DOCUMENT RESUME

ED 360 341 TM 020 207

AUTHOR Barrera, Marbella; Holst, Patricia
TITLE District-Wide Testing Results, Technical Report, 1991-92. OPET Report.
INSTITUTION Newark Board of Education, NJ. Office of Planning, Evaluation and Testing.
PUB DATE Aug 92
NOTE 109p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Academic Achievement; *Achievement Tests; Bilingual Education; Curriculum Evaluation; Elementary Secondary Education; Language Proficiency; Limited English Speaking; Mathematics Achievement; Public Schools; Reading Achievement; School Districts; Scores; Spanish; *Standardized Tests; State Programs; Student Evaluation; *Testing Programs; *Test Results
IDENTIFIERS Bilingual Testing Program NJ; Early Warning Test NJ; Newark School System NJ; *Newark Uniform Testing Program NJ; *New Jersey Statewide Testing System; Stanford Achievement Tests

ABSTRACT

Results from testing programs in the Newark (New Jersey) public schools for 1991-92 are discussed. The system used three testing programs in the school year: (1) the Newark Uniform Testing Program (Stanford Achievement Series, eighth edition); (2) the New Jersey Statewide Testing System; and (3) the Bilingual Testing Program. Results of the Newark Uniform Testing program indicate that students are steadily improving, but remain below grade level, with mathematics the weakest area and reading the strongest. Kindergartners were on grade level in reading and mathematics. The Early Warning Test component of the state testing system indicated that mathematics and writing scores for eighth graders improved overall, but that many students would benefit from remedial services. The High School Proficiency Test component of this program becomes a graduation requirement after one more administration. The district remained below state means on this test. Results from the Bilingual Testing Program indicate that limited English proficiency students had lower scores than did English proficient students and scored below grade level on a Spanish language achievement test. This review simplifies the complex pattern of test results, but is a first step to identifying strengths and weaknesses of the curriculum. Study findings are summarized in 25 tables and 10 figures. An appendix contains an additional seven tables of test results. (SLD)

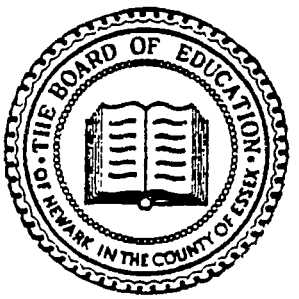
NEWARK BOARD OF EDUCATION
August, 1992

DISTRICT-WIDE TESTING RESULTS
Technical Report
1991-1992

Prepared by

Marbella Barrera
Testing Supervisor
and
Patricia Holst
Research Consultant

Office of Planning, Evaluation, and Testing
2 Cedar St.
Newark, NJ 07102



OFFICE OF PLANNING, EVALUATION, AND TESTING STAFF:

Elaine Walker, Ph.D.
Supervisor

Jann Azumi, Ph.D.
Director

Marbelia Barrera, M.S.
Supervisor

Rosemarie Kopacsi, Ph.D.
Sr. Research Assistant

Gwendolyn Johnson
Secretarial Assistant

Robert Mckenzie, M.S.
Sr. Research Assistant

Alvina Conyers, M.S.
Research Assistant

Sherleen Hillman
Statistical Typist

Beverly Collins
Research Assistant

Christine Santos
Sr. Bilingual Clerk Typist

Wendell Gibbs
Sr. Stock Clerk

Ernette Pinkston
Word Processor Operator

EXECUTIVE SUMMARY

- The Newark Public School System employed three specific testing programs during the 1991-1992 academic year: the Newark Uniform Testing Program, the New Jersey Statewide Testing System, and the Bilingual Testing Program.

- Results from the second administration of the Stanford 8, the Newark Uniform Testing Program, indicate that while students are steadily improving in all three areas they continue to be below grade level. Mathematics remains our strongest area while reading is our weakest. This year's results showed the greatest gain in reading which indicates progress toward strengthening that area.

- Kindergartners participated in the Stanford 8 this year. They performed very well and were on grade level in both reading and mathematics. The reading component of their exam consisted of only one subtest, Sounds and Letters.

- The second administration of the Early Warning Test (EWT) to our eighth graders presented some findings which were questionable. Math and Writing scores showed improvement over the year. Indications are that a substantial number of students would benefit from remedial services.

- Students in the eleventh grade participated in the second administration of a "due-notice" Grade 11 High School Proficiency Test. There will be one more administration of this test before it becomes a graduation requirement. Relative to the state mean, the district performed best in the area of writing followed by math and reading. Attendance rates improved slightly over the 1991 administration. The district was significantly below the state means in all three areas.

- The Bilingual Testing Program results indicate that Newark's Limited English Proficiency students tested with the Stanford 8 received lower scores than Newark's English proficient students. The strengths and weaknesses of both populations were similar. Students who took Aprenda, the Spanish language achievement test scored slightly below grade level. The Portuguese Achievement Test had most students scoring above the cutoffs.

- This summary review simplifies the complexity of the pattern of our district. Presentation of test results is the first step in identifying strengths and weaknesses of curriculum. The most vital element at the school level is to provide an optimal learning environment that emphasizes thinking and problem solving. Each student is entitled to the best education possible. This means fair and accurate identification of individual strengths and weaknesses as well as remediation and enrichment of these respective areas.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER II

CHAPTER III

CHAPTER IV

LIST OF FIGURES

CHAPTER II

CHAPTER III

CHAPTER IV

I. INTRODUCTION

This report presents the Office of Planning, Evaluation and Testing interpretation and assessment of the 1992 results of Newark's Uniform Testing Program, The New Jersey Statewide Testing Program, and the Bilingual Testing Program.

The information contained in any test result can act as a springboard for growth and change in our schools. Identification of strengths and weaknesses in curriculum can be clarified through test results. Academically talented students as well as students experiencing academic problems can be provided with enrichment and remediation as a result of testing.

Looking at test results is like looking at a snapshot. The snapshot is a representation of how the district looks at one point in time. It is a static representation of a dynamic process and as such can never perfectly depict what is happening. A well-focused snapshot can give a good picture but no matter how good the picture, a snapshot is not sufficient to describe a process. Like a snapshot, well focused test results can provide a good picture. The limitations, however, are that tests provide a static measure and can never be considered sufficient to completely represent what a student knows. Other criteria can, and should be considered when evaluating students.

Test results represent how a student is performing on specific objectives at a certain point in the school year. Attempting to teach directly to the test is analogous to spending all year getting ready to take a family picture. While

you may get a favorable picture, you have masked what is really happening and created an artificial facade. The elements which create a positive educational environment are greater than what is selected to be on the "test". It would be a sad day for everyone in education when the product of test results supersedes the process of learning.

Many districts throughout the United States are unhappy with the present state of assessment and are considering alternative methods. One method known as performance assessment requires students to perform the actual behaviors of interest. Students are then given a score based on their "performance". This method of assessment, like the learning process itself, is dynamic and is considered by some to be a more authentic measure of the skills and knowledge that a student has.

Many of the standardized tests employed by Newark this year have contained a performance element. Students taking the HSPT11 and the EWT are required to construct answers to certain questions rather than to select a multiple choice option. Complete information on the effectiveness of these alternative forms of assessment is not yet available but the Newark Office of Planning, Evaluation and Testing is investigating the feasibility of these methods by keeping updated on the current trends in this field.

The testing program used in the Newark district is designed to represent the different needs of the student population. The Bilingual Testing Program assesses the progress of students identified as having a limited proficiency of

English. A minimum passing score establishes this as a criterion referenced test where the goal is mastery of the English language. Spanish and Portuguese speakers in this population take academic achievement tests in their native language. Other students take these tests in English.

The New Jersey Statewide Testing Program is a group of tests (HSPT9, EWT and HSPT11) which are considered "high stakes" tests. This means there are consequences to the test which may affect a student's future. The motivational component of test-taking becomes very important in a "high stakes" testing situation. Current research indicates that motivation impacts test performance. One high stakes test which will become operational in the state of New Jersey is the HSPT11. All students in the state will be required to pass this test in order to receive a New Jersey state diploma. The present version of the HSPT11 is in an experimental stage so that there are no student consequences attached to the results. It is given to provide experience to students on the format of the test and to give information to schools and districts throughout the state on the strengths and weaknesses of their high school programs. There is one more year of experimental testing before the high stakes version of the HSPT11 is used. We should look at the results of this experimental testing carefully to determine what steps can be taken to improve the chances that all students in our district will pass the HSPT11. One way to do this is by teaching the higher order thinking and problem solving skills which this test is tapping. It will no longer be sufficient for students to

have basic mastery over factual and lower level skills. In order to pass the HSPT11 it will be necessary for students to be critical, independent thinkers able to solve problems, explain answers and communicate effectively. These are good goals to have but we as a district are far from achieving them.

Unfortunately testing often places a burden on teachers, principals, and administrators to "demonstrate" that their school or class has shown improvement. Test results are sometimes used as a measure of accountability. Parents want to see test scores rise. Teachers want to see test scores rise. Principals want to see test scores rise. This creates a 'high stakes' situation which may divert us from our primary purpose of teaching and evaluating students. The primary goal of testing is to provide the information necessary to help our students.

This report is organized in three additional chapters. Chapter 2 discusses the results of the Stanford 8. Chapter 3 presents the outcome of the state mandated tests. Chapter 4 analyzes performance on the Language Assessment Battery, Aprenda, and the Portuguese Achievement test.

II. THE NEWARK UNIFORM TESTING SYSTEM

The Stanford Achievement Series, Eighth Edition (Stanford 8), was the Newark district-wide test employed in the 1991-92 academic year. In 1990, members of a representative Test Review Committee evaluated a number of standardized tests and selected the Stanford 8 as the most appropriate for the district.

One of the key aspects considered when selecting the test was that it provides an optimum link with the Newark curriculum. A second consideration was that it is in line with the new state-mandated tests, the Early Warning Test (EWT), and the "due-notice", Eleventh Grade High School Proficiency Test. These tests explore reading, math, and language skills, with an emphasis on understanding and decision-making. They provide a variety of reading passage types, which allows for a more complete assessment of reading comprehension. The math subtests reflect the emphasis given to estimating and problem solving.

The Stanford 8 is a norm-referenced test. This means that students taking the test are compared to a norming population. Norms were collected in the Spring and Fall of 1988. The total standardization and research sample, also known as the reference group, consisted of approximately 600,000 students. The group represented 50 states, three socioeconomic statuses (high, middle, low), and three community types (urban, suburban, rural).

Scores for individual students are compared to this norming group and percentiles are computed.

The Stanford Achievement Series consists of 13 levels: SESAT (Levels 1 and 2 for Kindergarten), Primary (Levels 1, 2 and 3 for grades 1, 2 and 3), Intermediate (Levels 1,2 and 3 for grades 4, 5, and 6), Advanced (Levels 1 and 2 for grades 7 and 8) and TASK (Levels 1,2 and 3 for grades 9 through 12). Each grade takes a different level of the test with the exception of grades 11 and 12.

This section of the report examines student performance on the Stanford 8 in terms of the regular population (English proficient students) followed by the results of the limited English proficient (LEP) students. The outcomes are presented in terms of Normal Curve Equivalent (NCE) scores or percentile ranks. Percentile ranks range from a low of 1 to a high of 99 with 50 denoting average performance. The percentile rank corresponding to a given score indicates the percentage of a reference group obtaining scores less than or equal to that score. For example a student who performed as well as or better than 65 percent of the students in the reference group would earn a percentile rank of 65. Percentiles must always be interpreted with reference to the group from which they were derived. NCEs are derived from percentile ranks. In contrast to percentiles, NCEs provide an equal-interval scale; thus, they should be used instead of percentiles when interpolating or averaging scores. Use of NCEs allow for the comparison of performance across subtests, levels, and

forms, and for evaluation of students' performance in relation to their peers.

2.1. REGULAR STUDENTS

The District tested 35,530 regular students at the elementary, middle and secondary level in 71 schools. Their results are discussed on three levels: Total scores, subtest scores, and proficiency rates. Comparisons with last years' results will be examined.

2.1.1. Total Scores

Overall this year's performance was similar to last year's. Newark students scored highest in Mathematics (Mean NCE = 46.4), followed by Language (Mean NCE = 42.0) and Reading (Mean NCE = 39.5). The greatest growth, district wide, was seen in Reading (NCE change = +1.36). Grade 12 students are not presented here as only a limited sample (240) were tested.

Table 2.1.1 presents summary information regarding students' performance on the 1992 administration of the Stanford 8. Review of the results indicates that the best results are obtained by younger students. In mathematics, students up to the third grade are performing at grade level. The highest scores were obtained by second graders, with a mean NCE of 51.6. In language, first graders performed best with a mean NCE of 46.0. Kindergartners obtained the best results in reading with a mean NCE of 49.3.

TABLE 2.1.1
STANFORD 8
1992 MEAN NCE TOTAL SCORES FOR REGULAR STUDENTS

GRADE	READING		MATHEMATICS		LANGUAGE	
(*)	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
K	49.3	(23.7)	49.9	(23.8)	-	-
1	42.9	(19.3)	48.6	(22.6)	46.0	(23.1)
2	43.4	(18.4)	51.6	(21.7)	43.7	(19.8)
3	43.3	(17.9)	49.6	(19.8)	44.6	(17.9)
4	36.5	(18.5)	46.9	(19.0)	43.3	(16.8)
5	36.9	(15.9)	44.2	(18.8)	41.4	(16.9)
6	38.3	(16.2)	45.6	(18.5)	41.2	(15.4)
7	38.8	(16.1)	43.8	(16.7)	40.2	(15.6)
8	39.6	(16.5)	43.4	(18.0)	41.3	(16.8)
9	36.6	(16.3)	41.7	(17.0)	38.9	(16.0)
10	35.0	(16.9)	42.4	(16.8)	35.6	(16.7)
11	37.3	(17.1)	44.0	(16.9)	39.4	(16.4)
TOTAL	39.5	(17.1)	46.4	(19.7)	42.0	(17.8)

(*) Twelfth graders are not shown because only a small, non-representative sample of the population took the test.

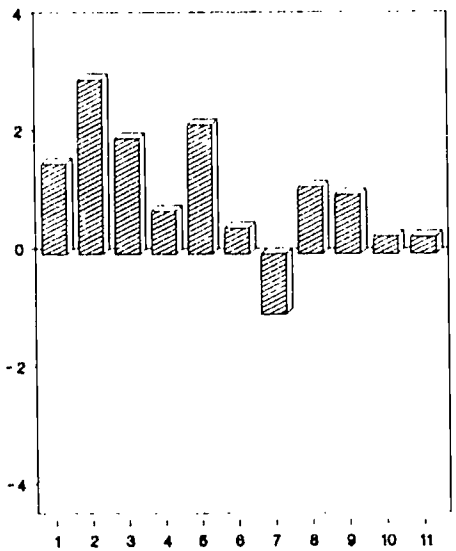
Their score, however, reflects only the Sounds and Letters subtest and not a total reading score.

