	438	83.7
42	78	79.6	107	87.0	468	89.5
43	66	67.3	99	80.5	460	88.0
44	57	58.2	91	74.0	446	85.3
45	72	73.5	106	86.2	467	89.3
46	70	71.4	105	85.4	474	90.6
47	72	73.5	110	89.4	499	95.4
48	51	52.0	80	65.0	427	81.6
49	58	59.2	86	69.9	431	82.4
50	66	67.3	101	82.1	484	92.5
51	70	71.4	96	78.0	479	91.6
52	77	78.6	109	88.6	504	96.4
53	63	64.3	101	82.1	488	93.3
54	44	44.9	73	59.3	430	82.2
55	62	63.3	107	87.0	498	95.2
56	59	60.2	101	82.1	494	94.5
57	21	21.4	47	38.2	327	62.5
58	50	51 0	83	67.5	424	81.1
59	61	62.2	100	81.3	445	85.1
60	65	66.3	98	79.7	479	91.6
61	57	58.2	102	82.9	473	90.4
62	65	66.3	106	86.2	503	96.2
63	62	63.3	93	75.6	435	83.2
64	28	28.6	56	45.5	320	61.2
65	40	40.8	62	50.4	415	79.3
66	41	41.8	73	59.3	409	78.2
67	34	34.7	62	50.4	420	80.3
68	38	38.8	90	73.2	457	87.4
69	69	70.4	106	86.2	491	93.9
70	72	73.5	109	88.6	509	97.3
71	64	65.3	99	80.5	483	92.4
72	30	30.6	73	59.3	449	85.9
73	59	60.2	106	86.2	491	93.9
74	77	78.6	114	92.7	508	97.1
75	59	60.2	105	85.4	458	87.6
76	68	69.4	98	79.7	453	86.6
77	57	58.2	102	82.9	496	94.8
78	61	62.2	101	82.1	458	87.6
79	50	51.0	88	71.5	453	86.6
80	67	68.4	95	77.2	486	92.9
81	33	33.7	53	43.1	387	74.0
82	25	25.5	42	34.1	416	79.5

Table 8 continued...

SMRT ITEM	Group 3 (low DRP scores) (n = 98) Number	Percent	Group 2 (marginal DRP scores) (n = 123) Number	Percent	Group 1 (high DRP scores) (n = 523) Number	Percent
83	25	25.5	58	47.2	407	77.8
84	27	27.6	34	27.6	320	61.2
85	40	40.8	67	54.5	439	83.9
86	24	24.5	33	26.8	269	51.4
87	35	35.7	71	57.7	449	85.9
88	78	79.6	104	84.6	515	98.5
89	77	78.6	107	87.0	510	97.5
90	72	73.5	106	86.2	515	98.5
91	74	75.5	92	74.8	438	83.7
92	70	71.4	98	79.7	486	92.9
93	18	18.4	33	26.8	292	55.8
94	78	79.6	108	87.8	493	94.3
95	21	21.4	33	26.8	309	59.1
96	73	74.5	103	83.7	454	86.8
97	38	38.8	69	56.1	388	74.2
98	25	25.5	40	32.5	252	48.2
99	25	25.5	33	26.8	200	38.2
100	17	17.3	46	37.4	302	57.7

Figure 3


SMRT SCORES CORRESPONDING TO DRP GROUPS
SMRT SCORES
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
DRP GROUPS
WORD ATTACK
WORD MEANING
LITERAL COMPREHENSION
REASONING COMPREHENSION
TOTAL

X. THE PROFESSIONAL PANEL

In order to ensure the relevance and usefulness of SMRT-STEPS to classroom instruction and school improvement, a professional panel was established. This panel was comprised of school and Community School District office professional educators, as listed on the next page.

Among other activities, the professional panel reviewed the School Mastery of Reading Test for bias and appropriateness (see Attachment # 8), provided judgments and opinions related to student performance criteria, and recommended ways of improving the usefulness of test results to enhance school progress.

In order to establish the professional panel, each Community Superintendent was asked to recommend two district office staff. From each participating school, also, two professional staff members including at least one fourth grade teacher were invited to join the panel. Specifically, each principal was invited to attend. The principal had the option of not participating and recommending either an Assistant Principal or a fourth grade teacher in his or her stead. In addition to the principal or designee, a fourth grade teacher was recommended by the Principal in conjunction with that school's United Federation of Teacher (UFT) representative. Both the UFT and Council of Supervisor and Administrators (CSA) were apprised of activities related to the selection of panel members.

During spring 1986, the Professional Panel met twice at 110 Livingston Street. The agenda for both the May 8 and June 19 meetings can be found in the appendix (see Attachments # 9 and # 10). Remuneration was provided at prevailing rates for both teachers (i.e., $21.15 per hour) and supervisors (i.e., $23.80 per hour).

C296U/1

PROFESSIONAL PANEL
Spring, 1986

Community School District and School Pedagogical Professionals

1.	Robert Boyce	Creative Writing Teacher
2.	Joan Byrd	Executive Assistant to the Superintendent
3.	Ronnie Cammeyer	Teacher
4.	Miriam Corn	Principal
6.	Shelley Freeman	District Director of Communication Arts
7.	Myrna Friedlander	District Test Liaison/Promotional Policy Facilitator
8.	Rose Genkin	Reading Teacher/Teacher Trainer
9.	Lillian German	Teacher
10.	Jeffrey Glicker	Assistant Principal
11.	Stuart Goldberg	Assistant Principal (Interim Acting)
12.	Thelma Harper	Principal
13.	David Henry	Assistant Principal
14.	Karen Homler	Teacher
15.	Ronnie Korenge	Teacher
16.	Barbara Levine	Teacher
17.	Winnifred Mayers	Teacher
18.	Rita Menkes	Teacher
19.	Naomi Miller	Teacher
20.	Vernita Patterson	Teacher
21.	Rhonda Plawner	Teacher
22.	Carol Rosen	Teacher
23.	Gilda Tesser	District Supervisor of Teacher Trainers
24.	Leon Weisman	District Director of Language Arts
25.	Bernice Wiley	Principal
26.	Michael Yagoda	Teacher

Educational Testing Service Staff

1.	Garlie Forehand	Director of Research, Educational Testing Service
2.	Myrtle Rice	Research Associate

Board of Education Headquarters Staff

1.	Richard Guttenberg	Director, Office of Educational Assessment
2.	Gary Kippel	Project Director, Office of Educational Assessment
3.	Charlotte Brown	Executive Assistant
4.	Ranjit Shivakumar	Computer Systems Consultant
5.	Lisa Solomon	Educational Consultant
6.	Ilene Wilets	Psychometric Consultant

Ex Officio Professional Panel Members

1.	Edna Cohen	Principal
2.	Roberta Cohen	Principal
3.	Heywood Feierstein	Principal
4.	Irwin Grossbard	Principal
5.	Maria S. Guasp	Principal
6.	Stanley Lavnick	Principal

0293U

X.A: PROFESSIONAL PANEL OPINIONS ON THE USEFULNESS OF STANDARDIZED TESTS

At the May 8th meeting, the Professional Panel was asked to complete the Participant Information Form (see Attachment #11). This form consisted of four questions in an attempt to gain an overall evaluation of standardized tests to use as a guide in the development and implementation of the SMRT-STEPS test.

Their responses were organized into the two categories of supervisors and teachers, in order to: 1) reflect areas of common agreement and concern between the two groups, and 2) identify the unique concerns of both supervisors and teachers.

The following is question #1 from the Participant Information Form: "In general, are standardized reading tests useful to you? If so, in what manner?"

Common Concerns of Both Supervisors and Teachers:

Both supervisors and teachers felt that standardized reading tests are useful for funded program evaluation and for identifying the need for professional and support staff including consulting services and supplies. They also felt that standardized reading tests are useful in helping to identify students in need of small group instruction; in identifying the level of the child in relation to grade standing; in identifying promotional gates and Limited English Proficiency (LEP) entitled population; and for general skills assessment.

Additional areas of common concern include the opinions that standardized reading tests do not reflect overall student performance and that they do not always reflect curriculum. In addition, both supervisors and teachers want very specific identification of deficit areas so that lesson plans can zero in on these areas.

Unique Concerns of Supervisors:

Supervisors stressed that standardized reading tests are useful for evaluating funded programs; for the comparative monitoring of schools; and for assessing accountability of principals. They also felt that these tests were valid in determining promotion. The opposite view was expressed in the teacher category. Specifically, teachers felt that tests represented too much pressure on both teacher and child when used for promotional purposes.

