

**ACADEMIC ACHIEVEMENT IN TWO EXPEDITIONARY LEARNING/
OUTWARD BOUND DEMONSTRATION SCHOOLS**

Submitted by:

**Polly Ulichny, Ed.D.
Brown University
Department of Education
Barus Hall
Providence, Rhode Island 02912**

Submitted to:

**Expeditionary Learning/Outward Bound USA
122 Mount Auburn Street
Cambridge, MA 02138**

September, 2000

TABLE OF CONTENTS

1. Executive Summary	1
• Methodology	1
• King Middle School Summary of Results	2
2. Introduction	7
3. Methodology	10
4. King Middle School, Portland Maine	12
• Background	12
• The King Middle School Today	16
• What the Tests Show	19
• MEA Data	21
• Revised-MEA Data	36
• Writing Assessments	41
• Conclusions	48
5. Rafael Hernandez School, Boston, Massachusetts	51
• Background	51
• ELOB Instructional Program	53
• Data Sources	55
• What the Tests Show	56
• Time 0 (1993-1994) MAT Scores	58
• Time 0 DRP Scores	61
• Time 1 (1994-1995) MAT Scores	62
• Time 2 (1995-1996) Stanford 9 Scores	67
• Time 3 (1997-1999) Stanford 9 Scores	74
• Median Percentile Stanford 9 Scores	74
• Stanford 9 Reading Performance Levels	76
• Stanford 9 Math Results	
• Open-Ended Stanford 9 Results	84
• Open-Ended Math Scores	87
• Summary of Standardized MAT and Stanford 9 Test Data	89

- MCAS Results 91
 - MCAS ELA 4th Grade 93
 - MCAS Math 4th Grade 97
 - MCAS ELA 8th Grade 100
 - MCAS Math 8th Grade 102
- Summary of MCAS Results 104
- Conclusions 105

LIST OF FIGURES AND TABLES

1. King Middle School Chart of Structural Changes	13
2. Old MEA and Revised MEA	20
3. Table of King Middle School MEA Cumulative Average Scores	22
4. King Middle School 3-Year Cumulative Average Scores 1992, 1996, 1998	23
5. Percent LEP and Bilingual at King Middle School	25
6. LEP Bilingual Enrollment Compared 1996-97	26
7. LEP & Bilingual/English Fluent Enrollment Compared 1997-98	26
8. Students in Title I Reading	27
9. SES Indicators 1997-98	31
10. SES Indicators (Raw Scores) 1995-1998	32
11. Table of Average Cumulative MEA Scores Compared 1996- 1998	33
12. Change in MEA Cumulative Scores 1996-1998	35
13. Percentage of King Middle School LEP and Handicapped Included in Test-Taking Population	36
14. Revised-MEA Average Scaled Scores Compared 1998-99	37
15. Percentages of Students Scoring at <u>Exceeds Standards</u> Level 1998-99	38
16. Percentages of Students Scoring at <u>Meets Standards</u> Level 1998-99	38
17. Percentages of Students Scoring at <u>Partially Meets Standards</u> 1998-99	39
18. Percentages of Students Scoring at <u>Does Not Meet Standards</u> 1998-99	39
19. Table of King Middle School Revised-MEA Performance Levels for LEP and Monolingual English Students	40
20. Writing Assessment Data Compared 1995-1999	42
21. Writing Sample and Portfolio Writing Scores Compared	45
22. Rafael Hernandez School Demographic Data	52
23. Metropolitan Achievement Scores - Reading	58
24. Metropolitan Achievement Scores - Math	59
25. Grades 1, 2, 3 MAT Cumulative Scores 1993-94	59
26. Grades 4 & 5 MAT Cumulative Scores 1993-94	60
27. Middle School MAT Cumulative Scores 1993-94	61
28. 5 th Grade DRP Results 1992-93	62