Figure 2.1.1 shows a comparison of students' performance on the 1992 and 1991 administrations of the Stanford test. Review of the reading results indicate that the greatest improvement in mean NCE scores was shown by Grade 2 (NCE change = +2.96) and Grade 5 (NCE change = +2.19). All grades experienced an improvement over last year with the exception of Grade 7 (NCE change =-1.02). Overall the total district improvement for Reading was NCE change +1.36. In math, the greatest improvement in mean NCE scores was shown by Grade 2 (NCE change = +3.49). Grade 9 experienced a substantial decrease (NCE change = -4.20). Overall the district showed improvement in Mathematics (NCE change = +0.87). In language, the greatest improvement in mean NCE scores was shown by Grade 1 (NCE change = +3.13), Grade 2 (NCE change = +2.80) and Grade 5 (NCE change = +2.00). All grades experienced a rise in NCE scores with the exception of Grades 9 and 10. Overall the district saw a rise in Language scores (NCE change = +1.25).

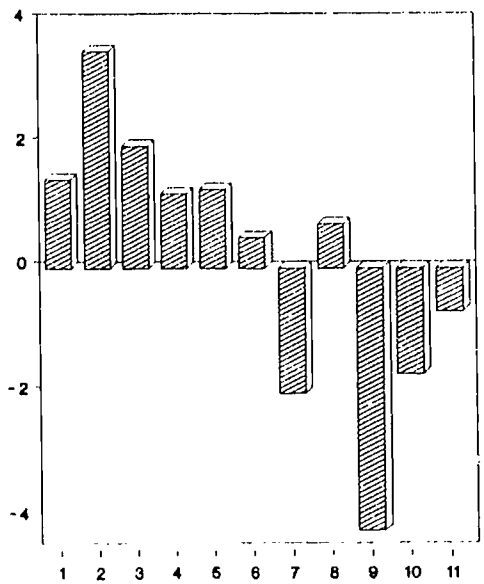
Tables A-1 through A-3 in the Appendix present detailed information on the comparison of the 1991-1992 mean NCE total scores for all areas.

Stanford 8
Comparison of 1991-1992 Results

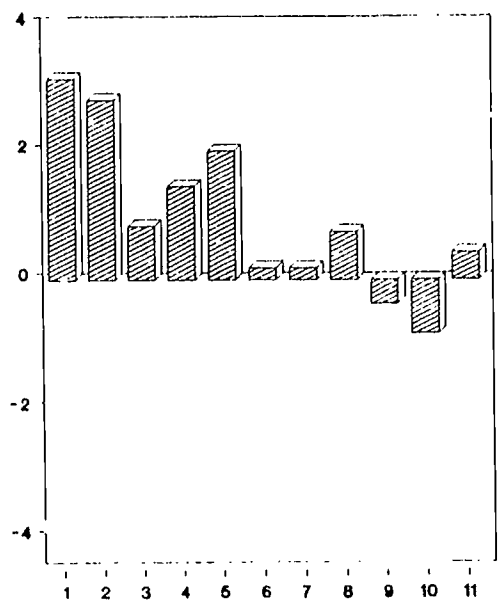
READING



MATHEMATICS



LANGUAGE



(X axis = Grades)
(Y axis = Change in Mean NCE Scores)

Figure 2.1.1

2.1.2. Subtest Scores

Stanford 8 uses subtests to assess the areas of reading, math and language achievement. In grades 1 through 3 reading is evaluated by three subtests: Word Study Skills, Reading Vocabulary, and Reading Comprehension. In grades 4 through 12 only 2 subtests Reading Vocabulary and Reading Comprehension are used. Math is measured in first through eighth grades with three subtests: Concepts of Numbers, Computations, and Applications. High school students are evaluated using only 1 subtest, Mathematics. Only 1 subtest (Language) is used to assess first and second graders. English is the sole subtest for high school students. Students in grade 3 through grade 8 are given 2 subtests; Language Mechanics and Language Expression.

Table 2.1.2 depicts the summary information for the Reading subtests along with comparisons of last year's results. First and third graders performed better in Word Study Skills. Higher scores were obtained in Reading Comprehension than in Reading Vocabulary for every grade except second through fourth, and eleventh.

Table 2.1.3 presents a comparison of the math subtests for the years 1991-1992. Students at all levels improved their math scores substantially, with the exception of high school. Their best scores were achieved on the Computations subtest, which was within grade level from first through sixth.

TABLE 2.1.2
STANFORD 8
MEAN NCE SCORES IN READING SUBTESTS FOR 1991 AND 1992

Grade	Word Study		Rd. Vocabulary		Rd. Comp.	
	1991	1992	1991	1992	1991	1992
1	42.8	46.3	40.7	42.3	42.5	43.4
2	43.4	44.8	41.3	46.0	39.4	42.0
3	44.4	46.3	41.3	44.0	41.1	42.4
4			37.1	38.4	37.4	37.6
5			32.4	36.5	38.7	39.3
6			36.3	38.0	40.3	39.9
7			39.7	38.4	40.9	40.3
8			37.9	39.6	40.6	41.2
9			36.8	37.0	36.5	37.6
10			34.4	34.7	37.0	36.8
11			37.9	37.3	37.1	37.0

TABLE 2.1.3
STANFORD 8
MEAN NCE SCORES IN MATH SUBTESTS FOR 1991 AND 1992

Grade	Num. Concept		Computations		Applications	
	1991	1992	1991	1992	1991	1992
1	48.6	49.3	47.5	50.6	46.6	47.8
2	49.7	51.0	49.4	54.6	45.4	48.2
3	47.5	49.4	51.4	55.1	45.1	45.3
4	45.2	46.2	48.4	51.3	44.4	44.3
5	43.1	44.8	46.7	48.5	40.9	40.8
6	43.6	44.7	48.3	49.3	43.9	43.5
7	43.4	43.8	48.1	46.5	44.3	46.5
8	40.1	41.5	44.8	45.7	43.8	43.9
9					41.7	38.5
10					38.5	37.1
11					40.3	39.6

As shown on Table 2.1.4, students performed better in the Language Expression than in the Language Mechanics subtest. These results are in line with last year's and seem to indicate that Newark students are performing better in the areas of reading and language that reflect higher order skills such as Reading Comprehension and Language Expression.

TABLE 2.1.4
STANFORD 8
MEAN NCE SCORES IN LANGUAGE SUBTESTS FOR 1991 AND 1992

Grade	L. Mechanics		L. Expression	
	1991	1992	1991	1992
3	43.0	44.5	45.7	45.7
4	40.1	41.0	44.2	46.1
5	38.7	40.0	41.7	44.1
6	38.2	39.0	45.2	45.0
7	39.5	39.5	41.3	41.7
8	38.2	38.6	44.3	45.3

2.1.3. Proficiency Rates

The State of New Jersey Department of Education has established minimum levels of proficiency (MLP) for each of the three global areas at grades three through nine. The Office of Planning, Evaluation, and Testing of the Newark Board of Education provided the cutoff scores for grades one and two. Thus, for every grade and each area (reading, language and math), there is a score representing the minimum level of proficiency. The percentage of

students who score at or above this value is referred to as the proficiency rate. The District established proficiency rates of 75 percent for reading and math as priority objectives for grades three and six. Since the HSPT9 is no longer a graduation requirement, the District also established proficiency rates of 65 percent in the reading, math, and writing subtests of Stanford for ninth graders. Students not achieving the MLP in a particular area are entitled to receive Chapter I services in that area. Table 2.1.5 lists the proficiency rates for the current year.

TABLE 2.1.5
STANFORD 8
PERCENTAGE OF REGULAR STUDENTS ABOVE STATE MLP

GRADE	N	READING	MATH	LANGUAGE
1	3,555	55.2%	53.1%	68.7%
2	3,359	62.7%	76.7%	74.6%
3	3,438	72.2%	55.6%	71.9%
4	3,306	66.5%	60.1%	71.0%
5	3,368	68.3%	48.2%	70.7%
6	3,366	62.2%	52.0%	71.7%
7	3,134	77.4%	55.6%	53.5%
8	2,967	73.2%	56.8%	66.6%
9	2,353	44.9%	38.4%	46.4%

As shown in Table 2.1.5, the district has not yet achieved its goals in any of the three targeted levels. However, the results are somewhat promising. The percentage of students meeting the state standards increased significantly

in all three grades. As to the district objectives, third graders are very close to reaching the reading goal. As curricular changes start producing their desired effect, and as our students become more familiar with the format of the Stanford 8, the district will get closer to achieving its objectives.

2.1.4. Study Skills

Study Skills is a subtest devised by the publishers of Stanford 8 for exploring academic skills independent of content. This subtest assesses student's ability to locate and use information in reference sources of tables of contents and indexes, which is considered critical to the learning process. The subtest is administered to students in grades three through twelve. Results of this year's administration were homogeneous across grades. Best performers were students in grade three (Mean NCE = 43.3), and lowest achievers were students in grade ten (Mean NCE = 37.8).

2.2. LIMITED ENGLISH PROFICIENT STUDENTS

A total of 3,018 students classified as of limited English proficiency (LEP) were given the Stanford 8. The group included high school students and the following subgroups at the elementary and middle level:

- Students with less than three years in a bilingual program, who achieved the District cutoff in the Spring administration of the LAB.
- Spanish- and Portuguese-speaking students with more than three years

in a bilingual program.

- Speakers of languages other than English, Spanish, or Portuguese - regardless of time in program.

This section will discuss proficiency rates and performance of the LEP students in relation to the regular students. Section 4.4 in the Bilingual Chapter reviews general information concerning this population, overall performance, and the impact of variables such as time in the program, language, school, and grade level.

2.2.1. Proficiency Rates

Minimum levels of proficiency (MLP) as described in Section 2.1.3 were also used to assess the eligibility of LEP students to Chapter I services. Based on their Stanford 8 performance, 1,489 students in the group of first through ninth graders will be offered remedial services in reading, 1,259 in language, and 1,114 in math. Table 2.2.1 presents the actual percentages for each area and grade level.

As indicated by Table 2.2.1, the best performers were second graders in math where 69.0 percent of the students achieved the MLP. The group that had the weakest performance in all three areas were the ninth graders. Less than 10 percent of them achieved the MLP in reading and language. About half of this group had been in the bilingual program less than 18 months.

TABLE 2.2.1
STANFORD 8
PERCENTAGE OF LEP STUDENTS ABOVE STATE MLP

GRADE	N	READING	MATH	LANGUAGE
1	355	50.0%	63.9%	58.6%
2	533	33.2%	69.0%	53.8%
3	338	46.2%	42.3%	53.3%
4	261	34.1%	55.6%	50.6%
5	206	24.8%	36.9%	34.0%
6	157	26.8%	32.5%	44.6%
7	144	38.2%	45.1%	21.5%
8	138	31.9%	41.3%	26.8%
9	155	2.2%	26.5%	8.4%

2.2.2. Comparison with Regular Students

Newark LEP population performed similarly to regular students, better in math than in any other area. Equally, they achieved higher scores in the lower grades than at the middle or high school level. Figure 2.2.1 shows both populations side by side.

Performance of both groups were similar in math. The smallest differences were seen at the first grade level. While LEP students had a mean of 48.2 NCEs, their English proficient counterparts averaged 48.6 NCEs, for a .4 NCE difference in favor of the regular students. Eighth grade LEP students achieved a mean of 36.4 NCEs, whereas the English proficient students obtained a mean of 43.4 NCEs. The difference of 7 NCEs was the largest in

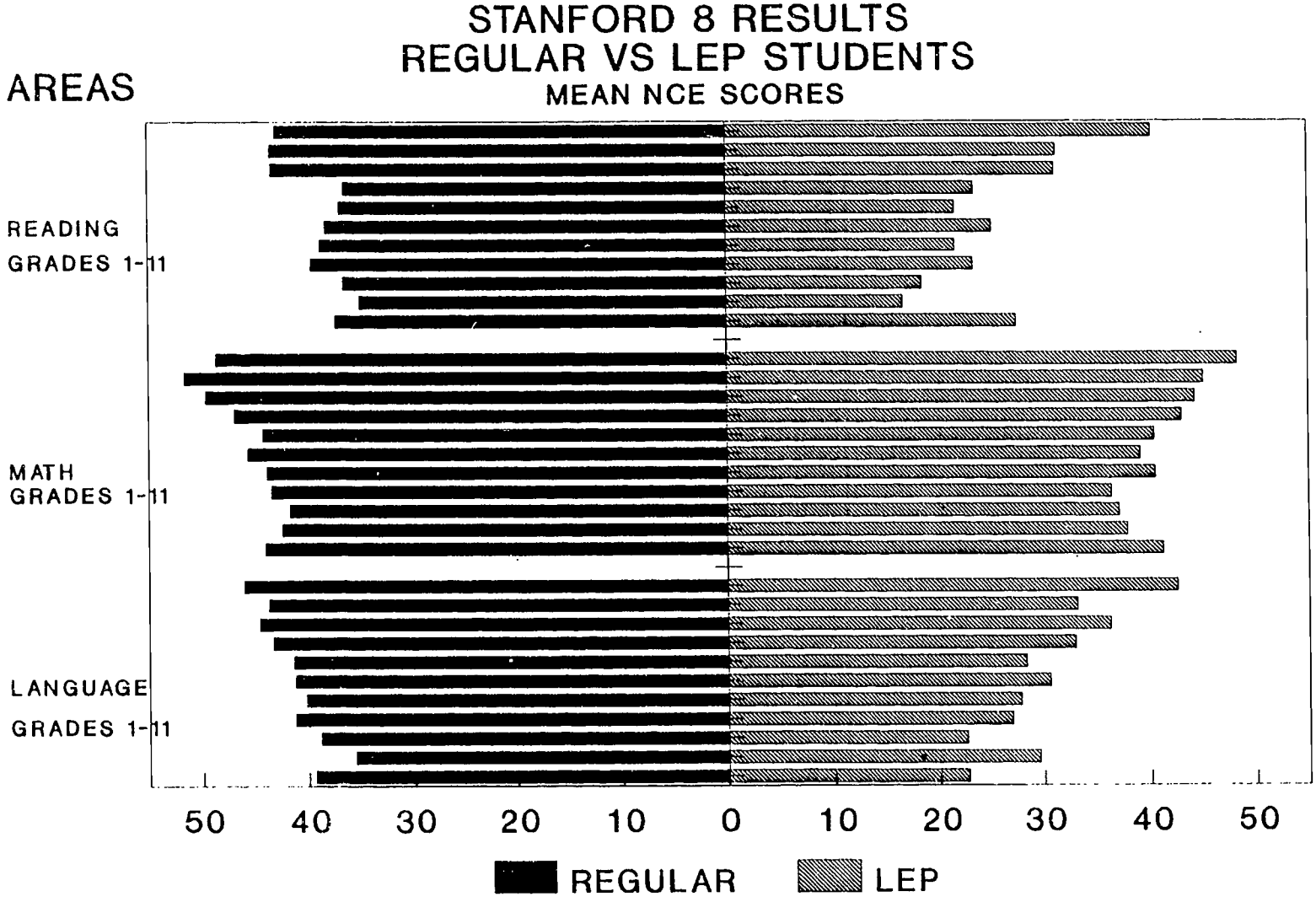

STANFORD 8 RESULTS
REGULAR VS LEP STUDENTS
MEAN NCE SCORES
AREAS
READING
GRADES 1-11
MATH
GRADES 1-11
LANGUAGE
GRADES 1-11
50
40
30
20
10
0
10
20
30
40
50
REGULAR
LEP

Figure 2.2.1

reading.