Unique Concerns of Teachers:

Teachers placed emphasis on skills assessment for individual instruction and stressed that standardized tests provide a fairly realistic assessment of the brighter child. For several teachers there was a great deal of concern that test results come too late in the year to be of much use as a pedagogical guide.

Question number two was: "In your professional opinion, how can standardized reading test results be made more useful?"

Common Concerns of Both Supervisors and Teachers:

Supervisors and teachers felt that test results should be reported earlier in the year providing accurate diagnosis of pupil needs toward the beginning of the term instead of at the end. It was felt that this specific feedback would serve as a pedagogical guide to individualize instruction. Both groups also felt that a pre/post format would be desirable; that is to say that the same test (or a parallel form) given at the beginning of the year is also given at the end of the year. This way, focus would be on areas indicated as weak at the beginning of the year and improvement could be ascertained easily with the pedagogical approach evaluated at the end of the term.

Unique Concerns of Supervisors:

Supervisors felt that standardized reading test results could be more useful if: the tests accurately reflected the curriculum; teachers were involved in test preparation; and that tests were field-tested for greater reliability and validity. Supervisors also stressed that test results should be used as one criteria of many in a child's evaluation and should not be used to compare schools and districts serving different populations.

Unique Concerns of Teachers:

Teachers felt that standardized reading test results could be more useful if the print-out data were easier to read and analyze and if the results were more precise in pinpointing areas of difficulty.

X.B: PROFESSIONAL PANEL RATINGS OF THE USEFULNESS OF THE SCHOOL MASTERY OF READING TEST

At the June 19 meeting, all Professional Panel members were asked to rate their opinions of the potential usefulness of SMRT for several specific purposes (see Attachment # 12). As depicted in Table 9, most ratings reflected the opinion that SMRT would be of "above average" usefulness. In this rating scale, "average" is defined as the middle rating category or value "4." It is noted, also, that SMRT was judged as being most useful for identifying remedial cases and for planning instruction for both individuals and groups.

Table 9

PROFESSIONAL PANEL MEMBER OPINIONS - THE USEFULNESS OF THE SCHOOL MASTERY OF READING TEST

	MEAN	S.D	PERCENTAGES OF RATINGS low 1	2	3	4	5	6	high 7
Identify remedial cases	5.61	1.01	0.00	0.00	0.00	16.70	27.80	33.30	22.20
Help plan instruction for individuals	5.39	1.30	0.00	5.60	5.60	11.10	11.10	55.60	11.10
Help plan instruction for class groups	5.22	1.23	0.00	5.60	5.60	11.10	22.20	50.00	5.60
Provide information for state funding	5.11	1.29	0.00	5.60	5.60	16.70	27.80	33.30	11.10
Report progress to students	4.78	1.55	5.60	11.10	0.00	5.60	44.40	27.80	5.60
Detect system-wide general strengths/weaknesses	4.75	1.25	5.90	0.00	0.00	29.40	29.40	35.30	0.00
Determine minimum competency standards	4.69	1.26	0.00	11.80	5.90	5.90	47.00	29.40	0.00
Measure educational status of individuals	4.67	1.05	0.00	0.00	22.20	11.10	44.40	22.20	0.00
Measure educational "growth" of individuals	4.56	1.30	0.00	5.60	11.10	44.40	5.60	27.80	5.60
Compare classes within a school	4.44	1.42	0.00	16.70	5.60	22.20	33.30	16.70	5.60
Compare schools within a system	4.28	1.59	0.00	22.20	11.10	16.70	22.20	22.20	5.60
Report to parents	4.28	1.33	5.60	5.60	11.10	27.80	33.30	16.70	0.00
Modify curriculum	4.22	1.58	11.10	5.60	11.10	16.70	33.30	22.20	0.00
Compare students with national peer groups	4.19	1.47	0.00	17.70	17.70	11.80	23.50	29.40	0.00
Assess school performance	4.11	1.49	5.60	16.70	11.10	5.60	50.00	11.10	0.00
Screen special educational students	3.78	1.58	5.60	16.70	27.80	16.70	16.70	11.10	5.60
Help evaluate teaching procedures or methods	3.72	1.69	11.10	22.20	5.60	22.20	27.80	5.60	[illegible]
Report to boards of education	3.67	1.63	16.70	11.10	11.10	22.20	27.80	11.10	[illegible]
Compare a system with systems across the country	3.38	1.32	5.90	23.50	17.70	29.40	11.80	5.90	5.90
Help evaluate teacher performance	3.28	1.41	11.10	22.20	22.20	22.20	16.70	5.60	0.00
Help evaluate instructional materials	3.22	1.51	11.10	27.80	22.20	16.70	11.10	11.10	0.00
Select students for bilingual training	2.75	1.20	11.80	35.30	23.50	11.80	17.70	0.00	0.00
Report to newspapers	2.38	1.05	23.50	29.40	23.50	17.70	5.90	0.00	0.00

X.C: PROFESSIONAL PANEL SUGGESTIONS FOR SCHOOL IMPROVEMENT PLANS

At the May 8th meeting, the Professional Panel was asked to complete the Participant Information Form (see Attachment #11). Items three and four of the Participant Information Form asked supervisors and teachers for their suggestions for various instructional and organizational strategies, materials, plans and programs likely to lead to improved school performance in reading.

Suggestions Assuming No Additional Resources

Item number three requested suggestions assuming that: "...current level of resources applies...that is, no additional resources are available." Supervisor and teacher Professional Panel member responses are elaborated below.

Common Concerns of Both Supervisors and Teachers:

Both teachers and supervisors felt that the following strategies would lead to improved school performance: smaller registers and small group instruction; better trained teachers and paraprofessionals; lessons and strategies based on specific needs determined by test results.

Unique Concerns of Supervisors:

Supervisors emphasized more effective use of faculty and grade conferences and more teacher input in ordering supplies.

Unique Concerns of Teachers:

Teachers focused on high interest reading materials and basal readers as well as more parent involvement; encouraging parents to foster a positive attitude in their children towards books and libraries, their schools and classroom.

Suggestions Assuming Unlimited Resources

Item number four asked: "If unlimited resources were available, what strategies, material, plans and/or programs would you recommend? Supervisor and teacher Professional Panel member responses are elaborated below.

Common Concerns of Both Supervisors and Teachers:

Both supervisors and teachers strongly indicated smaller class size as an extremely desirable goal. If unlimited funds were available, both groups would have more teacher trainers, more resource room teachers and resource people to give demonstration lessons in remedial techniques. They would have a paraprofessional in every classroom; high interest textbooks; multi-sensory materials to help with instruction; and high quality library collections.

Common Concerns of Supervisors:

Supervisors stressed classroom library collections; individualized reading kits; and a resource room for teachers. If unlimited funds were available, supervisors also would coordinate textbook subject matter to standardized tests. In addition, reading management systems correlated to Basal Reading Programs currently in use were suggested.

Common Concerns of Teachers:

Teachers suggested the use of computers; first, as part of a reading program using word game exercises and exercises in syntax; second, to store up-to-date and ongoing records for each student; and third, to store test scores and specific test results in a diagnostic and prescriptive format. Teachers expressed concern about greater input for staff into textbook selection, programming and content of teacher training sessions.

XI. VALIDITY OF THE SCHOOL MASTERY OF READING TEST

Validity of the School Mastery of Reading Test (SMRT) has been established based upon content and concurrent validation procedures. Specifically, the relationship between SMRT and New York City reading and language arts curriculum was assessed both by project staff and by curriculum, language arts and reading specialists of the Division of Curriculum and Instruction of the New York City Board of Education. Furthermore, teacher opinions of SMRT were elicited after they administered this test to their students. As discussed elsewhere in this report, also, the Professional Panel of New York City teachers and supervisors provided ratings reflecting their opinions of the usefulness of SMRT. In addition, the relationship between SMRT and the following other measures of reading is being assessed: the Degrees of Reading Power (DRP) test, the Interactive Reading Assessment System (IRAS), the Metropolitan Achievement Test, and the National Assessment of Educational Progress (NAEP). More specific details are provided on the following pages.