First graders also had the smallest differences in language and reading scores. Differences in language scores for other grades ranged from 8.3 NCEs for second graders to 16.3 NCEs for the ninth grade level. Differences in reading scores ranged from 9.8 NCEs for eleventh graders to 18.2 NCEs for ninth graders. (We should note that, unlike the regular students, the performance of LEP eleventh graders was higher than that of the ninth and tenth graders.) What these differences indicate is that the gap between LEP and regular students grows wider as the students progress through the school levels. They also suggest that this gap may be a function of the limited proficiency in the English language.

Kindergartners of limited English proficiency (not shown in Figure 2.2.1) were also evaluated with SESAT 1. As a group, the 307 students achieved a mean score in Reading of 43.3, six NCEs below the mean of their peers in the regular program. In math, their average score was 44.2, 7.7 NCEs lower than the English proficient students. Since this test is basically graphic, their difficulties most likely stem from misinterpreting teachers' directions.

III. THE NEW JERSEY STATEWIDE TESTING SYSTEM

The New Jersey Statewide Testing System has its legal basis on the 1976 Amendment of the Public School Education Act which established uniform standards of minimum achievement in basic communication and computation skills. Beginning in 1981 ninth grade students were required to pass the Minimum Basic Skills (MBS) test as one of the conditions for a high school diploma. For students entering ninth grade in 1985, the requirement was changed to passing the more rigorous High School Proficiency Test (HSPT9).

Students who entered high school in 1991 or later, have their graduation requirement moved to the eleventh grade (HSPT11). This test involves the assessment of higher order skills such as thinking, problem solving, reasoning, and decision making. Since this new test will be administered at a later point in the student's high school career, the State considered it critical to make an early identification of those at risk of not passing it. Thus, the eighth grade "Early Warning Test" (EWT) came into existence.

During the academic year 1991-1992, in compliance with the Statewide Testing System, the District administered the HSPT9, the HSPT11, and the EWT. Each test consisted of three sections: Reading, Mathematics, and Writing, which included the writing of an essay on a given topic. In criterion-referenced tests such as these, a minimum standard of performance is

established and students achieving it are deemed to be successful on the test. The State has established such standards for the HSPT9 and for the EWT but since the HSPT11 is still a "due-notice", or practice test, no standard has yet been determined for it. What follows is a summary of the district's results in these tests.

3.1. GRADE 8 EARLY WARNING TEST (EWT)

The purpose of the EWT is to determine whether students are making sufficient progress in mastering the skills they will need to pass the Grade 11 High School Proficiency Test. The test, administered for the second time in March 1992, emphasizes students' skills in problem solving and decision-making. The State Department of Education reports EWT results as scale scores or percent correct responses, and warns that they are not to be compared across content areas. We will discuss the outcome in terms of scale scores, because they permit a direct comparison of results over time, which provides an adequate frame of reference. We will also present the proficiency rates, which indicate the percentage of students who achieved the State-established minimum levels of proficiency (MLP) for this test.

3.1.1. Overall Results

The metric selected to report student performance is known as scale scores. EWT scale scores have a mean of approximately 500 and a standard deviation of 50-60. It is important to keep in mind that these scores are not comparable across content areas; that is, a scale score of 412 in Reading is not the same as a scale score of 412 in Math. Table 3.1.1 presents mean scale scores for the District and each individual school in reading, math, and writing. Schools are ranked based on a combined score of all three areas, computed exclusively for this purpose. Nine of the top 10 schools last year are part of

Table 3.1.1
EWT 1992 RESULTS (MEAN SCALE SCORES)

SCHOOL	READING	MATH	WRITING
UNIVERSITY H.S.	508.0	501.9	501.9
LAFAYETTE STREET	462.3	474.2	498.5
OLIVER STREET	448.4	458.7	501.2
ANN STREET	454.3	476.9	476.6
WILSON AVENUE	436.9	469.7	498.0
FIRST AVENUE	456.6	475.5	469.7
RIDGE STREET	443.4	466.8	463.1
ABINGTON AVENUE	438.6	461.7	459.9
MILLER STREET	449.8	441.7	467.6
MOUNT VERNON	439.3	442.9	459.6
HAWKINS STREET	428.6	432.7	476.3
DR. E. ALMA FLAGG	432.9	430.4	463.3
QUITMAN STREET	430.7	429.8	463.3
NEWTON STREET	417.7	413.5	492.2
CHANCELLOR AG.	431.0	432.1	450.9
G. W. CARVER	423.7	438.7	451.4
BRAGAW AVENUE	416.5	443.1	446.1
MCKINLEY	406.9	435.6	455.3
MAPLE AVENUE	420.8	422.2	453.2
SUSSEX AVENUE	416.2	431.0	447.4
DAYTON STREET	418.9	411.7	461.4
CAMDEN M.S.	417.2	429.4	443.5
BURNET STREET	408.1	430.3	442.9
PESHINE AVENUE	406.2	439.6	435.5
DR. WILLIAM HORTON	408.3	431.1	440.7
WARREN STREET	400.3	408.7	470.9
BERGEN STREET	406.4	409.7	444.7
MORTON STREET	390.0	428.5	441.1
HAROLD WILSON	396.8	415.5	446.9
THIRTEENTH AVENUE	399.2	411.6	441.3
VAILSBURG M.S.	409.8	420.2	419.2
HAWTHORNE AVENUE	406.8	409.7	423.9
LUIS MUNOZ MARIN	398.4	410.9	430.8
SOUTH SEVENTEENTH	392.7	412.7	434.4
FIFTEENTH AVENUE	384.7	399.5	452.3
LOUISE A. SPENCER	402.0	409.2	412.4
DR. M. L. KING, JR	386.7	409.1	422.6
AVON AVENUE	395.0	399.7	418.2
CHESTNUT STREET	352.7	355.8	408.6

the same group. The exception is Maple Avenue, which moved from the seventh to the nineteenth place and it was replaced by Miller Street which went from the twentieth to the ninth position. At the lower end, five of last year's bottom schools remain there. Since Cleveland and Eighteenth Avenue no longer have eighth grade classes and three schools moved out of the group (Bragaw, George Washington Carver, and Dayton), five new schools entered it (Chestnut Street, Avon Avenue, Dr. Martin Luther King Jr., Thirteenth Avenue, and Vailsburg M.S.).

District mean scores for reading, math, and writing are 419.0, 431.3, and 453.5, which represent a change of -24.8, +2.7, and +6.1 scale scores in relation to last year's results. These mean scores are more than one standard deviation below the State average, which means that the District's average score in reading, math, and writing is among the lowest 16 percent of the State. Variation in school performance is large, the range of the mean scale scores is over 150 in reading and math, and over 100 in writing.

3.1.2. Cluster Results

Each of the EWT areas assessed several skills which were represented by items consolidated as clusters. In addition, there were some open-ended questions which also contributed to the total score.

Reading skills were assessed through ten clusters; four of them dealt with different kinds of reading texts: Narrative, Informational, Persuasive/

Argumentative, and Everyday; three related to the level of comprehension shown when reading the texts: Reading the Lines, Reading Between the Lines, and Reading Beyond the Lines; one focused on reading strategies: Knowledge about Reading; and two addressed student background information: Attitudes about Reading, and General Information. Only items in the first seven clusters contributed to the reading scores, items on the Knowledge about Reading, Attitudes about Reading, and General Information clusters did not.

Six clusters were used for Mathematics: Numerical Operations, Measurement and Geometry, Patterns and Relationships, Data Analysis, Pre-Algebra, and General Information. Equally, only the first five contributed to students' math scores.

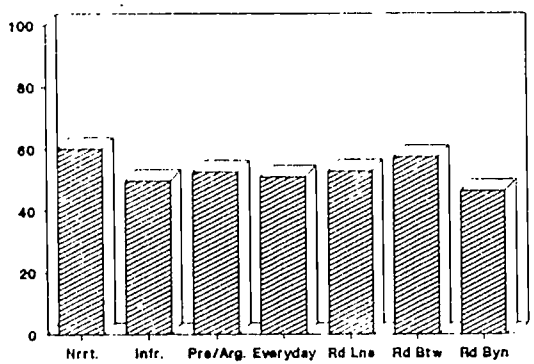
To obtain information about students' writing abilities, a combined measure of essay and multiple-choice was used. The multiple choice component (Reconstructing Meaning by Revising/Editing the Written Text of another Writer), dealt with the following seven clusters: Sentence Mechanics, Sentence Construction, Precision and Coherence, Sentence Combining, Transitional and Logical Progressions, Focus and Organization, and General Information. Again, only the first six contributed to students' writing score.

Cluster score information was offered as percent correct scores, which do not allow for a direct comparison with the overall information. Nevertheless, we have provided a general picture of the results in Figure 3.1.1. As shown in the figure, eighth graders seem to have a better understanding of the

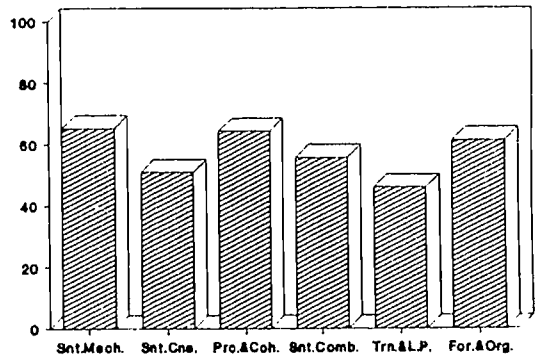
EWT AVERAGE PERFORMANCE

(Percent Correct Responses)

Reading Section



Writing Section



Mathematics Section

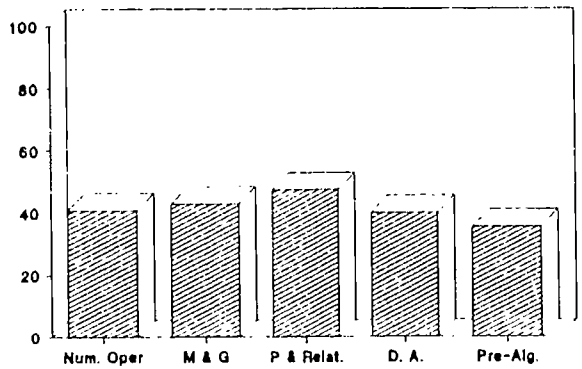


Figure 3.1.1

narrative, which is explored by questions based on the text elements that the readers internalize and come to expect of a story, i.e., characters, setting, central problem. Equally, students performed better in Reading Between the Lines (answer is implied by the text) than in Reading the Lines (answers to the questions are explicitly stated within the text.)

Mathematics results were generally poor, with little variation among the clusters. The range of percent correct responses went from 35.4 for Pre-Algebra to 47.3 for Patterns and Relationships.

Writing results portray a mean total essay equal to 6.8 on a scale from 2 to 12. In the Revising/Editing section, students performed best at identifying and correcting common mechanical and/or usage errors within a sentence (Sentence Mechanics), and at selecting words, phrases, or clauses that logically completed a partially constructed sentence in written text (Precision and Coherence).

3.1.3. Considerations on Reading Results

Comparability of performance over time on statewide tests such as the EWT is critical. In order to assess progress towards meeting state standards a consistent frame of reference is needed. This was the purpose of the development of scale scores. Under the design, each year's test is statistically equated to the previous year's test and, therefore, comparable to the 1991 base test. Thus, improvement of students' performance should be reflected in

improvement of scores.

As discussed above, district performance improved slightly in math and writing, but decreased dramatically in reading. It is highly unlikely, statistically, that a sample as large as the Newark district can drop about half of a standard deviation due solely to chance. Our eighth graders are administered the Stanford 8 a month later than the EWT. Eighth grade performance in the Stanford this year increased 1.1 NCE's in reading. A correlational study performed last year found a high degree of association between performance on the EWT and performance on the Stanford 8. This would lead to the expectation that results on the 1992 EWT would parallel results on the Stanford 8. This was not the case. Eighth grade students had a slight increase on the Stanford reading, but a dramatic decrease on the EWT reading. This indicates a problem of equivalence between the 1991 and the 1992 versions of the EWT. We strongly recommend that the State reevaluate the equating study.

3.1.4. Mastery of Benchmark Skills

In compliance with the statute which created the EWT, the New Jersey State Board of Education established score standards for that test. (See Table 3.1.2). These standards are to be applied to the scores of eighth graders who took the EWT in March 1992 and those who take it thereafter. Students whose scores fall below these standards do not meet state requirements for

mastery of benchmark skills and are candidates for supplemental remedial instruction. The final decision about assigning students to basic skills instruction is also based on other relevant information as well, in Newark's case, performance on the Stanford 8.

Table 3.1.2
Score Standards for the EWT

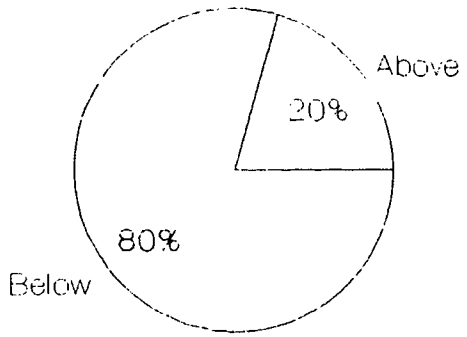
Subject	Scale Score
Reading	466
Mathematics	464
Writing	471

Figure 3.1.2 shows how these standards are going to affect our school population. Students eligible for basic skills based on the EWT results represent 79.5 percent of the population in reading, 70.2 percent in math, and 62.8 percent in writing.

EARLY WARNING TEST (EWT)

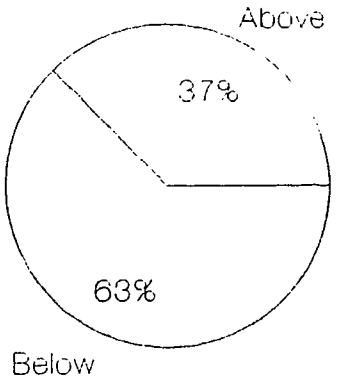
PERCENTAGE OF STUDENTS ABOVE AND BELOW STATE STANDARDS

READING



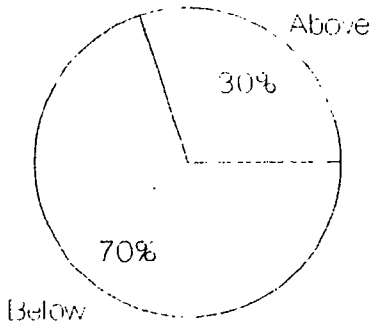
N = 3,060

WRITING



N = 3,044

MATHEMATICS



N = 3,046

Figure 3.1.2

3.2. GRADE 11 HIGH SCHOOL PROFICIENCY TEST

The New Jersey Legislature passed a law in 1988 which authorized the transfer of the High School Proficiency Test (HSPT) from the ninth to the eleventh grade. All public school students entering the ninth grade on, or after September, 1991 will have to attain a passing score on the HSPT11 as a high school graduation requirement for the State of New Jersey. A three year experimental phase of testing was instituted to allow schools to prepare for this requirement. The experimental phase known as due-notice testing spans the academic years of 1990 through 1992. The results of due-notice testing are not meant to provide individual student assessments. Results can best be used to assess strengths and weaknesses of schools and districts. Results can also be useful in determining acceptable school and district levels of performance in specific curriculum areas. This experimental phase allows schools time to examine their curriculum and determine if their objectives are in alignment with the objectives of the HSPT11. If objectives are not aligned, test results may reflect this. Students who have not experienced specific curriculum objectives in the classroom will be disadvantaged when items representing these objectives are presented to them on the HSPT11. Schools with aligned curriculum will provide their students a greater opportunity to do well on the HSPT11 than schools with curriculum objectives different from those assessed by the HSPT11.

State recommendations for aligning curriculum evolved from subject area

development committees which were established to clarify the skills and knowledge areas that should be tested. These areas are based on current research and philosophy relevant to reading, mathematics, and writing. The testing design implemented by the state for use in the due-notice phase of HSPT11 administration was matrix-sampling. This technique involves utilization of a large test with a multiple-choice format. This large test is broken down into separate forms which are sequentially distributed to students. Each student receives only one section of the whole test. This technique reduces the amount of time each student is tested while providing the school with a complete picture of how it did on each item of the entire test. The school receives feedback on all items while the student experiences a shorter test. This testing format will not be used in the high stakes version of HSPT11.

The present report reflects the second (1991) administration of the HSPT11 due-notice testing. Although the 1990 and 1991 HSPT11 test items assess the same skills, items may vary in difficulty. It is not appropriate to directly compare the 1990 and 1991 scores, as a completely new set of items was developed each year. Without items common to both tests, it is not possible to determine if the two tests were equally difficult. The results in Table 3.2.1 would be incorrectly interpreted if they were read as absolute. A statement such as: The reading test performance on HSPT11 in Newark rose from 1990 to 1991 is inaccurate. A statement such as: The HSPT11 reading test performance in Newark relative to the state rose in 1991 would be

accurate.

TABLE 3.2.1
DISTRICT MEAN SCALE SCORES
HSPT 11

YEAR	READING	MATH	WRITING
1990	145	155	170
1991	150	160	165

While the due-notice tests do not provide a measure of absolute performance, they do provide an avenue to determine how schools, and the district, have performed relative to the state as a whole.

The results of HSPT11 are reported in mean scaled scores which have been converted from raw scores. The first step in the scaling process is to compute a school mean which is then used to determine the state mean. The mean scaled score values range from 100 to 400 for each school. The state mean was set at 250 with a standard deviation of 50 points.

A standard deviation is typically interpreted in terms of the normal curve. The whole distribution of means is included in the interval between -3 and +3 standard deviations. This means that 68 percent of all schools scored between 200 and 300, while 95 percent scored between 150 and 350. A score of 100 would be among the lowest .13 percent of the distribution and a score of 400 would be among the top .13 percent of the distribution.

It should be remembered that the scaled scores are reflective of school

means not of individual scores. This is common practice in matrix-sampled tests such as the due-notice HSPT11 where the emphasis is on school level rather than student level scores.

Due to the method of testing and the method of scaling, direct comparisons between tests and schools cannot be made. For example, if school "X" obtained a reading mean score of 120 and a math mean score of 140, it is not correct to say that the performance of students in school X is better on the math test than on the reading test. The only types of comparisons that can be made are those that are indirect.

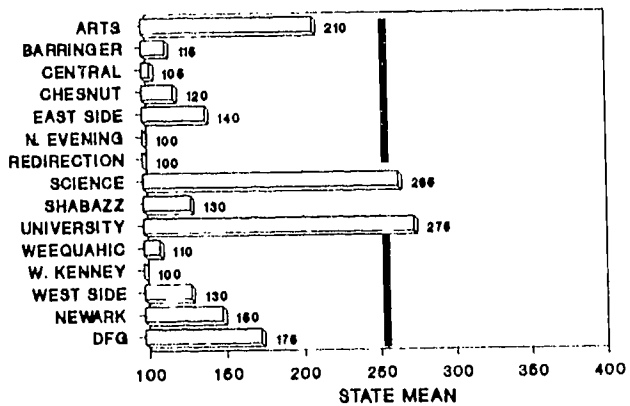
3.2.1. Overall Results

Figure 3.2.1 graphically depicts the mean scale scores for each high school, the District, and the District Factor Group (DFG). The State mean is highlighted at 250 to make comparisons easier. The state standard deviation is 50. The district mean scaled score of 150 for reading indicates that the average school score was 2 standard deviations below the state mean. This would generally indicate that in reading, the average school in the district is among the lowest 2.5 percent of the schools in the State.

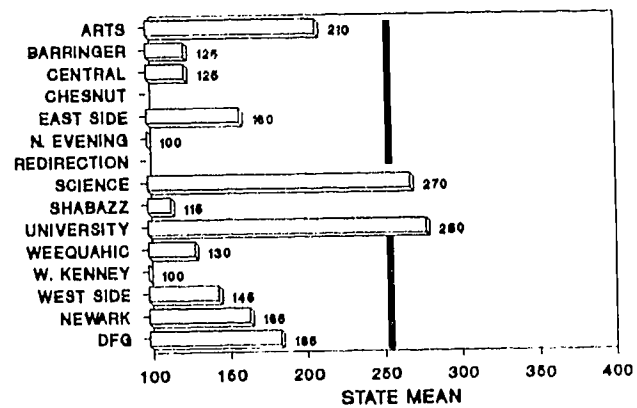
University (275) and Science (265) performed above the state mean. Arts (210) was less than 1 standard deviation below the state mean, while East Side (140), West Side (130), Shabazz (130), Chestnut (120), Barringer (115), Weequahic (110) and Central (105) performed between 2 and 3 standard

1991 HSPT 11 RESULTS

READING SECTION



WRITING SECTION



MATHEMATICS SECTION

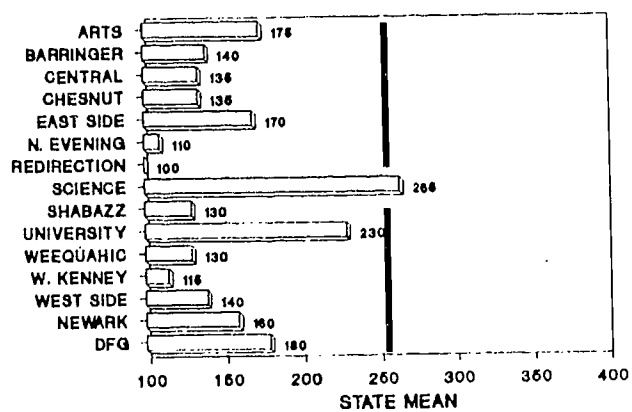


Figure 3.2.1

deviations below the state mean. Newark Evening, West Kinney and Redirection were 3 standard deviations below the state mean.

The average math score (160) was less than 2 standard deviations below the state mean. Science (265) scored above the state mean while University (230) scored less than one-half of a standard deviation below the state mean. Arts (175) and East Side (170) were less than 2 standard deviations below the state mean. Barringer (140), West Side (140), Chestnut (135), Center (135), Shabazz (130), Weequahic (130), West Kinney (115) and Newark Evening (110) ranged between 2 and 3 standard deviations below the state mean. Redirection scored 3 standard deviations below the mean.

The average writing score (165) was less than 2 standard deviations below the state mean. University (280) and Science (270) were above the state mean while Arts (210) was less than 1 standard deviation below the state mean. East Side (160) was less than 2 standard deviations below the state mean and West Side (145), Weequahic (130), Barringer (125), Central (125) and Shabazz (115) were between 2 and 3 standard deviations below the state mean. Newark Evening and West Kinney were 3 standard deviations below the state mean. (Averages were not reported for Chestnut and Redirection as fewer than five students took the test.)

Overall, Newark is between 1 and 2 standard deviations below the state mean on each of the three tests.

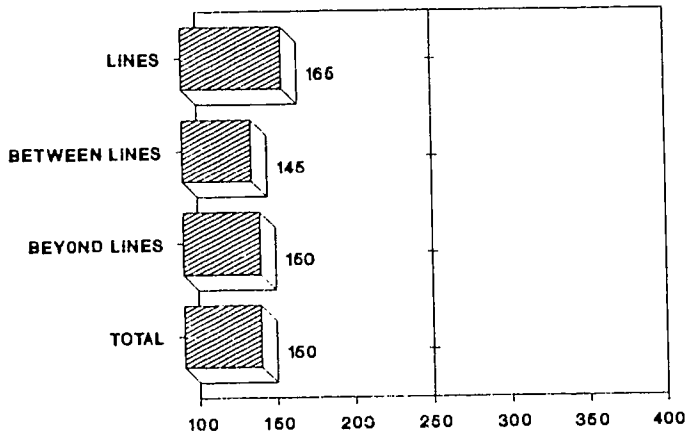
3.2.2. Cluster Results

Items within each test of the HSPT11 contain clusters which represent general skill areas. Figure 3.2.2 depicts the district's average performance for each of these clusters. In the Reading test the skill areas included three levels of Comprehension (Reading the Lines, Reading Between the Lines and Reading Beyond the Lines), Reading strategies (Knowledge About Reading) and Personal background (Attitudes Towards Reading). Figure 3.2.2 presents the average scaled scores of the three levels of comprehension. The average scores ranged between 145 in Reading Between the Lines and 165 in Reading The Lines. Reading comprehension was assessed through four types of text (Narrative, Informational, Persuasive/Argumentative and Workplace). Figure 3.2.2 also depicts the reading passage cluster performance averaged by passage type. Average scores within the passage cluster ranged from 145 in the Narrative passage to 155 in the Workplace passage. The average district score on this cluster was 150 which is 2 standard deviations below the state mean.

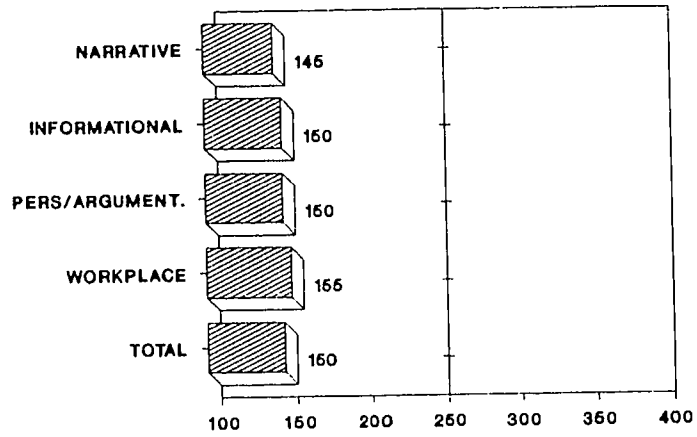
The mathematics section of the HSPT11 emphasized understanding, applications, problem solving and thinking. This holistic view of mathematics tapped into the cognitive domains of: conceptual understanding, integration of mathematical knowledge, procedural knowledge, communication and reasoning, and problem solving. These cognitive domains were sampled within five content clusters: Numerical Operations, Measurement and Geometry, Patterns and Functions, Data Analysis and Fundamentals of Algebra. The total

HSPT 11 OVERALL CLUSTER RESULTS

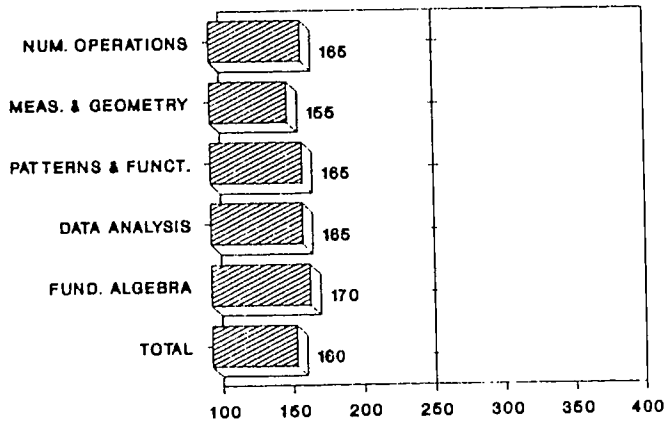
READING COMPREHENSION LEVELS



READING CLUSTERS



MATHEMATICS CLUSTERS



WRITING CLUSTERS

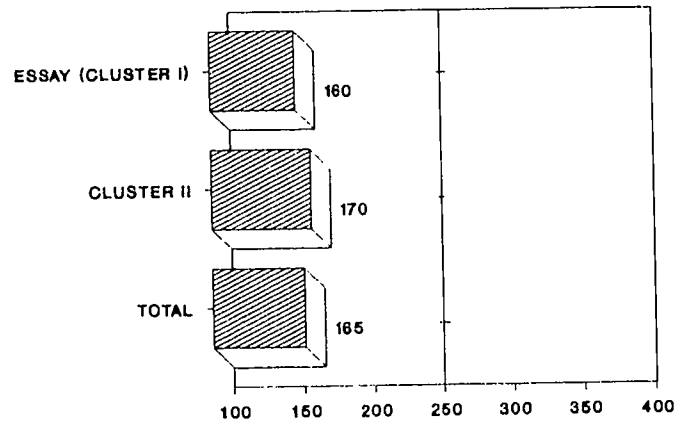


Figure 3.2.2

mathematics score is the average of these five cluster scores. As shown in Figure 3.2.2, the average cluster scores ranged from 155 in Measurement and Geometry to 170 in Fundamentals of Algebra. The overall District average in Mathematics was 160 which was slightly under two standard deviations below the state mean.

The writing section of the HSPT11 is divided into two clusters: Constructing Meaning By Writing, and Reconstructing Meaning By Revising and Editing the Written Text of Another Writer. The first cluster was measured by an essay written by the student in response to one of nine different writing tasks. These nine writing tasks were divided along two dimensions: Purpose, and situation. The three purposes for writing were: Solution to a problem, speculations about cause and effect, and a controversial issue. The three levels of situation established an audience: Peers, school, and large community. The essay score which was averaged by the dimension purpose, ranged from 155 for Controversial to 170 for Solution to a Problem. Figure 3.2.2 displays the average cluster and total writing score for the district. The overall essay average for the District was 160. The second writing cluster averaged 170 while the average District Writing Total was 165.