XI.A: NEW YORK CITY CURRICULUM AND THE SCHOOL MASTERY of READING TEST

The School Mastery of Reading Test is related specifically to New York City public school curriculum. In order to establish the congruence between SMRT and New York City public school fourth grade reading curriculum, the following three Board of Education of New York City (1968, 1969, 1980) publications were used to define curriculum: Minimum Teaching Essentials - Grades 3-5, (MTE), Sequential Levels of Reading Skills (SLRS), and the Handbook for Language Arts - Grades 3 and 4 (HLA). In addition, guidance and assistance were provided by citywide curriculum specialists of the Division of Curriculum and Instruction.

First, each SMRT item was reviewed and related to MTE. If it was not possible to relate the item to MTE, then SLRS was used. As a third resort, HLA was reviewed. When the item was located, record was made of both the descriptive category and the grade level to which that item was related.

Specific examples will illustrate the manner in which the reading category of each test item was determined. Question thirty on page seven of the test booklet reads as follows:

> Astronauts visited the moon. They brought back rocks. These rocks are being studied.
>
> 30. Who visited the moon?
> (A) rocks
> (B) astronauts
> (C) nobody

Initially, the descriptive categories described in Attachment 1 were used to classify test items. For example, the definition under "Literal Comprehension" reads:

> "The student reads a sentence, several sentences or a short story and (1) chooses a sentence that has the same meaning. (2) chooses a picture that best represents the meaning of what was read, and (3) answers factual questions about what was read by choosing from a list of possible answers. The reading material includes: simple sentences, compound subjects and objects, compound and complex sentences."

Question thirty on page seven, concerning astronauts visiting the moon, would fall under category (3) above: "answers factual questions about what was read by choosing from a list of possible answer".
The reading material would be "simple sentences."

Next, the Communication Arts-Grade Four section of MTE was examined. In the judgment of the professional educators involved, this test item was related to "Literal Comprehension - details may be used to support a point" (see Attachment # 13).

In fact, MTE further breaks this down by referring to page 26 in the SLRS book (see page 26 after: "details may be used to support a point", on Attachment # 13). Subsequently, page 26 in SLRS (see Attachment # 14) "Level C" under "Finding and relating details" says: "Develops the concept that details fill out, extend, make specific, exemplify, clarify and document the main idea." Also, the description under "Language Arts" right below it further clarifies this by stating: "Reads vivid passages, recalls details, and describes or draws simple pictures including details."

Another example is question eighty-eight on page twenty-two of the test booklet. The question reads as follows:

88. Janet put on her bathing suit. She walked to the pool. What happened next?

(A) Janet went swimming
(B) Janet went fishing
(C) Janet went to school

As indicated earlier, Attachment # 1 was used to describe test items. The descriptive categories under "Reasoning Comprehension" reads as follows: "The student reads a sentence, several sentences or a short story and answers inferential questions by choosing from a list of pictures or written answers." The reading materials include: "simple paragraphs, a short story, causal and all or some relationships, predicted outcomes, comparisions and sequencing."

Question eighty-eight, concerning what "Janet did after putting on her bathing suit and walking to the pool," would fall under "answering inferential questions by choosing from a list of written answers." The reading material would be "sequencing."

Next, the "Fourth Grade Communication Arts" section of the MTE's was reviewed, but this type of sequencing question was not listed there. Knowing that fourth grade curriculum naturally comprises some of the third grade material, the "Third Grade-Communication Arts" section was examined. It was listed there under "Inferential Comprehension" questions. Attachment # 15 indicates: "Sequential order is a factor in predicting outcomes and drawing conclusions."

The "page 28" at the end of this sentence refers to the SLRS section on "Determining Sequence." Attachment # 16 specifies: "Level D: Uses sequential order to make interpretations and predictions". Also, right below under "Language Arts", it states: "Predicts outcomes of story stemming from logical sequence of events."

The purpose in matching the SMRT items with the New York City Board of Education fourth grade curriculum is to demonstrate the extent to which SMRT is congruent with New York City curriculum. All of the SMRT test items have been analyzed and matched reading curriculum in the three curriculum publications mentioned earlier. In other words, SMRT has been constructed to reflect curriculum taught in New York City schools.

0294U/14 & 15

XI.B: TEACHER OPINIONS OF THE SCHOOL MASTERY OF READING TEST

After administering SMRT, each teacher was asked to complete a one-page survey (see Attachment # 5) designed to assess their opinions of the test and testing procedures.

In most instances, as indicated in Table 10, teacher opinions were favorable. For example, the majority of teachers agreed that:

The test items are easy to understand for the students.

The test questions seem free of technical flaws.

The test items are free of bias (for example, sex or ethnic bias).

The test directions are clear to the test-taker.

The test administration manual is easy to use.

The layout of the test booklet is convenient for the examinees.

The reading difficulty levels of test questions are appropriate to fourth grade students.

The majority of the test questions correspond well to fourth grade curriculum (for example, Minimum Teaching Essentials).

The majority of teachers indicated that there is no need for additional practice questions for students.

At the same time, the majority of teachers indicated that:

They somewhat disagree that time limits are adequate, and

They disagree that the layout of the test booklet is attractive.

In other words, there is some evidence that time guidelines or limits should be improved. In addition, the layout of the test booklet should be made more attractive.

Table 10

TEACHER OPINIONS OF THE SCHOOL MASTERY OF READING TEST

	Strongly Agree	Agree	Somewhat Agree	Somewhat Disagree	Disagree	Strongly Disagree
The test items are easy to understand for the students.	1 (18.9%)	2 (45.9)	3 (18.9)	4 (8.1)	5 (5.4)	6 (2.7)
The test questions seem free of technical flaws.	1 (13.9%)	2 (38.9)	3 (27.8)	4 (13.9)	5 (5.6)	6 ---
The test items are free of bias (for example, sex or ethnic bias).	1 (36.1%)	2 (50.0)	3 (13.9)	4 ---	5 ---	6 ---
The testing time limits are adequate.	1 (16.2%)	2 (24.3)	3 (18.9)	4 (27.0)	5 (8.1)	6 (5.4)
The test directions are clear to the test-taker.	1 (24.3%)	2 (32.4)	3 (16.2)	4 (21.6)	5 (2.7)	6 (2.7)
The test administration manual is easy to use.	1 (32.4%)	2 (48.6)	3 (16.2)	4 (2.7)	5 ---	6 ---
There is a need for additional practice questions for the student.	1 (19.4%)	2 (11.1)	3 (13.9)	4 (11.1)	5 (38.9)	6 (5.6)
The layout of the test booklet is attractive.	1 (5.4%)	2 (37.8)	3 (32.4)	4 (8.1)	5 (16.2)	6 ---
The layout of the test booklet is convenient for the examinees.	1 (16.2%)	2 (40.5)	3 (18.9)	4 (13.5)	5 (8.1)	6 (2.7)
The reading difficulty levels of test questions are appropriate to fourth grade students.	1 (8.6%)	2 (51.4)	3 (25.7)	4 (11.4)	5 (2.9)	6 ---
The majority of the test questions correspond well to fourth grade curriculum(for example, Minimum Teaching Essentials).	1 (8.3%)	2 (61.1)	3 (25.0)	4 (5.6)	5 ---	6 ---

XI.C: OTHER READING TESTS AND THE SCHOOL MASTERY OF READING TEST

The initial version of the group administered SMRT was validated using the individually administered Interactive Reading Assessment System (IRAS) developed by Robert and Kathryn Calfee. The IRAS instrument was based on a review of research and practice in the teaching of reading. It assesses a set of component skills which have generally been accepted as necessary for proficient reading. A very strong relationship was found between the initial version of SMRT and the IRAS.

The relationship between the School Mastery of Reading Test (SMRT) and both the Degrees of Reading Power (DRP) test and the Metropolitan Achievement Test (MAT) was examined. Total test scores were available on all three tests for 744 students. Correlation coefficients were calculated. The correlation between SMRT and DRP was .791 and the correlation between SMRT and the MAT was .819.

The relationship between SMRT and the National Assessment of Educational Progress (NAEP) is discussed in detail in another section of this report.

XII. RELIABILITY OF THE SCHOOL MASTERY OF READING TEST

Indices of reliability provide an indication of the extent to which a particular measurement is consistent and reproducible (Thorndike & Hagen, 1977, p.56). In other words, reliability refers to the necessity for dependability in measurement (Kerlinger, 1973, pp.442-443). Reliability implies stability, consistency, predictability and accuracy. In more technical terms, reliability is the proportion of true variance in obtained test scores (see, for explanation, Guilford, 1954, pp. 349-354).