3.2.3. Confidence Intervals

Confidence intervals are bands around a score which indicate the range of values within which a true score can fall. These intervals are determined by

the level of confidence one wishes to have in a particular score. A typical band would contain the scaled score plus and minus 2 standard error units. This computation would yield a range of scaled scores within the 95 percent confidence interval. This means that if a test is given to a group of students an infinite number of times, 95 percent of the computed confidence intervals would contain the true score. The range of the confidence interval increases as the standard error increases. The standard error is affected by the size of the sample, so that larger samples have smaller standard errors than smaller samples.

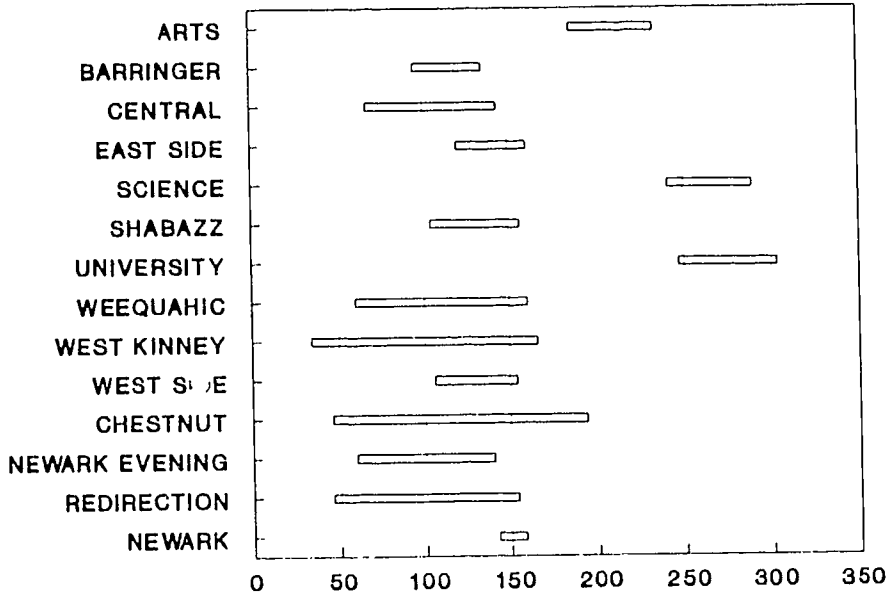
Figure 3.2.3 displays the confidence intervals for the Reading, Mathematics and Writing sections of the test. It can be noted from these figures that schools with a small number of students have a large standard error which increases the band around a score, indicating less precision in the mean estimate. Schools with a large number of students have a smaller standard error and consequently decreases the band around the score indicating a greater precision in the mean estimate. The District's confidence intervals are very small due to the large number of students participating in the tests.

3.2.4. Distribution of Individual Scores

The major purpose of due-notice testing is to give schools information on their curriculum. Once identified, strengths and weaknesses within the curriculum can be addressed. The state provides a frequency distribution of

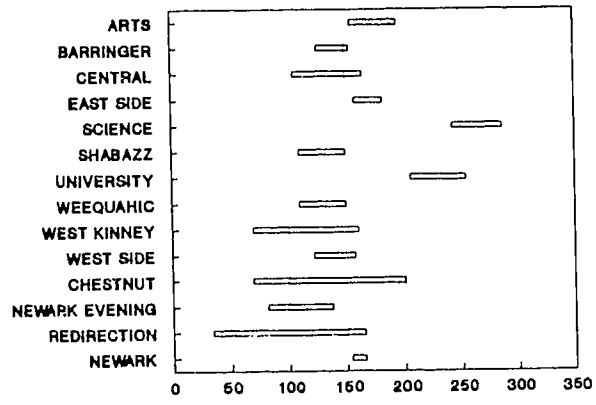
HSPT11 CONFIDENCE INTERVALS

READING SECTION



(MEAN SCALE SCORES)

MATHEMATICS SECTION



WRITING SECTION

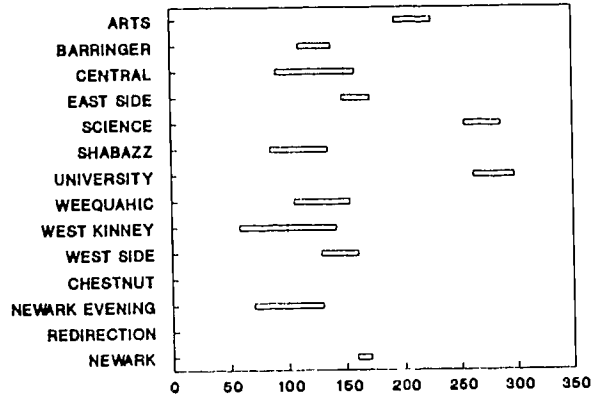


Figure 3.2.3

percentage of individual students falling within six levels of performance. Figure 3.2.4 depicts the percentage of individual students falling within these specified ranges for the District, District Factor Group and the State.

It is apparent from the charts that the district has a larger percentage of students falling within the lower ranges than the state. While 47 percent of district students scored below 151 in reading, only 19 percent fell in that group at the state level. In math the percentages were 53 for district students and 21 for state students. In language the figures were 44 percent for Newark and 18 percent for the state.

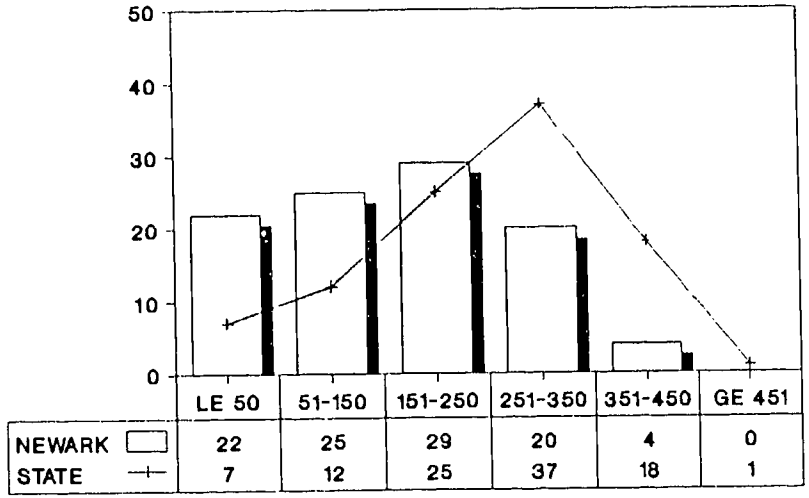
3.2.5. Attendance

According to the State's direction, the HSPT11 was to be administered to all eleventh grade students except those special education students whose individual education plan exempts them from this requirement. This year, schools were given the authority to decide about the LEP students. Special education students and students with limited English proficiency were not included in school summary reports.

Attendance at the 1991 HSPT11 has improved over the 1990 administration. Newark's eleventh grade enrollment at the time of the test was 2,062. District-wide attendance at the Reading subtest was 62 percent, approximately 2 percent higher than 1990. Attendance at the Math subtest was 60 percent, approximately equal to 1990. Attendance at the Writing

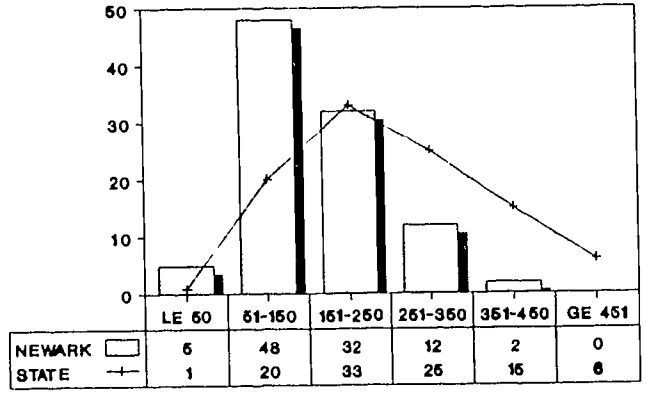
DISTRIBUTION OF HSPT11 INDIVIDUAL SCORES

READING SECTION



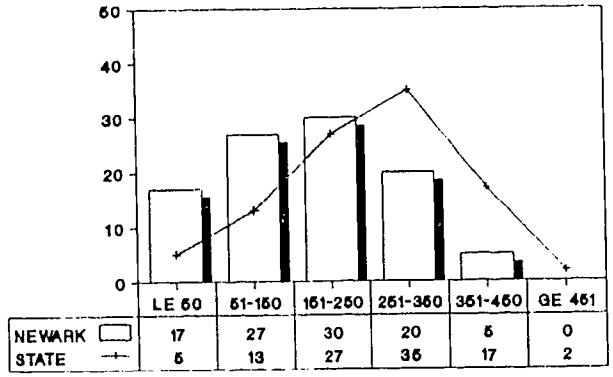
	LE 50	51-150	151-250	251-350	351-450	GE 451
NEWARK	22	25	29	20	4	0
STATE	7	12	25	37	18	1

MATHEMATICS SECTION



	LE 50	51-150	151-250	251-350	351-450	GE 451
NEWARK	5	48	32	12	2	0
STATE	1	20	33	25	15	6

WRITING SECTION



	LE 50	51-150	151-250	251-350	351-450	GE 451
NEWARK	17	27	30	20	5	0
STATE	5	13	27	35	17	2

Figure 3.2.4

subtest was 50 percent, approximately 10 percent higher than last year. It has been shown that students who would not do well on the test were more likely to be absent. This trend could cause district averages to be inflated and lead to under representation of low scoring students. It is important that all students be present at testing so that each will be represented in determining the state average.

Figure 3.2.5 depicts the attendance rates, by test, at the school level as well as the overall District attendance rate.

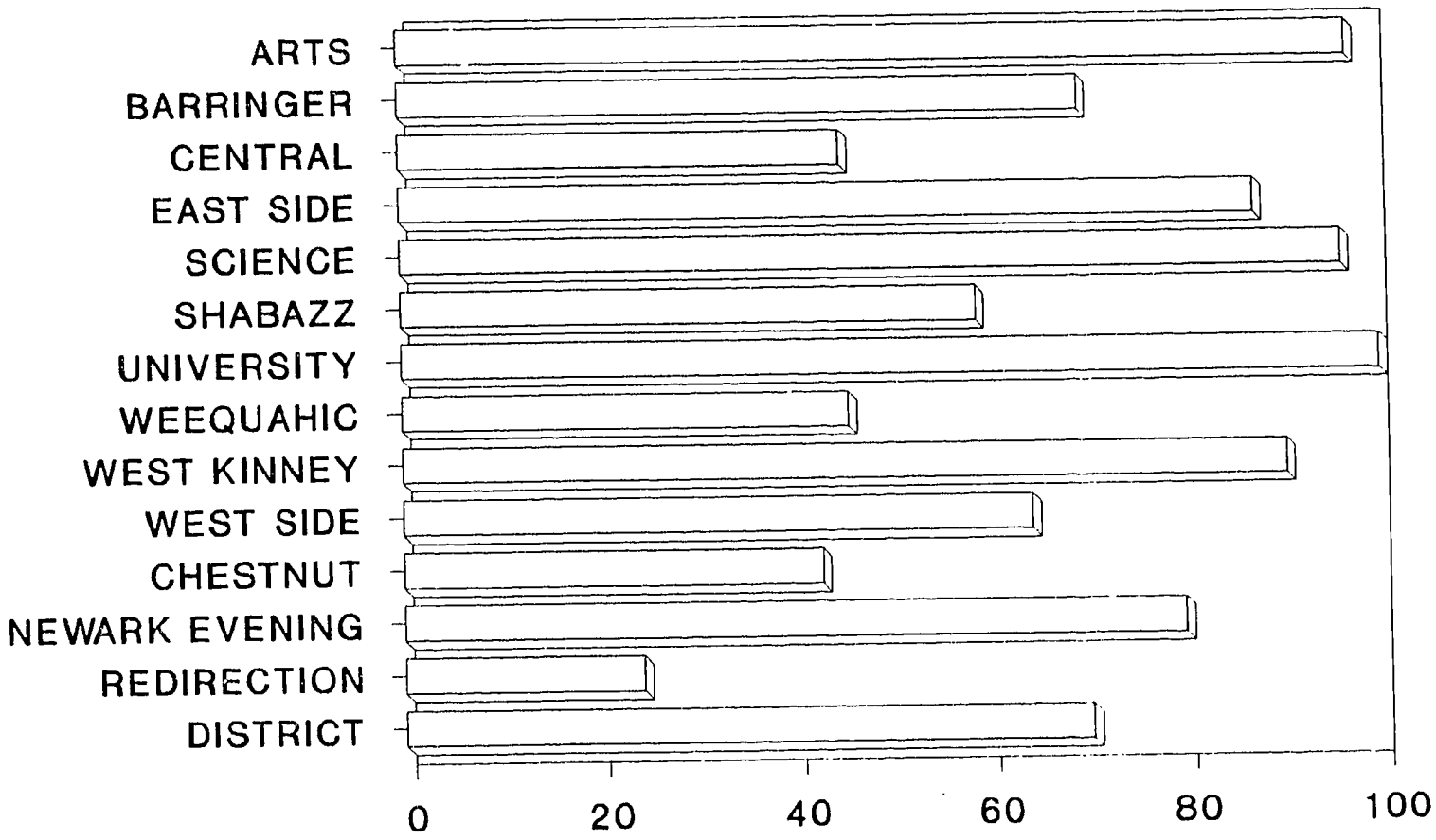
3.2.6. Reflections on HSPT 11 Results

While a statewide standard of passing has not been determined for the HSPT 11, prediction can be made based on accumulated information. A conservative estimate would assume that at least 40 percent of the population statewide will fall below any state established HSPT standard. This means that these scores will be above 160 for all three sections.

If this were the case, and this year's results were valid towards granting a high school diploma, the results would have been devastating for our district: More than half of our seniors would be at risk of not graduating. If we consider that the larger number of students who did not take the test (about 40 percent of the Newark population) would have failed it, the situation becomes even more dramatic.

Now is the time for each school to assess its performance, set

HSPT 11 ATTENDANCE RATES



(Include all regular students who took at least 1 subtest)

Figure 3.2.5

reasonable goals, and determine how to help students who will be scoring at unacceptable levels. Now is the time for teachers and administrators to realize the importance of this test. Now is the time for the District to act.

We have the information on strengths and weaknesses in curriculum. We need to use it. Tomorrow will be too late. Tomorrow, 60 percent of our students will fail to qualify for a New Jersey State diploma.

3.2.7. Use of "Due-notice" Test Information

Due-notice testing presents each school with an opportunity to gauge how well its students are meeting the future high school requirements for graduation. There is time, before the high stakes version of this test becomes operational, to address strengths and weaknesses of curriculum. These results are a report of how well each school's curriculum is aligned with the objectives of the HSPT11. Students will be at a disadvantage if the school curriculum is poorly aligned with the test objectives. School administrators and teachers should carefully inspect their test results, identify problem areas and realign curriculum to fit the objectives. Acceptable levels of passing should be established by each school so that administrators, teachers, and students can work together towards achieving these goals.

The HSPT11 is not a basic skills test. It requires the use of higher order cognitive skills such as thinking, problem solving, reasoning and decision making. These skills are additional to what is usually presented in standard

texts. Our students are expected to go beyond basic math and reading, to application and integration of academic materials into real life situations. This calls to each of us to present these opportunities to our students before they are required to take the HSPT11. We want our students to be successful.

One aspect of due-notice testing which can help schools is the option to view school results by specified groups within the school. Students in different programs may perform differently. If a school wanted to determine the effectiveness of a program they could do so by employing this option. Results would be reported for that subgroup on a separate sheet as well as in the total. Schools could also use this option to identify the "feeder" schools students graduated from. This would help distinguish elementary schools which are producing high or low achieving students allowing high schools to prepare for them. Feedback could also be provided to elementary schools on their effectiveness in relationship to this test.

(For further information on how to utilize "Due-notice" test information consult your copy of School District Guidelines: How to Interpret and Use Grade 11 HSPT Reports).

3.3. GRADE 9 HIGH SCHOOL PROFICIENCY TEST

All New Jersey public school students who entered ninth grade from September 1985 until September 1990 were required to pass the HSPT9 as one of their graduation requirements. Created to assess reading, mathematics, and writing basic skills, the test was to be administered for the first time in the Spring of the student's freshman year, and every Fall and Spring thereafter until the student successfully completed each subtest.

The HSPT9, no longer a graduation requirement, will take a few years to phase out. In the interim, test results will continue to be used to classify students in need of remediation and provide the means to identify strengths and weaknesses of each student.

Results for the HSPT9 are comparable on a year to year basis because they are reported in the form of equated scores, which represent equal levels of achievement based upon the first year's scores. This section presents the results of the two administrations of the test during the 1991-1992 academic year.

3.3.1 Student Performance

The High School Proficiency Test was administered to all eligible tenth, eleventh, and twelfth grade students in October 1991 and in April 1992. Eligible students were those who had previously failed any of the sections of the test, and those who for any reason had never taken it. Table 3.3.1 shows

the number of students who took the test and the respective passing rates in each administration. This table suggests that the highest passing rates were in reading while the lowest were in math. Overall, there was not a great deal of variability in the performance of the different grades.

TABLE 3.3.1
HIGH SCHOOL PROFICIENCY TEST
PASSING RATES 1991-1992

GRADE	STUDENTS	FALL 1991			SPRING 1992		
		READING	MATH	WRITING	READING	MATH	WRITING
10	TESTED PASSING	333 56.2%	795 23.5	741 49.9	158 49.0%	522 33.5%	356 32.0%
11	TESTED PASSING	108 56.5%	344 26.5%	222 52.7%	73 56.6%	253 28.1%	147 32.7%
12	TESTED PASSING	67 55.2%	179 27.4%	127 42.5%	24 62.5%	85 27.1%	68 32.4%

3.3.2. Proficiency rates

The District's criteria for success established the minimum percentage of sophomores, juniors, and seniors who were expected to attain the State prescribed MLPs on the HSPT: 95 percent in reading, 75 percent in math, and 85 percent in writing for tenth graders; 95 percent, 85 percent, and 90 percent in reading, math, and writing, respectively, for eleventh graders; and 95 percent in all subtests for twelfth graders. Analyses of the 1991-1992 results indicate that the district exceeded its objectives for all grades and all components of the HSPT9. Table 3.3.2 summarizes the results.

TABLE 3.3.2
HIGH SCHOOL PROFICIENCY TEST
CUMULATIVE PASSING RATES

GRADE	READING	MATH	WRITING
10	96.9	78.5	86.3
11	99.6	90.1	97.0
12	99.9	94.8	98.5

IV. THE BILINGUAL TESTING PROGRAM

The Bilingual Testing Program evaluates the English language skills of bilingual students and the academic achievement of those with limited English proficiency (LEP). The student identification process begins with the administration of a Home Language Survey, used to determine the primary language of incoming students. If this is other than English, the Language Assessment Battery (LAB) is administered to assess the students' level of English proficiency. Students who score at or above the state mandated cutoff scores are considered proficient in English and subsequent testing is administered in English. Students scoring below the cutoff are considered to be of limited English proficiency and are placed in a bilingual program. If the primary language of these students is Spanish or Portuguese, achievement tests are administered in their native language. If the primary language of the bilingual student is other than Spanish or Portuguese, achievement tests are administered in English.

For current students, time in the program, as well as score on the LAB, play a role in determining the language in which the achievement tests are administered. Current guidelines dictate that achievement tests will be administered in English under the following circumstances:

(1) To any student who scores at or above the district cutoff in the LAB.

(2) To elementary school students in the bilingual program 3 years or more.

(3) To high school students in the bilingual program one year or more.

This section of the report discusses the results of the 1992 administration of the bilingual testing program. It is arranged in four parts:

- Evaluation of English Proficiency.
- Evaluation in Spanish.
- Evaluation in Portuguese.
- Other Evaluation of LEP Students.

4.1. EVALUATION OF ENGLISH PROFICIENCY

English proficiency is assessed in the district with the Language Assessment Battery (LAB). This instrument, developed by the New York City Board of Education, has four levels covering the span from kindergarten through the 12th grade. (See Table 4.1.1.) Each level has two parallel forms, which are alternated each year so that it takes 2 years for both forms to be administered. Form A is always used in the Fall screening of new students.

English proficiency is assessed in four areas: Reading, Writing, Listening, and Speaking. Levels II, III and IV of the LAB assess these areas in separate tests. Level I combines Listening and Speaking into one

TABLE 4.1.1
LEVELS AND GRADES FOR LAB

Level	Grade
I	K-2
II	3-5
III	6-8
IV	9-12

subtest and Reading and Writing into another. Both subtests are administered to Grades 1 and 2. Students in Kindergarten take only the Listening/Speaking component of the test.

Demonstrating proficiency on the LAB is the first step in exiting the bilingual classroom. A multi-criteria review process is employed to ensure the readiness of a bilingual student to function successfully in an English-only classroom. These criteria are in keeping with the New Jersey Assembly Bill No. 1371, Bilingual Education Act supplementing P.L. 1974 (Approved: January 24, 1991). The first criterion is for LEP students to achieve a LAB score at or above the state determined cutoff (See Table 4.1.2).

TABLE 4.1.2
STATE CUTOFFS FOR LAB (Form A)

	K	1	2	3	4	5	6	7	8	9	10	11	12
Fall	26	51	53	87	100	102	111	113	118	110	115	119	123
Spring	29	54	56	94	104	105	115	117	120	115	118	121	124

Student readiness is further assessed by :

1. Class performance
2. English reading level
3. Performance in achievement tests administered in English
4. Teacher judgment

If a student does not meet these criteria he/she will continue in the bilingual program.

4.1.1. Proficiency Rates

During the Spring of 1992, the Newark district administered form "A" of the LAB to 5,182 LEP students. There was a Special Education population of 66 students which were excluded from this analysis so that the number of student scores examined was 5,116. Table 4.1.3 presents the 1991 and 1992 LAB results.

TABLE 4.1.3
LANGUAGE ASSESSMENT BATTERY
PERCENTAGE OF STUDENTS MEETING THE STATE CUTOFF

	1991		1992	
GRADE	NUMBER TESTED	PERCENT PASSING	NUMBER TESTED	PERCENT PASSING
K	832	11.2	808	11.0
1	881	11.1	911	5.6
2	694	34.5	681	29.2
3	528	23.9	489	23.5
4	427	15.0	410	11.0
5	395	19.5	309	18.1
6	315	12.1	292	15.4
7	293	16.4	267	17.2
8	257	22.6	256	18.4
9	208	6.7	191	14.7
10	197	15.7	194	5.2
11	181	12.2	165	12.7
12	162	16.7	143	20.3

The greater percentage of students passing in 1992 were in the second, third, fifth, eighth, and twelfth grades. Students in grades one and ten had the lowest passing rates. Overall, 15.3 percent of the students met the state

criteria in the LAB. Within normal variations, these results are consistent with the 1991 analyses.

4.1.2. Performance by Primary Language

The district's multicultural composition is reflected in the native languages of the student population. The number of languages spoken by district students has been growing steadily, with the vast majority of the bilingual students speaking Spanish, Portuguese, or Haitian Creole. Together, they account for almost 96 percent of this population. Table 4.1.4 illustrates the composition of the bilingual program.

TABLE 4.1.4
PRIMARY LANGUAGE OF BILINGUAL POPULATION

LANGUAGE	PERCENT IN PROGRAM
Spanish	67.4
Portuguese	23.9
Haitian Creole	4.5
Pashto	.7
Polish	.6
Urdu	.5
Vietnamese	.5
Bengali	.4
Gujarati	.3
Cantonese	.2
Others	1.0

Table 4.1.5, below, show the mean raw scores for the three larger language groups. It is evident from this table that they differ in their

Table 4.1.5
LAB RESULTS BY PRIMARY LANGUAGE

	SPANISH (N = 3,449)				PORTUGUESE (N = 1,225)				HAITIAN (N = 228)			
GRADE	NUMBER TESTED	MEAN SCORE	S.D.	PERCENT PASSING	NUMBER TESTED	MEAN SCORE	S.D.	PERCENT PASSING	NUMBER TESTED	MEAN SCORE	S.D.	PERCENT PASSING
K	605	18.64	8.13	0.00	160	20.08	8.09	10.00	16	20.94	9.53	18.80
1	671	36.85	11.41	5.50	192	41.02	12.04	3.60	19	41.95	10.59	0.00
2	500	47.38	10.88	22.50	148	51.01	11.74	48.60	14	51.71	6.57	50.00
3	365	74.08	23.15	20.60	106	77.93	25.67	33.00	7	75.57	20.37	28.60
4	299	75.16	27.72	7.70	84	80.31	27.54	22.60	13	66.23	22.83	7.70
5	221	80.73	28.42	15.80	63	89.24	21.43	20.60	13	87.00	18.20	23.10
6	176	81.30	31.81	11.90	89	86.34	31.34	21.30	14	83.79	36.76	28.60
7	173	83.62	30.53	14.50	60	93.72	27.98	23.30	22	80.77	26.48	13.60
8	163	87.94	29.15	14.70	60	100.52	31.46	33.30	17	69.95	26.54	5.90
9	87	71.75	34.94	21.80	69	79.22	28.19	7.20	14	76.79	32.08	21.40
10	80	74.35	28.11	7.50	69	92.01	22.39	5.80	30	71.53	19.92	0.00
11	62	86.40	26.96	8.10	64	92.25	30.18	21.90	23	71.83	18.99	0.00
12	47	87.23	27.89	10.60	61	113.21	16.89	34.40	26	83.65	20.12	0.00

performance. The Portuguese-speakers obtained the highest overall results in the LAB; their mean score in grades three through twelve were the highest of the three groups. In grades K through two, the Haitian-Creole speakers were highest while Spanish-speakers ranked in second and third place, depending upon the grade level.

Tables A-4 through A-7 in the Appendix present the mean raw scores for all other language groups, and Table 4.1.6, below, presents the overall passing rate by language. It should be noted when reading these results that some languages have a minimum number of students in their program so that the percentages may be misleading.

TABLE 4.1.6
PASSING RATE ON LAB BY PRIMARY LANGUAGE

PRIMARY LANGUAGE	NUMBER TESTED (*)	PERCENT PASSING
Spanish	3,499	11.25
Portuguese	1,225	19.40
Haitian Creole	228	10.53
Pashto	35	17.14
Polish	31	19.35
Urdu	28	7.00
Vietnamese	25	28.00
Bengali	21	10.00
Gujarati	13	15.40
Cantonese	10	20.00
Others	46	22.00

(*) Some students did not report on primary language.

4.1.3. Impact of Time in Program

The main goal of transitional bilingual education programs is to provide structured English language instruction as well as other subject instruction in the student's native language with the intent of facilitating English language competence. Students remain in these programs until they reach such competence. More than 75 percent of Newark's bilingual population have been in the bilingual program between 8 and 37 months. Table 4.1.7 presents the overall distribution of students by time in the program.

Table 4.1.7
Time in Bilingual Program

Months in Program	Number of Students	Percent Students
Less than 8	468	9.1
8 - 17	1769	34.6
18 - 27	1335	26.1
28 - 37	817	16.0
38 - 47	364	7.1
48 - 57	213	4.2
58 - 67	79	1.5
68 - 77	48	.9
78 - 87	8	.2
88 or More	7	.1
Total	**5108**	**99.8**

A correlation analysis which looks at the magnitude of association between two variables was computed between time in the program and score on the LAB. Correlations can range from -1 to +1 with higher positive numbers associated with a positive association and higher negative numbers associated with a reverse association. Table 4.1.8 displays the correlations between these variables. They ranged from .22 to .56, depending on grade level and were significant at the .001 level, which suggests that the association between these two

variables was unlikely to have been caused by chance factors.

TABLE 4.1.8
TIME IN THE BILINGUAL PROGRAM AND PERFORMANCE IN THE LAB
CORRELATION COEFFICIENTS

GRADE	K	1	2	3	4	5	6	7	8
CORRELATION	.22	.27	.39	.48	.56	.42	.51	.44	.46

The strength of the correlations varied depending on grade level. One of the reasons why this happens in the lower grades is that time in the program has very low variability. Eighty-six percent of the kindergartners were in their first year in the program (between 8 and 17 months). Approximately the same percentage of first graders were either in their first (25 percent) or second year (61 percent).

These findings led to a further analysis which looked at whether time in the program could predict success or failure on the LAB, and if so at which grade levels. A logistic regression was the statistical technique employed for this purpose. Scores on the LAB were dichotomized to be 0 for a score falling below the state cutoff and 1 for a score falling at or above the state cutoff. A regression analysis was used to analyze these results with time in the program used as a predictor of success. Individual grade analysis revealed that time predicted success on LAB, significantly for all grades except grade 5, grade 10, grade 11, and grade 12. This means that, except for those grades, time in the program has an impact on whether or not students reach the proficiency levels in the LAB.

4.2. EVALUATION IN SPANISH

Spanish-speaking students in kindergarten through eighth grade who scored below the district cutoff in the LAB and had been in a bilingual program less than three years were given a test in their native language: Aprenda, La Prueba de Logros en Español. High school students who scored below the LAB cutoff for the district and were in their first year in the bilingual program were also given the option of taking the Aprenda math subtest.