Coefficient alpha is the basic formula for determining the reliability based on obtained consistency (Nunnally, 1978, pp. 229-230). Coefficient alpha is the expected correlation of one test with an alternative form of the test, of the same length, when the two tests purport to measure the same thing (Nunnally, 1967, pp.196-197).

In order to obtain reliability estimates for SMRT, coefficient alpha was calculated. Coefficient alpha reliability coefficients are presented in Table 11 . These data provide support for the contention that SMRT can be used reliably.

0293U/2

Table 11

Reliability of the School Mastery of Reading Test

(N=752)*1

Aggregate of Test Items	Number of Items	Chronbach's Alpha
Total Test*2	97	.9281
Part One	47	.8611
Part Two	50	.8809
Word Attack Subtest	18	.7610
Word Meaning Subtest	21	.7880
Literal Comprehension Subtest	31	.8202
Reasoning Comprehension Subtest	27	.7775

*1 The number of students providing data for these reliability analyses was 752 as determined by the "listwise deletion" requirements of the SPSS-X (1986, p.872) computer program. Specifically, these analyses were comprised of item responses only from test papers where students answered all of the 100 questions.

*2 The three word recognition items at the very beginning of the School Mastery of Reading Test were eliminated from these analyses. The Word Recognition items were used as "sample" items to orient students. Total test reliability, therefore, was based upon 97 rather than 100 items.

0293U/4

XIII. PREDICTING DEGREES OF READING POWER TEST SCORES FROM SCHOOL MASTERY OF READING TEST SCORES

It is likely that the School Mastery of Reading Test (SMRT) will be of most usefulness if it is administered at the beginning of the school year. When administered at that time, it can provide diagnostic information to assess needs and to help guide planning and instruction. To some extent, in addition, SMRT administered at the beginning of the school year may be predictive of Degrees of Reading Power (DRP) test scores from the annual citywide reading testing program conducted near the end of the school year. This can be accomplished by generating appropriate regression equations.

For example, a linear regression equation was calculated based upon May 1986 test results. The illustrative equation expresses the relationship between SMRT and DRP as follows:

$$y = .690x - 16.291,$$

where "x" represents SMRT scores achieved by students and "y" represents predicted DRP scores. This equation could be used to predict or estimate fourth grade DRP score from fourth grade SMRT scores available before the DRP has been administered. It is noted that more sophisticated and more accurate multivariate models using SMRT could be developed for predictive purposes.

XIV. NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS AND THE SCHOOL MASTERY OF READING TEST

National Assessment of Educational Progress (NAEP) has been developed to measure how effectively 9-, 13- and in-school 17-year-old American students can read (Messick, Beaton, & Lord, 1983). For this purpose, nationally representative samples of students within various demographic subgroups are tested (National Assessment of Education Progress, 1985).

NAEP bases each assessment on a wide range of materials and asks questions requiring use of a variety of reading skills and strategies. Reading selections range from simple sentences expressing a single concept to complex articles about specialized topics in science or social studies.

A scale ranging from 0 through 500 has been developed. Both items and tests span a wide range of difficulty and are presented in a variety of formats. Various points on the scale have been provided criterion-referenced interpretations. Items have been calibrated using the three-parameter latent trait model (See, for discussion; Messick, Beaton, & Lord, 1983, pp. 43-55).

Items are reviewed for potential bias before being accepted by NAEP for administration. Specifically, NAEP items are reviewed by educators on the basis of their academic appropriateness, effectiveness, freedom from bias or stereotyping, and sensitivity to racial, ethnic, religious and political groups. After test administration, item response curves are analyzed for potential bias.

The relationship between SMRT and NAEP is being determined. In effect, the current study is designed to improve local school level diagnosis and prescriptions for progress by using NAEP items and norms (See footnote #1). The primary intent is to determine the feasability of:

1) obtaining norm-referenced interpretations of SMRT results with respect to NAEP national norms,

2) demonstrating the extent to which SMRT results relate to NAEP absolute performance,

3) establishing a cost-effective source of items by incorporating NAEP items within SMRT, and

4) estimating costs and resources required.

---

[1] In addition, it is noted that a somewhat different potential role for NAEP in assisting the development and implementation of local educational standards has been defined by Messick (1985).

In order to achieve these objectives, NAEP items were evaluated with regard to item content, format and general appropriateness for fourth grade students. It was determined, consequently, that some NAEP items could be incorporated within SMRT. This decision was based upon the fact that current elementary school level NAEP items were designed for grade three students and have sufficient range for grade four students. In the recent past, elementary school level NAEP items were designed for grade four students. It is noted that SMRT is designed for relatively low achieving fourth grade students. Furthermore, SMRT is most likely to be administered early in the school year for maximum diagnostic usefulness.

Some NAEP items are so similar in format and content to some SMRT items that, if mixed together, it would be difficult to determine the source of each. At the same time, some SMRT item types are not matched by NAEP items.

A more specific assessment of the congruence between NAEP and SMRT requires detailed review of item content, format and psychometric characteristics of items from both sources. In addition, analyses of scales and assessment of the dimensionality of both NAEP and SMRT are required. Such reviews are being conducted.

At our request, the Educational Testing Service (ETS) obtained permission for use of NAEP items within SMRT. Permission was granted to use NAEP items under "reasonable constraints". Specifically, it is understood that: 1) NAEP items will not be published or inappropriately disseminated, 2) NAEP items will not be used for pre-test practice or instruction, and 3) appropriate steps will be taken to insure adequate security of NAEP items.

Selection of particular items were based upon item scale value, content and format. A total of 16 NAEP items were selected for testing. These NAEP items were embedded within both Parts I and II of the two-part SMRT test.

In addition to obtaining traditional item statistics, latent trait analyses are being conducted. The intent is to use the NAEP items to determine the SMRT item parameters. First, it must be demonstrated that the test data fits the Latent Trait Model (See, for discussion; Hambleton & Cook, 1977; Wright & Stone, 1979). Subsequently, SMRT item calibrations will be calculated. In effect, the overall goal is to establish a common SMRT-NAEP scale with a Latent Trait calibrated item pool. Consequently, new forms and levels of SMRT can be designed which will be based upon New York City curriculum and will yield NAEP norm-referenced information. It is noted that a similar strategy (See, for discussion; Lenke, Oswald, & Kippel, 1982) was followed in the development and administration of the customized New York City mathematics test for five years from Spring 1981 through Spring 1985.

XV. DEVELOPMENT OF SUBTESTS FOR THE SCHOOL MASTERY OF READING TEST

Test items were categorized by subtest based upon the professional opinions of several curriculum, reading, research and teaching specialists. Subtests were developed using the definitions provided in Attachment # 1.

The School Mastery of Reading Test (SMRT) is comprised of two parallel sections (i.e., Parts I and II), each containing 50 items. The items within each part serve to measure a particular facet of reading ability. The five following types of items were included in the May, 1986 test administration: 1) word recognition, 2) word attack, 3) word meaning, 4) literal comprehension, and 5) reasoning comprehension.

As indicated earlier, the numbers of items within the five categories were: word recognition--3 items, word attack--18 items, word meaning--21 items, literal comprehension--31 items, and reasoning comprehension--27 items.

In order to validate and confirm the placement of items within the particular subtests, the use of factor analytic statistical techniques was considered. Consequently, we conferred with an expert in such techniques, Dr. Phil Merrifield, who is a professor and measurement specialist at New York University. Dr. Merrifield recommended an alternate statistical approach.

Factor analytic procedures do not appear to be appropriate for the development and confirmation of subtests on mastery tests such as SMRT. The factor analytic technique relies on the assumption that test scores are normally distributed, i.e. some students did very well, some did poorly, and the majority fell somewhere in-between the two extremes. The SMRT, however, is a test measuring fourth grade reading mastery, which was administered to fourth graders at the end of their academic year. As a result, most students obtained relatively high scores because they had mastered fourth grade reading skills. As expected, consequently, the distribution of test scores was "negatively skewed" rather than normally distributed. This departure from the bivariate normal distribution might confound any results obtained through factor analytic methods

Dr. Merrifield recommended that the subtests be validated by rank-ordering the items in each of Parts I and II within subtests based upon item difficulty. For example, the Literal Comprehension subtest items in Part I would be rank-ordered based upon item difficulty. Secondly, the additional Literal Comprehension subtest items in Part II would be rank-ordered separately from Part I Literal Comprehension items. Subsequently, the same procedure would be followed with items in Parts I and II for each of the other subtests.

After items are matched on item difficulty, correlations between Parts I and II for each subtest must be calculated. For example, to what extent the Literal Comprehension items in Part I of the SMRT correlate with those in Part II? A high correlation suggests the items are all measuring a common reading skill. The group of items may then be considered a cluster or factor, representative of one of the various dimensions comprising reading ability. This is similar, in effect, to establishing two alternate or parallel forms of the Literal Comprehension subtest.

In addition, correlations between different subtests for each of Parts I and II of SMRT should be obtained. For example, what is the correlation or relationship betwen Literal Comprehension and Reasoning Comprehension within Part I? In this case, low correlations would suggest distinct subtests. In contrast, high correlations might indicate that certain subtests might be combined. This type of information would be extremely useful in the test validation process.

These particular statistical techniques may be thought of in terms of split-half reliability methods. In effect, the underlying structure of the test is being examined by determining the extent to which the items relate to each other. In other words, are the items measuring the same or different facets of reading ability? How strong is this relationship? How weak? All these questions are important in terms of the validation of the SMRT subtests.

The relationship between the four primary subtests is presented in Table 12 . Review of the correlations reveals that the highest correlation (i.e., .764) is between the Literal Comprehension and Reasoning Comprehension subtests. The lowest correlation is between Word Attack and Reasoning Comprehension.

Table 12

Intercorrelation Matrix of School Mastery of Reading Test Subtests

(N = 889)

SUBTEST	WORD ATTACK	WORD MEANING	LITERAL COMPREHENSION	REASONING COMPREHENSION
WORD ATTACK	1.000	.670	.644	.628
WORD MEANING		1.000	.734	.698
LITERAL COMPREHENSION			1.000	.764
REASONING COMPREHENSION				1.000

0293U/3

XVI. Potential Short- and Long-Term Objectives

Short-Term Objectives

Office space must be acquired and computer terminal hardware and software must be set-up and available for use.

A budget must be established for the 1986-1987 school year.

Consultants must be acquired for the 1986-1987 school year.

In order to obtain necessary funds for all aspects of SMRT, a technically sophisticated research proposal must be developed for submission to the Office of Educational Research of the Department of Education, and possibly, to various foundations.

Latent trait analyses to determine the feasibility of determining common scaling between SMRT and National Assessment of Educational Progress must be conducted. A common scale between SMRT and NAEP will enable: cost-effective source of items, comparisons with National NAEP results, and reference to NAEP absolute performance.

School improvement plans of participating school must be reviewed.

SMRT and the test administration manual must be reviewed and revised based upon the May 1986 test administration.

Time guidelines for administration SMRT will be reviewed and the possibility of administering SMRT as a "power" rather than as a "speed" test will be determined.

The graphics, layout, format and the quality of both the SMRT booklet and the test administration manual will be improved.

Implications of SMRT for the Chancellor's Minimum Standards Program must be determined.

SMRT may be administered to more schools in both October and May to develop local norms and/or expectations of performance and to further elaborate standardization and technical test characteristics.

Item analyses will be performed and psychometric characteristics of SMRT will be established. For example, the following may be included for items, subtests and total test, as appropriate: difficulty levels, distractor analysis, central tendency, variability, reliability, item intercorrelation, item-subtest-total test correlation, and proportion of responses achieving mastery or manifesting non-mastery.

A computerized SMRT item bank will be established. Procedures for writing or acquiring new test items required for different forms and levels of SMRT must be established.

Descriptive characteristics of participating students, staff and schools will be obtained and summarized from New York City Board of Education School Profiles reports.

0296U/13

Long-Term Objectives

Mastery criteria must be developed for SMRT items, subtests and/or total test score.

A SMRT student answer sheet must be developed and appropriate scoring software must be developed.

Prototypes of meaningful reports for school administrators and teachers must be developed.

Relevance of SMRT for special education and limited English-proficient students must be determined.

Demonstration of the relationships between SMRT, the Degrees of Reading Power (DRP) test and the Metropolitan Achievement Test (MAT) must be conducted.

Innovative uses of computer technology for test delivery, scoring and interpretation will be explored.

A new SMRT answer key must be developed to establish standardized response categories.

SMRT score profiles must be related to prescriptive choices of instructional programs, strategies and materials for local school improvement.

SMRT should be extended to either grade three or grade seven students.

Both Fall and Spring Citywide norms should be developed.

The relevance of SMRT to the Promotional Gates Program must be demonstrated. SMRT may assist in individual student diagnosis, prescription and Promotional Gates "appeals" decision-making.

The relationship between SMRT and Comprehensive Assessment Report results and conclusions must be determined. In particular, the relationship with Degrees of Reading Power test results must be determined.

0296U/14

BIBLIOGRAPHY

Board of Education of New York City. (1968). Sequential levels of reading skills; Prekindergarten-grade 12. Brooklyn, New York: Division of Curriculum and Instruction.

Board of Education of New York City. (1969). Handbook for language arts; Grades three and four. Brooklyn, New York: Division of Curriculum and Instruction.

Board of Education of New York City (1980). Minimum teaching essentials; Grades 3-5. Brooklyn, New York: Division of Curriculum and Instruction.

Ebel, R.L. (1965). Measuring educational achievement. Englewood Cliffs, New Jersey: Prentice-Hall.

Guilford, J.P. (1954). Psychometric Methods (2nd ed.). New York: McGraw-Hill.

Hambleton, R.K., & Cook, L.L. (1977). Latent trait models and their use in the analysis of educational test data. Journal of Educational Measurement, 14, 76-96.

Kerlinger, F.N. (1973). Foundations of behavioral research (Second Edition). New York: Holt, Rinehart & Winston.

Lenke, J.M., Oswald, J., & Kippel, G. (1982). An application of Rasch model technology to a large-city testing program. Florida Journal of Educational Research, 24, 23-28.

Livingston, S.A. & Zieky, M.J. (1982). Passing scores: A manual for setting standards of performance on educational and occupational tests. Princeton, New Jersey: Educational Testing Service.

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

Messick, S., Beaton, A., & Lord, F. (1983). National Assessment of Educational Progress Reconsidered: A new design for a new era. (NAEP Report 83-1). Princeton, New Jersey: National Assessment of Educational Progress at Educational Testing Service.

National Assessment of Educational Progrss (1985). The reading report card: Progress toward excellence in our schools; Trends in reading over four National Assessments, 1971-1984 (Report No. 15-R-01). Princeton, New Jersey: National Assessment of Educational Progress at Educational Testing Service.

New York City Board Of Education - School Profiles, 1982-1983 (1984). Brooklyn, New York: New York City Board of Education.

New York State Education Department (1984). Proposed new part of 100 of the Commissioner's Regulations. Albany, New York: Deputy Commissioner for Elementary, Secondary and Continuing Education.

New York State Education Department (1985). Review of the comprehensive assessment report. Albany, New York: Deputy Commissioner for Elementary, Secondary and Continuing Education.

Nunnally, J.C. (1967). Psychometric theory. New York: McGraw-Hill.

Nunnally, J.C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.

The Psychological Corporation (Undated). A glossary of measurement terms (Test service notebook 13). San Antonio, Texas.

Seashore, H.G. (1980). Methods of expressing test scores (Test service notebook 148). San Antonio, Texas: The Psychological Corporation.

SPSS-X: User's Guide. (2nd ed.) (1986). Chicago, Illinois: SPSS, Inc.

Thorndike, R.L., & Hagen, E.P. (1977). Measurement and evaluation in psychology and education (Fourth Edition). New York: John Wiley & Sons.

Wright, B.D., & Stone, M.H. (1979). Best test design: Rasch measurement. Chicago: Mesa Press.