Aprenda assesses the same educational objectives that are measured by Stanford 8: reading, mathematics, language, and study skills. Its subtests, with the exception of mathematics computation, were developed directly in the Spanish language, which allows for a better use of the language and for a more accurate reflection of the diverse cultural backgrounds of Spanish-speakers. Aprenda includes seven levels: Preprimer, for kindergarten students; Primary 1 through 3 for students in first through third grades; Intermediate 1 for fourth graders; Intermediate 2, which the district uses for students in fifth and sixth grades; and Intermediate 3, administered to seventh and eighth graders.

One of the main objectives pursued by the District in choosing Aprenda was to obtain a more uniform assessment than it had in previous years when speakers of Spanish were tested with several different tests. An additional consideration towards the selection of Aprenda was the equating of this test to the Stanford, which provides a link between both tests and allows for comparison of their scores. Aprenda provides two sets of norms necessary for

complete score representation. One set allows the comparison of Aprenda scores to the national school population. The other is referenced for comparison among the Spanish-speaking population attending school in the United States. National scores should be used when a comparison with the English-speaking population is desirable, or when reference to State-imposed criteria of performance is required. Reference scores allow for a comparison of this population of students with a more analogous group, both in language and in socio-economic background.

4.2.1. General Information

A total of 1,637 students took Aprenda this Spring. This increase of 385 students over last year reflects the inclusion of kindergarten students. Overall, the number of students taking the test decreased between 2 and 18 percent in all grades except eight. The number of schools administering this test increased from 35 to 45, with Franklin testing 212 students and Roberto Clemente, Dr. W. Horton, and Elliot testing over one hundred students each.

Analysis of time in the bilingual program showed that about 18 percent of the students in the sample had less than 8 months in the program, 47 percent had between 8 and 17 months, and 24 percent had between 18 and 27 months. Information on time in the program was not provided for the rest of the students. A large proportion of the students taking this test (874 or 53.4 percent) were receiving Chapter I services, the majority of them in

Spanish. The supplemental ESL program serviced 396 of these students.

4.2.2. Student Performance

As shown in Table 4.2.1, mean scores for 1992 reflect little variation in relation to the 1991 figures. Computation of a weighted mean for grades one through eight yielded results similar to last year: The average score for reading and language was 47.3 NCEs, and for math it was 46.7 NCEs. Third graders had the highest performance across areas, and eighth graders had the lowest with math being their weakest point. Students in the second and sixth grades increased their averages in all three subtests. Students in the fourth grade saw a decline in all three of them. Students in grades one, three, five, seven, and eight improved some of their scores and decreased others.

Similar results occurred in the subtest 'Study Skills', which measures a broad range of library and research skills. The average score in this subtest increased for students in grades six through eight, while the reverse occurred for students in grades three through five. The mean national NCE scores for these students, in grades three through eight were 48.4, 44.9, 41.1, 49.8, 53.9, and 52.5. This indicates that most students are performing very close to grade level, which would be 50.

Table 4.2.2. presents the 1992 results in terms of reference scores (referred to the Spanish-speaking population). It is clear from the table that reference scores are higher than national scores especially in the higher grades.

TABLE 4.2.1
APRENDA
MEAN NATIONAL NCE SCORES 1991-1992

GRADE	NUMBER TESTED (*)		READING (**)		MATHEMATICS		LANGUAGE	
	1991	1992	1991	1992	1991	1992	1991	1992
K	-	414	-	44.8	-	44.3	-	-
1	454	446	43.4	43.4	50.1	48.6	47.5	47.1
2	147	125	46.1	48.7	46.7	52.4	44.0	45.1
3	134	120	47.4	49.3	49.7	51.2	50.1	48.8
4	122	113	56.2	54.6	51.1	46.6	49.1	46.1
5	101	82	57.7	57.2	46.5	43.3	44.3	45.0
6	93	85	51.3	54.1	42.9	45.5	42.8	46.7
7	116	96	46.0	45.5	40.6	42.1	48.7	51.2
8	85	90	43.1	44.5	39.6	33.4	46.1	48.8

(*) The number of students may vary slightly for each subtest.

(**) Kindergartners took only the Sounds and Letters subtest, therefore they do not have a total reading score.

TABLE 4.2.2
APRENDA
1992 MEAN REFERENCE NCE SCORES FOR REGULAR STUDENTS

GRADE	READING		MATHEMATICS		LANGUAGE		STUDY SKILLS	
(*)	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
K	49.8	(23.4)	52.2	(24.4)	-	-	-	-
1	47.1	(21.8)	52.1	(22.4)	51.0	(21.8)	-	-
2	51.4	(24.6)	56.2	(23.9)	49.8	(21.7)	-	-
3	52.2	(20.1)	56.3	(18.8)	52.4	(19.1)	51.8	(18.4)
4	60.7	(17.2)	57.4	(19.2)	51.9	(19.3)	50.7	(19.3)
5	66.9	(17.5)	59.1	(17.1)	59.2	(15.6)	54.2	(14.8)
6	65.4	(25.0)	61.5	(17.5)	62.2	(16.4)	60.7	(17.7)
7	58.7	(21.4)	56.8	(18.2)	63.3	(18.9)	64.6	(20.1)
8	57.6	(18.7)	50.1	(16.5)	61.2	(17.1)	63.7	(20.3)

(*) Twelfth graders are not shown because only a small, non-representative sample of the population took the test.

Except for reading at the first grade, all other means are at or above grade level. This difference is explained because of differences in the norming populations.

Analyses of variance revealed that the mean differences between grades were significant for reading and math but not for language. Study skill scores also were significantly different between some of the grades.

4.2.2.1. Kindergartners

For the first time Kindergarten students were administered the Preprimer level of Aprenda. This test has a higher difficulty than SESAT 1, the test given to the English-and Portuguese-speaking kindergartners. It is recommended for grades K.5 to 1.5, while SESAT 1 is recommended for grades K.0 to K.5. Due to the increased difficulty of this test only two subtests were administered: Mathematics, and Sounds and Letters.

Performance of kindergartners was in line with other grade means. In Sounds and Letters, the 418 students who took the section obtained a mean NCE score of 44.8, with a standard deviation of 22.8. In Mathematics, where only 375 students had a valid score, the mean NCE was 44.3, and the standard deviation was 24.3. The English population mean of SESAT was slightly higher in both areas which may be an artifact of the more difficult test.

4.2.2.2. High School Students

Sixty-five high school students took Aprenda this year: 41 ninth graders at Barringer Prep, and 13 sophomores, 8 juniors, and 3 seniors at Barringer High. They were administered the math sections of level Intermediate 3, which was standardized on a population of students in grades 6.8 through 8.8. This makes the norms unusable on a high school population. Therefore, results from their testing are presented as percentage of correct responses. The ninth graders showed certain weaknesses, with a 57.1 percent of correct responses. Higher grades showed stronger results, with 68, 71, and 75 percent correct responses for 10th, 11th, and 12th graders, respectively. In the same manner, the best subtest for all grades was Math Applications, which yielded 65, 74, 79, and 87 percent of correct responses for students in grades 9 through 12.

4.2.3. Impact of Time in the Program

Analyses of variance were performed to determine whether time in the program affected achievement in Aprenda. Three groups of student were considered, those with less than 8 months, those between 8 and 17 months, and those between 18 and 27 months. Although the overall statistic was significant for the three subtests, the differences in scores among any two groups were not significant. This indicated that time in the bilingual program is not a factor contributing to scores in this test.

4.3. EVALUATION IN PORTUGUESE

Portuguese speakers who scored below the cutoff score in the LAB and had been in a bilingual program for less than three years were administered the Portuguese Achievement Test (PAT) in grades one through eight. Kindergarten students were given the PAT in reading and the SESAT 1 in mathematics. High school students who scored below the LAB cutoff and were in the bilingual program for a year or less had the option to take the PAT math subtest.

The Portuguese Achievement Test (PAT) is an in-house translation of the CTBS U/V, which has been in use in the District for the past three years. It measures reading, language and mathematics, and is organized in eight levels, to cover the entire grade range. Test results are reported as raw scores.

4.3.1. General Information

Four hundred and sixty students took the 1992 PAT, 49 percent were females and 51 percent were males. Information on instructional setting was available for 365 students. The majority (77.5 percent) were receiving services in a self-contained classroom, while the rest were in a pull-out situation. Some students obtained Chapter I services during the year: 45 of them received basic skills instruction in their native language, 30 in all three areas, 5 in reading and math, and 10 in one of the three. There were 114 students in the group who were receiving Supplemental ESL.

It is important to note that the trend of decreasing numbers of students

being tested in grades one through eight, which was noted last year, is continuing. This Spring, 328 students took the PAT at these grade levels. This is a reduction of 15 percent in relation to 1991. Two reasons can be offered as an explanation for the decreasing numbers; one is that the Portuguese population has diminished throughout the district in the past few years; the second is that as Portuguese-speakers tend to perform well in the LAB, they take their achievement tests in English, thus reducing their numbers in the PAT.

4.3.2. Proficiency Rates

After a number of revisions, OPET developed permanent cutoff scores in 1990. These were used again this year. Their main purpose is to establish eligibility for the Communication/Computation Native Language Component of the Basic Skills Program. Table 4.3.1 presents the outcome for students in grades one through eight, in 1991 and in 1992.

Table 4.3.1 shows that, for both years, the majority of students performed above the cutoff scores. This year's results indicate improvement in all subtests in grade 8; and in the reading and math subtests in grades 1, 5, 6, and 7. Students in 1992 performed lower in all tests in grades 2 and 3; in language in grades 1, 5, 6, and 7; and in reading in grade 4. We cannot yet extrapolate from these results. We hope to be able to validate them against other criteria such as teacher judgment and correlations with performance in other tests.

TABLE 4.3.1
PORTUGUESE ACHIEVEMENT TEST
PERCENTAGE OF STUDENTS EXCEEDING CUTOFF SCORE

GR	NUMBERS		READING		MATH		LANGUAGE	
	1991	1992	1991	1992	1991	1992	1991	1992
1	99	102	84%	92%	83%	96%	86%	83%
2	39	28	64%	46%	87%	82%	97%	89%
3	44	31	46%	32%	82%	68%	80%	61%
4	39	39	72%	54%	95%	95%	82%	82%
5	46	31	87%	90%	85%	90%	91%	90%
6	45	46	78%	89%	82%	91%	82%	80%
7	40	27	93%	96%	85%	93%	88%	85%
8	35	24	83%	100%	83%	96%	83%	88%

4.3.3. Kindergarten Assessment

Students in kindergarten were assessed with two instruments. The PAT subtests Visual Recognition and Sound Recognition were used in the area of Reading, while the SESAT 1 (with the directions for administering translated to Portuguese) was selected to assess math.

One hundred and three kindergartners were tested this year. This number reflects a reduction of 27 percent over last year in the number of students tested. A mean raw score of 14/18 in Visual Recognition and of 11.3/17 in Sound Recognition was achieved. Mean raw score in math was 21.6/42.

4.3.4 Impact of time in the program

Students taking the PAT test were classified according to time in the bilingual program. The largest number of students (149 or 46.1 percent) had been in the program between 8 and 17 months, 88 students (27.2 percent) had been serviced between 18 and 27 months, and 86 students (26.7%) had been in the program for less than 8 months. Fifteen students did not have data available.

Results were independent of time in the program except for the lower grades. In kindergarten, students who had been in the program between 8 and 17 months performed significantly better in reading than those who had been in the program less than 8 months (their respective mean scores were 27.4 and 23.6). In math, results were reversed. Newly arrived students performed significantly better that those of longer stay (27.0 and 18.7, respectively). At the first and second grades, the only statistically significant difference was in the language subtest. Here, students with less than 8 months in the program obtained better results than those with 8 months or more (the average score was 20.1 for the first group and 14.4 for the other). Time in the program did not make a difference in the performance of the students in other grades.

4.4. OTHER EVALUATION OF LEP STUDENTS

All LEP students with more than three years in a bilingual program, as well as those whose native language is other than Spanish or Portuguese, or those who scored at or above the district cutoff in the LAB were administered the Stanford 8 (in English), as were all the regular students in the District.

The Stanford 8 test was administered to 3,018 students identified as having limited English proficiency. The group included kindergartners through twelfth graders and the numbers reflect the decline in the bilingual population this year. The test was given in 45 schools district-wide. Schools were unevenly represented in the sample; some had very few students and others, such as Franklin, Elliott, Ann, and Mount Vernon had a substantial amount (about 200). The largest bilingual population was in East Side, where 386 LEP students were assessed.

4.4.1. General Information

The group of LEP students assessed with the Stanford 8 came from divergent backgrounds. As previously mentioned, there were three paths leading to this test: time in the program, language, and LAB score, creating a heterogeneous testing group. Table 4.4.1 presents the composition of this group by language. There were 445 students for whom information on language was missing and almost 94 percent of the remaining group spoke one of the three main languages.

Table 4.4.1
PRIMARY LANGUAGE OF LEP STUDENTS TAKING STANFORD

LANGUAGE	PERCENT TESTED
Spanish	63.1
Portuguese	27.4
Haitian Creole	3.2
Pashto	1.4
Vietnamese	0.5
Others	4.4

As shown on Table 4.4.2, variability of time in program was also a factor It ranged from less than 8 months to more than 88 months. The greatest percentage of students (72.5 percent) have been in the program between 8 and 37 months.

Table 4.4.2
TIME IN THE PROGRAM OF LEP STUDENTS TAKING STANFORD

TIME (MONTHS)	PERCENT STUDENTS
Less than 8	4.3
8 - 17	23.0
18 - 27	23.3
28 - 37	26.2
38 - 47	11.6
48 - 57	6.7
58 - 67	2.7
68 - 77	1.7
78 - 88	0.3
More than 88	0.2

About 60 percent of the students in this group were receiving Chapter

1 services, the majority of them (703) in all three areas of reading, math, and writing. The number receiving supplemental ESL was 387 students.

4.4.2. Student Performance

Students scored below grade level at all grades and in all areas. Math performance, however, was relatively better with a mean NCE of 42.4. Language came in second place, for a mean NCE of 31.8, and reading was last with a mean NCE of 27.5. The encouraging note is that, overall, students scored better than last year, for a combined increase of 6.4 NCE's, 1.8 in math, 1.9 in language, and 2.7 in reading. Results by area and grade are presented in Table 4.4.3. Comparison with regular students is discussed in Section 2.2.2.

Kindergarten students are not reflected on Table 4.4.3 because they did not take the whole test. In Sounds and Letters (the only Reading subtest the 307 LEP kindergartners were administered), they achieved an average score of 43.3 with a standard deviation of 24.4. In math, their mean NCE was 44.2 and their standard deviation was 24.6. According to these results, the typical LEP kindergarten in the district is performing below grade level. The large standard deviation points to the fact that even within the group of kindergartners there is a large degree of variability.