APPENDIX

Description of School Mastery of Reading Test Category/Subtest

Category/Subtest	Description
1. Word recognition	The student (1) hears a word and chooses that word from a list of words, (2) reads a word and chooses a matching picture, and (3) looks at a picture and chooses a matching word. The following words are included: of, was, cat, dog, four, from, one, what, some, know, might, flower, night, automobile, piano, birdcage, castle, swords.
2. Word attack	The student (1) hears a word and chooses a word with the same sound from a list of words (ŭ, ā, ō, oi, ow, f, ch, t, gh), and (2) reads a word with a portion underlined and chooses from a list a word with the same sound as the underlined portion (hard c (k), gh, ch, sh, ow, oi (oy), silent b, wr, silent e, soft g).
3. Word meaning	The student (1) matches words to definitions, (2) chooses synonyms and antonyms for words, and (3) chooses words for blank spaces in sentences. The following words are included: ring, cry, chair, night, above, glad, slow, sick, shut, narrow, big, cent, their, children, men, highest, unlike, retell, lost, hide, enjoyed, seen, worked.
4. Literal comprehension	The student reads a sentence, several sentences, or a short story and (1) chooses a sentence that has the same meaning, (2) chooses a picture that best represents the meaning of what was read, and (3) answers factual questions about what was read by choosing from a list of possible answers. The reading material includes: simple sentences, compound subjects and objects, compound and complex sentences.
5. Reasoning comprehension	The student reads a sentence, several sentences, or a short story and answers inferential questions by choosing from a list of pictures or written answers. The reading materials include single paragraphs, a short story, causal and all/some relationships, predicted outcomes, comparisons and sequencing.
6. Comprehension: cloze	The student reads two long stories (six or seven paragraphs each) with seven words missing in each story. For each missing word, the student chooses from a list of five words the word that best completes the meaning of the story.

ATTACHMENT #1

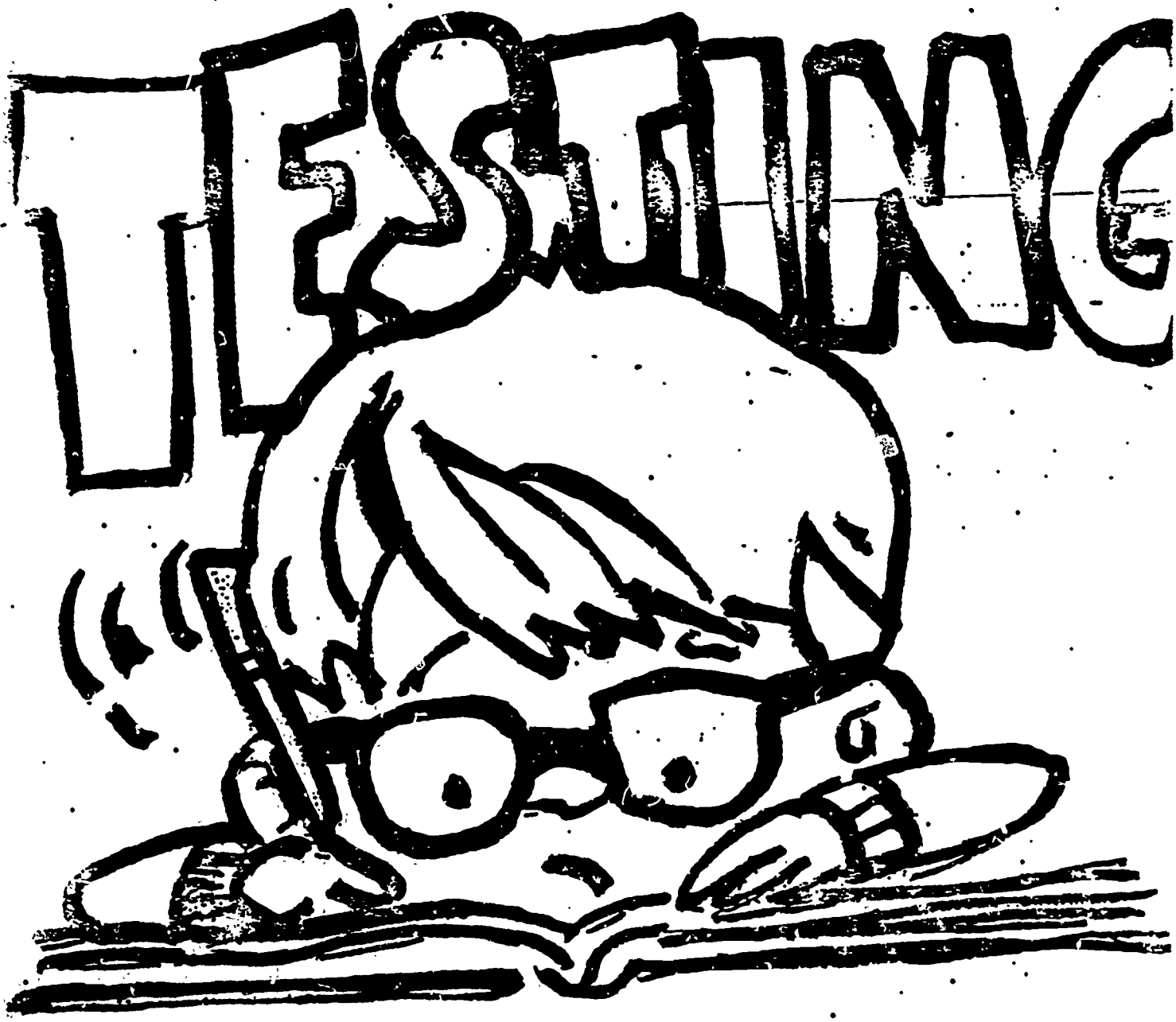
ATTACHMENT #2

QUANTITY AND IDENTIFICATION NUMBERS OF TESTS AND TEST ADMINISTRATION MANUALS

	School Number	Potential Number of Students on Register	25% Overage	Number of Tests Required	Range of of Tests Numbers	Range of Admin. Manuals Numbers
CSD #17	191	134	34	168	1 - 168	1 - 10
	289	157	39	196	169 - 364*	11 - 20
	398	197	49	246	365 - 610**	21 - 30
(Subtotal)				(610)		(30)
CSD #19	213	129	32	161	611 - 771	31 - 40
	290	134	34	168	772 - 939	41 - 50
	328	82	21	103	940 - 1042	51 - 60
(Subtotal)				(432)		(30)
CSD #21	90	76	19	95	1043 - 1137	61 - 70
	212	89	22	111	1138 - 1248	71 - 80
	329	67	17	84	1249 - 1332	81 - 90
(Subtotal)				(290)		(30)
Total				(1332)		(90)

* Box 1 of 2, #'s 169-266; Box 2 of 2, #'s 267-364.
** Box 1 of 2, #'s 365-488; Box 2 of 2, #'s 489-610.

ATTACHMENT #3



DO NOT DISTURB

SIDE 1

GENERAL PURPOSE - NCS - ANSWER SHEET

FOR USE WITH ALL NCS SENTRY OPTICAL MARK READING SYSTEMS
SEE IMPORTANT MARKING INSTRUCTIONS ON SIDE 2

NAME (Last, First, M.I.)

SEX (M) (F)

GRADE OR EDUC (0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10) (11) (12) (13) (14) (15) (16)

BIRTH DATE			IDENTIFICATION NUMBER										SPECIAL CODES					
MO.	DAY	YR.	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Jan																		
Feb																		
Mar	0 0	0 0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Apr	1 1	1 1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
May	2 2	2 2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Jun	3 3	3 3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Jul	4 4	4 4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Aug	5 5	5 5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Sep	6 6	6 6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
Oct	7 7	7 7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
Nov	8 8	8 8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
Dec	9 9	9 9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9

1 A B C D E	11 A B C D E	21 A B C D E	31 A B C D E	41 A B C D E
2 A B C D E	12 A B C D E	22 A B C D E	32 A B C D E	42 A B C D E
3 A B C D E	13 A B C D E	23 A B C D E	33 A B C D E	43 A B C D E
4 A B C D E	14 A B C D E	24 A B C D E	34 A B C D E	44 A B C D E
5 A B C D E	15 A B C D E	25 A B C D E	35 A B C D E	45 A B C D E
6 A B C D E	16 A B C D E	26 A B C D E	36 A B C D E	46 A B C D E
7 A B C D E	17 A B C D E	27 A B C D E	37 A B C D E	47 A B C D E
8 A B C D E	18 A B C D E	28 A B C D E	38 A B C D E	48 A B C D E
9 A B C D E	19 A B C D E	29 A B C D E	39 A B C D E	49 A B C D E
10 A B C D E	20 A B C D E	30 A B C D E	40 A B C D E	50 A B C D E
51 A B C D E	61 A B C D E	71 A B C D E	81 A B C D E	91 A B C D E
52 A B C D E	62 A B C D E	72 A B C D E	82 A B C D E	92 A B C D E
53 A B C D E	63 A B C D E	73 A B C D E	83 A B C D E	93 A B C D E
54 A B C D E	64 A B C D E	74 A B C D E	84 A B C D E	94 A B C D E
55 A B C D E	65 A B C D E	75 A B C D E	85 A B C D E	95 A B C D E
56 A B C D E	66 A B C D E	76 A B C D E	86 A B C D E	96 A B C D E
57 A B C D E	67 A B C D E	77 A B C D E	87 A B C D E	97 A B C D E
58 A B C D E	68 A B C D E	78 A B C D E	88 A B C D E	98 A B C D E
59 A B C D E	69 A B C D E	79 A B C D E	89 A B C D E	99 A B C D E
60 A B C D E	70 A B C D E	80 A B C D E	90 A B C D E	100 A B C D E

Attachment #3

ATTACHMENT #5

We are interested in developing and improving upon the School Mastery of Reading Test (SMRT). Your expertise as a teaching professional would aid us in this endeavor. Below are several statements concerning various aspects of the test. Please indicate the extent to which you either "agree" or "disagree" with EACH statement by circling the number corresponding to your opinion. Please DO NOT SIGN YOUR NAME on the questionnaire as all responses are to remain anonymous. Remember, there are no right or wrong opinions.

	Strongly Agree	Agree	Somewhat Agree	Somewhat Disagree	Disagree	Strongly Disagree
The test items are easy to understand for the students.	1	2	3	4	5	6
The test questions seem free of technical flaws.	1	2	3	4	5	6
The test items are free of bias (for example, sex or ethnic bias).	1	2	3	4	5	6
The testing time limits are adequate.	1	2	3	4	5	6
The test directions are clear to the test-taker.	1	2	3	4	5	6
The test administration manual is easy to use.	1	2	3	4	5	6
There is a need for additional practice questions for the student.	1	2	3	4	5	6
The layout of the test booklet is attractive.	1	2	3	4	5	6
The layout of the test booklet is convenient for the examinees.	1	2	3	4	5	6
The reading difficulty levels of test questions are appropriate to fourth grade students.	1	2	3	4	5	6
The majority of the test questions correspond well to fourth grade curriculum(for example, Minimum [illegible]ing Essentials).	1	2	3	4	5	6

ATTACHMENT # 6

Types of Test Scores

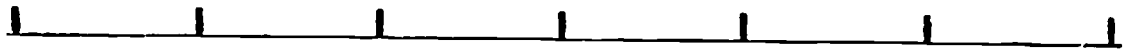
In your opinion, how useful are each of these types of scores?

a. Raw scores

1 2 3 4 5 6 7

Not at all useful — Extremely useful

b. Normal curve equivalents



1 2 3 4 5 6 7

Not at all useful — Extremely useful

c. Percentile rankings

1 2 3 4 5 6 7

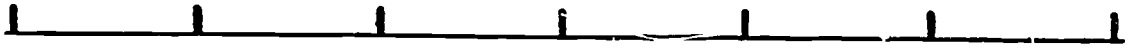
Not at all useful — Extremely useful

d. Stanines

1 2 3 4 5 6 7

Not at all useful — Extremely useful

e. Grade Equivalents



1 2 3 4 5 6 7

Not at all useful — Extremely useful

ATTACHMENT #7

Types of Test Scores

Raw Score.
The first quantitative result obtained in scoring a test. Usually the number of right answers; number right minus some fraction of number wrong; time required for performance; number of errors, or similar direct, unconverted, uninterpreted measure. (see source # 1 below)

Normal Curve Equivalent (NCE).
This score, used primarily for reporting in federally funded programs such as Title I, has a mean of 50 and a standard deviation of 21.06. (see source # 2 below)

Percentile.
A point (score) in a distribution at or below which fall the percent of cases indicated by the percentile. Thus a score coinciding with the 35th percentile (P35) is regarded as equaling or surpassing that of 35 percent of the persons in the group, and such that 65 percent of the performances exceed this score. "Percentile" has nothing to do with the percent of correct answers an examinee makes on a test. (see source # 1 below)

Stanine.
One of the steps in a nine-point scale of standard scores. The stanine (short for standard-nine) scale has values from 1 to 9, with a mean of 5 and a standard deviation of 2. Each stanine (except 1 and 9) is 1/2 standard deviation in width, with the middle (average) stanine of 5 exte3nding from 1/4 S.D. below to 1/4 S.D. above the mean. (see source # 1 below)

Grade Equivalent (GE).

The grade level for which a given score is the real or estimated average. Grade-equivalent interpretation expresses obtained scores in terms of grade and month of grade, assuming a 10-month school year (e.g., 5.7). Since such tests are usually standardized at only one (or two) point(s) within each grade, grade equivalents between points for which there are data-based scores must be "estimated" by interpolation. (see source # 1 below)

Sources: 1. The Psychological Corporation (Undated). A Glossary of Measurement Terms. (Test Service Notebook 13). San Antonio, Texas.

2. Seashore, H.G. (1980) Methods of Expressing Test Scores. (Test Service Notebook 148). San Antonio, Texas: The Psychological Corporation.

NAME: _____ DISTRICT: ____

ITEM BIAS AND COMMENT SHEET

ATTACHMENT # D

The items in this test were developed to be as unbiased as possible with respect to sex, race, ethnicity, and urban/rural origins. In the event that we have inadvertently included items that might discriminate against some group or groups of students, we would like you to indicate by item number any item that you feel is basically unfair. For each item so indicated, please explain the source of the bias that you feel exists. Under "Other Comments", add any further reaction you may have to the item(s).

Page & Item #'s	Source of Bias	Other Comments

(Use other side for additional comments)

ATTACHMENT # 9

SMRT-STEPS

School Mastery of Reading Test
System To Enhance Progress of Schools

PRELIMINARY PROFESSIONAL PANEL MEETING

May 8, 1986
4:00 p.m. to 8:45 p.m.

AGENDA

Registration and Distribution of Folders

Welcome, Introductions and Brief Overview of Current Activities	Gary M. Kippel
Greetings	Richard Guttenberg
Overview of Proposed Long-Term Activities	Garlie A. Forehand Myrtle Rice
Test Administration	Ilene Wilets Charlotte Brown
Self-Scoring of Answer Sheets	Panel Members
Discussion of SMRT Test and Administration Procedures	All Participants
Written Reactions Provided on "Item Bias and Comments" Sheets	Panel Members
Discussion of the Usefulness of Test Information to Improve Schools	All Participants
Written Responses to Questions on Improving Diagnosis and Prescription	Panel Members
Summary, Plans for Next Meeting and Conclusion	Garlie A. Forehand Gary M. Kippel

New York City
Board of Education

110 Livingston Street
Brooklyn, New York 11201

ATTACHMENT # 10

Nathan Quinones
Chancellor

Louise Latty
Chief Executive for Instruction

Office of Educational Assessment
Richard Guttenberg
Director
(718) 596-4045

SMRT-STEPS

School Mastery of Reading Test
System To Enhance Progress of Schools

PRELIMINARY PROFESSIONAL PANEL MEETING

June 19, 1986
4:00 p.m. to 8:45 p.m.

AGENDA

Registration and Distribution of Folders

Usefulness of Standardized Reading Tests

Relationship Between School Mastery of Reading Test and New York City Curriculum

Improving the School Mastery of Reading Test and Test Administration Procedures

Panel Judgements of Mastery Criteria

Preliminary Field Test Results and Future Analyses

Enhancement of School Progress Efforts

Test Scores and Reports

Plans for the Future

ATTACHMENT # 11

PARTICIPANT INFORMATION FORM

SPRING, 1986 PRELIMINARY PROFESSIONAL PANEL

Improving Diagnosis and Prescription

May 8, 1986

ATTACHMENT #11
CONTINUED...

SPRING, 1986 PRELIMINARY PROFESSIONAL PANEL

May 8, 1986

USEFULNESS OF TEST RESULTS FOR INSTRUCTION

In general, are standardized reading test results useful to you? (Check one box)

[_] YES [_] NO

If yes, in what manner? Please specify how these test results are useful to you?

SPRING, 1986 PRELIMINARY PROFESSIONAL PANEL

May 8, 1986

USEFULNESS OF TEST RESULTS FOR INSTRUCTION

In your professional opinion, how can standardized reading test results be made more useful?

ATTACHMENT #11 Continued...

SPRING, 1986 PRELIMINARY PROFESSIONAL PANEL

May 8, 1986

PANEL SUGGESTIONS FOR PRESCRIBED SCHOOL ACTIVITIES

Effective use of diagnostic information to improve school progress requires the implementation of specific activities in response to test scores. For example, when reporting reading test results, it may be desirable to provide school supervisors and teachers with a brief description or computerized "menu" of various instructional and organizational strategies, materials, plans and programs which are likely to lead to improved school performance in reading.