Table 4.4.3
STANFORD 8
1992 MEAN NCE TOTAL SCORES FOR LEP STUDENTS

GRADE	N TESTED(*)	READING Mean	READING S.D.	MATHEMATICS Mean	MATHEMATICS S.D.	LANGUAGE Mean	LANGUAGE S.D.
1	355	40.3	(17.4)	48.2	(20.3)	42.5	(23.0)
2	533	31.4	(16.4)	45.0	(19.3)	33.2	(17.6)
3	338	31.2	(15.2)	44.2	(17.6)	36.3	(14.8)
4	261	23.5	(13.2)	42.9	(15.2)	33.0	(12.0)
5	206	21.6	(11.7)	40.4	(15.6)	28.3	(13.5)
6	157	25.2	(12.2)	39.1	(16.6)	30.6	(13.7)
7	144	21.6	(12.8)	40.5	(18.0)	27.8	(13.9)
8	138	23.4	(15.6)	36.4	(19.3)	27.0	(13.7)
9	155	18.4	(10.8)	37.1	(16.5)	22.6	(11.9)
10	178	16.6	(11.1)	37.9	(15.2)	22.0	(12.8)
11	187	27.5	(14.9)	41.2	(14.7)	29.6	(15.6)
12(**)	59	18.0	(9.3)	33.8	(11.6)	22.8	(12.7)
TOTAL	2,801	27.5	(16.1)	42.4	(18.6)	31.8	(16.9)

(*) Since some of the students did not take all portions of the test, sample size (N) for each of the subtests may vary.

(**) The only seniors required to take the test were those in Chapter 1.

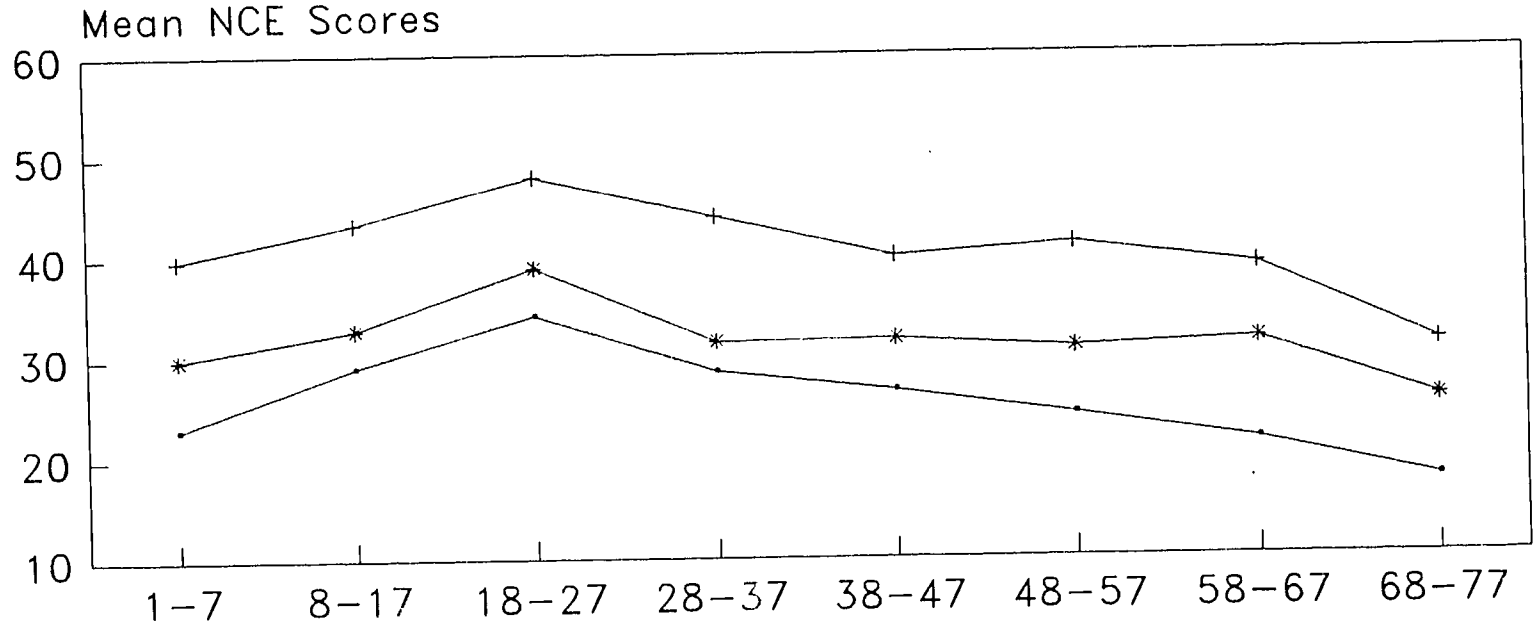
4.4.3. Impact of Other Variables

Statistical analyses were performed to determine the impact of two independent variables (time in the program and grade) on students' performance in reading, math, and language. The analyses indicate a significant effect of the two independent variables in all three areas. An interaction effect was also found in two of the areas indicating that time in program had a differential effect across grade. In other words, the effect of time in the program on student performance in reading and math was not the same across grades. No interaction effect was found for language, which means that the impact of time in program for this variable was homogeneous across grades.

Performance on this test is not directly related to time in the bilingual program. Trend analyses of subject by time in program for each grade did not show a linear trend in any of the subject areas. As depicted in Figure 4.4.1. performance increases in the first and second year and decreases thereafter. Students who had been in the program between 18 and 27 months (2 years) had the best performance in all three areas. The difference between their mean scores and that of the majority of the other groups was large enough to be statistically significant.

School was also a significant variable. Students were grouped according to the school they attended, thus obtaining three groups for the high schools and 9 for the elementary (at this level only schools with more than 100

OVERALL STANFORD RESULTS LEP STUDENTS



	1–7	8–17	18–27	28–37	38–47	48–57	58–67	68–77
Reading	23	29.1	34.2	28.6	26.6	24.2	21.5	17.6
Math	39.7	43.3	47.9	43.9	39.9	41.1	38.9	31.1
Language	29.9	32.8	38.9	31.5	31.7	30.8	31.5	25.6

Time in Program (Months)

Reading Math Language

Figure 4.4.1

students were considered). At the high school level, East Side students ranked first and Central students ranked last in all three areas. At the elementary school level, best performers in the three subtests came from Oliver Street and Ann Street.

Language was also a significant variable affecting performance. There were six groups in total. Vietnamese-speakers scored the highest, but their numbers were so small that these results were not statistically significant. Portuguese-speakers came second in reading and language, whereas speakers of other languages came second in math. Spanish-speakers came consistently next to last, followed only by Haitian Creole-speakers who had the poorest performance in all subtests. To put these results in proper perspective, we must remember that speakers of languages other than Spanish or Portuguese must take the achievement test in English test whether or not they are proficient in the language.

APPENDIX

TABLE A-1
READING STANFORD 8
COMPARISON OF 1991 AND 1992 MEAN NCE TOTAL SCORE

Grade	Number Tested	Reading 1991 Mean	S.D.	Reading 1992 Mean	S.D.	NCE Change
K	2,834	-	-	49.3	(23.7)	-
1	3,555	41.4	(19.2)	42.9	(19.3)	+1.53
2	3,295	40.4	(16.9)	43.4	(18.4)	+2.96
3	3,397	41.3	(17.2)	43.3	(17.9)	+1.96
4	3,295	35.8	(17.4)	36.5	(18.5)	+0.73
5	3,352	34.7	(15.4)	36.9	(15.9)	+2.19
6	3,342	37.9	(16.3)	38.3	(16.2)	+0.44
7	3,118	39.8	(16.5)	38.8	(16.1)	- 1.02
8	2,943	38.5	(16.5)	39.6	(16.5)	+1.14
9	2,300	35.6	(16.1)	36.6	(16.3)	+1.01
10	1,635	34.7	(16.5)	35.0	(16.9)	+0.30
11	1,275	37.0	(16.5)	37.3	(17.1)	+0.29
Total	31,719	38.1	(17.1)	39.5	(17.5)	+1.36

TABLE A-2
MATHEMATICS STANFORD 8
COMPARISON OF 1991 AND 1992 MEAN NCE TOTAL SCORE

Grade	Number Tested	Math 1991 Mean	Math 1991 S.D.	Math 1992 Mean	Math 1992 S.D.	NCE Change
K	2,833	-	-	49.9	(23.8)	-
1	3,555	47.2	(22.7)	48.6	(22.6)	+1.43
2	3,315	48.1	(20.6)	51.6	(21.7)	+3.49
3	3,410	47.6	(19.8)	49.6	(19.8)	+1.97
4	3,269	45.7	(18.6)	46.9	(19.0)	+1.20
5	3,348	42.9	(18.1)	44.2	(18.8)	+1.27
6	3,339	45.1	(18.1)	45.6	(18.5)	+0.49
7	3,109	45.8	(15.9)	43.8	(16.7)	- 2.01
8	2,933	42.7	(17.8)	43.4	(18.0)	+0.71
9	2,353	45.9	(17.4)	41.7	(17.0)	- 4.20
10	1,701	44.1	(16.6)	42.4	(16.8)	- 1.70
11	1,329	44.7	(16.9)	44.0	(16.9)	- 0.70
Total	34,623	45.5	(18.9)	46.4	(19.7)	+0.87

TABLE A-3
LANGUAGE STANFORD 8
COMPARISON OF 1991 AND 1992 MEAN NCE TOTAL SCORE

Grade	Number Tested	Language 1991 Mean	S.D.	Language 1992 Mean	S.D.	NCE Change
1	3,555	42.9	(22.6)	46.0	(23.1)	+3.13
2	3,312	40.9	(18.3)	43.7	(19.8)	+2.80
3	3,347	43.8	(18.0)	44.6	(17.9)	+0.83
4	3,197	41.8	(16.4)	43.3	(16.8)	+1.45
5	3,308	39.4	(16.7)	41.4	(16.9)	+2.00
6	3,290	41.0	(15.5)	41.2	(15.4)	+0.18
7	3,101	40.0	(15.0)	40.2	(15.6)	+0.18
8	2,940	40.6	(16.2)	41.3	(16.8)	+0.74
9	2,409	39.3	(16.4)	38.9	(16.0)	- 0.38
10	1,723	36.4	(16.0)	35.6	(16.7)	- 0.84
11	1,354	39.0	(16.2)	39.4	(16.4)	+0.42
Total	31,702	40.7	(17.4)	42.0	(17.8)	+1.25

Table A-4
LAB RESULTS BY PRIMARY LANGUAGE

GRADE	PASHTO (N=35)				POLISH (N=31)			
	Number Tested	Mean Score	Stand. Deviat.	Percent Passing	Number Tested	Mean Score	Stand. Deviat.	Percent Passing
K	3	6.93	6.93	0.00	2	22.50	3.54	0.00
1	7	13.69	13.69	28.60	3	49.67	6.11	33.30
2	5	22.74	22.74	0.00	-	-	-	-
3	5	3.05	3.05	40.00	2	87.00	28.28	50.00
4	3	18.72	18.72	0.00	-	-	-	-
5	3	40.80	40.80	33.30	1	96.00	0.00	0.00
6	1	0.00	0.00	0.00	2	55.50	17.68	0.00
7	1	0.00	0.00	100.00	3	104.67	30.04	66.70
8	-	-	-	-	8	85.00	66.47	50.00
9	3	11.85	11.85	0.00	7	77.14	24.13	14.30
10	-	-	-	-	4	85.75	10.46	0.00
11	3	13.32	13.32	0.00	2	80.50	27.58	0.00
12	1	0.00	0.00	0.00	3	128.33	4.16	100.00

Table A-5
LAB RESULTS BY PRIMARY LANGUAGE

GRADE	VIETNAMESE (N=28)				URDU (N=25)			
	Number Tested	Mean Score	Stand. Deviat.	Percent Passing	Number Tested	Mean Score	Stand. Deviat.	Percent Passing
K	1	16.00	0.00	0.00	6	12.00	9.76	16.70
1	2	28.54	10.61	0.00	5	42.40	15.08	40.00
2	1	54.00	0.00	0.00	2	58.00	0.00	100.00
3	1	77.00	0.00	0.00	-	-	-	-
4	1	88.00	0.00	0.00	2	43.00	0.00	0.00
5	2	100.00	12.73	50.00	3	104.00	6.56	66.70
6	-	-	-	-	3	72.00	31.32	0.00
7	1	63.00	0.00	0.00	1	34.00	0.00	0.00
8	-	-	-	-	3	44.00	7.00	0.00
9	2	86.53	19.09	0.00	-	-	-	-
10	9	64.11	23.59	0.00	-	-	-	-
11	5	86.60	20.77	20.00	-	-	-	-
12	3	84.00	14.00	0.00	-	-	-	-

Table A-6
LAB RESULTS BY PRIMARY LANGUAGE

GRADE	BENGALI (N = 21)				GUJARATI (N = 13)			
	Number Tested	Mean Score	Stand. Deviat.	Percent Passing	Number Tested	Mean Score	Stand. Deviat.	Percent Passing
K	4	11.00	12.00	0.00	-	-	-	-
1	1	29.00	0.00	0.00	2	57.00	1.41	100.00
2	1	1.00	0.00	0.00	1	43.00	0.00	0.00
3	1	21.00	0.00	0.00	-	-	-	-
4	5	86.20	31.97	40.00	-	-	-	-
5	-	-	-	-	-	-	-	-
6	2	28.50	2.12	0.00	-	-	-	-
7	1	35.00	0.00	0.00	1	32.00	0.00	0.00
8	-	-	-	-	3	91.00	30.27	0.00
9	4	64.75	15.82	0.00	2	42.00	0.00	0.00
10	1	55.00	0.00	0.00	1	49.00	0.00	0.00
11	-	-	-	-	2	81.00	4.24	0.00
12	1	85.00	0.00	0.00	1	112.00	0.00	0.00

Table A-7
LAB RESULTS BY PRIMARY LANGUAGE

GRADE	CANTONESE (N = 10)				OTHER LANGUAGES (N = 46)			
	Number Tested	Mean Score	Stand. Deviat.	Percent Passing	Number Tested	Mean Score	Stand. Deviat.	Percent Passing
K	1	16.00	0.00	0.00	10	18.80	11.83	30.00
1	-	-	-	-	9	33.44	16.06	0.00
2	1	58.00	0.00	100.00	8	10.26	10.26	50.00
3	-	-	-	-	2	33.23	33.23	0.00
4	-	-	-	-	3	28.57	28.57	0.00
5	-	-	-	-	3	26.58	26.58	33.30
6	1	66.00	0.00	0.00	4	33.68	33.68	25.00
7	1	29.00	0.00	0.00	2	24.04	24.04	50.00
8	3	114.67	7.37	33.30	5	30.12	30.12	0.00
9	1	51.00	0.00	0.00	-	-	-	-
10	-	-	-	-	-	-	-	-
11	2	81.5	36.06	0.00	-	-	-	-
12	0	-	-	-	-	-	-	-