Inasmuch as you are exemplary professional educators, we would appreciate your suggestions of such strategies, materials, plans and/or programs. Please assume that your current level of resources applies...that is, no additional resources are available.

ATTACHMENT #11 continued...

SPRING, 1986 PRELIMINARY PROFESSIONAL PANEL

May 8, 1986

PANEL SUGGESTIONS FOR PRESCRIBED SCHOOL ACTIVITIES

If unlimited resources were available, what strategies, materials, plans and/or programs would you recommend?

For each objective, please circle the number which best expresses your opinion of the degree of usefulness of the School Mastery of Teading Test.

ATTACHMENT # 12

	NOT AT ALL USEFUL						EXTREMELY USEFUL
Measure educational status of individuals.....	1	2	3	4	5	6	7
Measure educational "growth" of individuals.....	1	2	3	4	5	6	7
Identify remedial cases.....	1	2	3	4	5	6	7
Select students for bilingual training.....	1	2	3	4	5	6	7
Screen special education students.....	1	2	3	4	5	6	7
Help plan instruction for individuals.....	1	2	3	4	5	6	7
Help plan instruction for class groups.....	1	2	3	4	5	6	7
Detect system-wide general strengths/weaknesses...	1	2	3	4	5	6	7
Help evaluate teaching procedures or methods.....	1	2	3	4	5	6	7
Help evaluate instructional materials.....	1	2	3	4	5	6	7
Help evaluate teacher performance.....	1	2	3	4	5	6	7
Compare students with national peer groups.....	1	2	3	4	5	6	7
Compare classes within a school.....	1	2	3	4	5	6	7
Compare Schools within a system.....	1	2	3	4	5	6	7
Compare a system with systems across the country..	1	2	3	4	5	6	7
Assess school performance.....	1	2	3	4	5	6	7
Provide information for state funding.....	1	2	3	4	5	6	7
Determine minimum competency standards.....	1	2	3	4	5	6	7
Modify curriculum.....	1	2	3	4	5	6	7
Report to newspapers.....	1	2	3	4	5	6	7
Report to boards of education.....	1	2	3	4	5	6	7
Report to parents.....	1	2	3	4	5	6	7
Report progress to students.....	1	2	3	4	5	6	7

****
THANK YOU FOR YOUR COOPERATION.
****

0295U

MTE's

CURRICULUM PRESENTATION

COMMUNICATION ARTS GRADE 4

SAMPLE 1

CONTENT AREA AND TOPIC	SKILLS AND KNOWLEDGE
Exposition	Factual materials give useful knowledge about the world.
PROCESSES Following skills generally applicable to content areas as well as to literature	
Comprehension	
Literal	Details may be used to support a point. **p. 26
Inferential	Details may be inferred from the context. **p. 31 Characters can be analyzed by their actions. **p. 30 Outcomes may be predicted based upon stated fact. **p. 30
Critical	Personal experiences may be related to a narrative or expository selection. **p. 34 Fact and opinion may be distinguished based upon clue words such as: in my opinion and I believe that. **p. 34
Word Meaning	
Vocabulary Development	Vocabulary also expands through including content area words. p9 Words evoke emotional and sensory responses. **p. 9
Contextual Clues	Word meaning may be found contextually through comparison and/or contrast. **p. 22
Structural Analysis	Words of more than one syllable can be divided between the syllables. **p. 20
Phonic Analysis Should have been completed by this grade except for individual students who need reinforcement	(For detailed information on the various phonemes see the tables appended to this bulletin.)

ATTACHMENT #13

CURRICULUM PRESENTATION

SAMPLE 2

SKILL	LEVEL B	LEVEL C	LEVEL D
B. Finding and Relating Details	Develops the understanding that a series of connected sentences (spoken or written on this level of maturity and understanding) contain details which are related to each other and to the main thought.	Develops the concept that details fill out, extend, make specific, exemplify, clarify, and document the main idea.	Develops the ability to use details in preparing material for oral or written presentation; develops the ability to use details in making a point.
	Language Arts: Learns to find specific details in simple reading material by finding answers to questions, e.g., What did the children bring to the party?	**Language Arts:** Reads vivid passages, recalls details, and describes or draws simple pictures including details.	**Language Arts:** Answers questions (taken from material read) asking for details—who, what, where, when, how big, what color.
	Science: Picks out, from an experience chart, details related to the main idea, e.g., What were some of the things our magnet could pick up?	**Science:** Reads and follows detailed step-by-step directions in order to perform a simple experiment, e.g., Plants Need Water Materials: 2 four-inch pots 2 similar plants soil water Directions: 1. Place soil in each pot. 2. Place plants in soil. 3. Every day, water one pot only. Question: What did the plants look like on the first day? after a week? after 2 weeks?	**Science:** Uses details from printed materials to prove or refute a point, e.g., wheels move things more easily.
	Social Studies: Reads simple text for answers to questions, e.g., In what way does the policeman help us?	**Social Studies:** Answers questions based on material read, e.g., We all know that different materials are used in constructing buildings. In what ways are these materials used?	**Social Studies:** Supplies details from material read under a stated main idea, e.g., In different cultures around the world, some types of workers are always needed: 1. Workers to get food. 2. Workers to make clothing. 3. Workers to build homes.

ATTACHMENT #14

CURRICULUM PRESENTATION

SAMPLE 4

COMMUNICATION ARTS GRADE 3

CONTENT AREA AND TOPIC	SKILLS AND KNOWLEDGE
Exposition	Factual materials give useful knowledge about the world.
PROCESSES Following skills generally applicable to content areas, as well as to literature	
Comprehension	
Literal	Stated details in content area materials may be isolated as sub-headings in chapters or parts of chapters.** p. 27 The main idea of a selection may be directly stated in various locations within the selection. ** p.24 Clue words such as first, next, and last indicate sequential order of events. ** p.28 Clue words such as because and as signal cause and effect relations. ** p.29
Inferential	Cause and effect relations not directly stated may be determined. ** p.31 Sequential order is a factor in predicting outcomes and drawing conclusions. ** p.28
Critical	Experiential background and knowledge affect evaluation of written materials. ** p.34
Word Meaning	
Vocabulary Development	Vocabulary also expands through including content area words. ** p.9
Contextual Clues	Word meaning may be found from the context in a direct statement, through a synonym or through apposition. ** p.22
Structural Analysis	A prefix comes before a root word and changes the meaning of the word. ** p.18 A suffix comes at the end of a root word and changes the meaning of the word. ** p.18 Comparatives and superlatives change the meanings of adjectives. (er, est, more, most). ** p.18

CURRICULUM PRESENTATION

SAMPLE 5

SKILL	LEVEL B	LEVEL C	LEVEL D
C. Determining Sequence	Develops the understanding that a series of events occurring within one episode may be ordered sequentially. **Language Arts:** Listens to a simple story and is guided to develop an understanding of sequence by answering such questions as: What happened first? What happened next? **Social Studies:** Places pictures in sequential order to develop a simple time line. **Science:** Participates in the writing of experience charts based on firsthand observations: How long is a day in fall, winter, and spring?	Recognizes the relationship of sequence to meaning. **Language Arts:** Reproduces story heard or read by drawing several consecutive pictures of events in sequential order. **Science:** Plans and lists (in proper sequence) steps needed to carry out a science experiment. **Social Studies:** Reads several sentences describing an event connected with the unit under study; arranges these in correct order, e.g., When people needed to travel, they used stagecoaches drawn by one or two horses. The first European settlers traveled from place to place on foot, on horseback, or in wagons drawn by horses. When the Indians lived in New York, they traveled on foot. Later, people traveled long distances on trains drawn by steam engines instead of horses.	Uses sequential order to make interpretations and predictions; recognizes key words as clues to sequential development. **Language Arts:** Predicts outcome of story stemming from logical sequence of events; picks out key words or phrases which highlight sequence, e.g., now, then, after a while, today, yesterday, tomorrow, in the first place, in the second place, etc. **Science:** Is able to renumber or rewrite in sequential order facts taken from text. **Social Studies:** Studies two pictures of related events in growth of towns, cities, or nations to determine time span and sequence of events between them. **Physical Activities:** Gives facts sequentially when explaining how to play a game.

ATTACHMENT #1