29. 8 th Grade DRP Results 1992-93	62
30. 1994-95 MAT Scores	63
31. Grades 1, 2, 3 MAT Scores 1994-95	64
32. Grades 4 & 5 MAT Scores 1994-95	64
33. Middle School MAT Scores 1994-95, Hernandez and District	65
34. 5 th Grade DRP Results 1994-95	66
35. 8 th Grade DRP Results 1994-95	66
36. Reading 1995-96 Stanford 9 Performance Levels	68
37. Tables of 1995-96 Stanford 9 Performance Levels	71
38. Math 1995-96 Stanford 9 Performance Levels	72
39. Stanford 9 Reading Cumulative Median Percentiles 1996-99	75
40. Stanford 9 Math Cumulative Median Percentiles 1996-99	75
41. Grade 4 Stanford 9 Reading Changes 1997 to 1998	78
42. Grades 5, 6, 7 Stanford 9 Reading Change 1996 to 1999	80
43. Grade 8 Stanford 9 Reading Changes 1997 to 1999	80
44. Grades 3, 5, 6 Stanford 9 Math Changes in Performance Levels 1996 to 1999	82
45. Grades 5, 6, 7 Stanford 9 Math Changes in Level 1 1996 to 1999	83
46. Grades 5, 6, 7 Stanford 9 Percentages of Ls 3&4 Math 1996 to 1999	84
47. Reading Open-Ended Stanford 9 Grade 2 1997-98-99	86
48. Reading Open-Ended Stanford 9 Grade 6 1998 and 1999	87
49. Math Open-Ended Stanford 9 Grade 5 1998 - 1999	88
50. Math Open-Ended Stanford 9 Grade 7 1998 - 1999	89
51. MCAS ELA Grade 4 Performance Levels 1998 - 1999	94
52. MCAS ELA Grade 4 Pass & Fail 1998 - 1999	95
53. MCAS ELA Grade 4 Change in Performance 1998 - 1999	96
54. MCAS ELA Grade 4 Average Scores 1998 - 1999	96
55. MCAS Math Grade 4 Performance levels 1998 - 1999	98
56. MCAS Math Grade 4 Change 1998 to 1999	99
57. MCAS Math Grade 4 Average Scores 1998 - 1999	99
58. MCAS ELA Grade 8 Performance Levels 1998 and 1999	101
59. MCAS ELA Grade 8 Changes 1998 to 1999	102

60. MCAS ELA Grade 8 Average Scores 1998 - 1999	102
61. MCAS Math Grade 8 Performance Levels 1998 - 1999	103
62. MCAS Math Grade 8 Changes 1998 to 1999	104
63. MCAS Math Grade 8 Average Scores 1998 - 1999	104

EXECUTIVE SUMMARY

This report was commissioned by Expeditionary Learning Outward Bound (ELOB) to examine primarily the academic achievement of two ELOB demonstration schools that have been engaged in implementing ELOB since the 1993-94 academic year, the King Middle School in Portland, Maine and the Rafael Hernandez School in Boston, Massachusetts. In the initial years of ELOB, the demonstration schools received considerable support in implementing the design principles and practices that define ELOB and by 1995-96 the schools were considered to be implementing the design well. ELOB evaluations of design implementation are expressed on a three-point scale indicating the extent to which the design principles and core practices appear in the school (1=Beginning; 2=Implementing; 3=Realizing). Recent evaluations of the King Middle School and the Rafael Hernandez place them within the 2--2.5 range, indicating that they are continuously implementing the design.

To better understand the results of the quantitative analyses of various test results over the time period of ELOB implementation, we also include some information about the context of the school prior to ELOB implementation as well as current conditions in the school.

METHODOLOGY

The main question this report seeks to answer is: What is the record of academic achievement, as revealed by standardized tests and District-wide assessments, over the years of ELOB implementation in each of the schools chosen for this study? While we can not prove a cause and effect relationship, a second question that we attempt to answer is: What has been the impact of ELOB on the schools' academic performance? The answer to this second question is inferred from the achievement data and interpreted within the context of demographic data as well as qualitative data gathered on site.

The quantitative data used came from secondary sources, meaning that the data we used were analyzed and reported from individual student scores by each of the Districts we examine and their respective State Departments of Education. For the King Middle School we relied on State and District reports of the annual MEA (Maine Educational Assessment)

and the District reports of the Middle School Writing Assessment, a performance evaluation conducted by the District involving a selected sample of all middle school grade students. For the Rafael Hernandez school we mainly relied on yearly District reports and analyses prepared by the Boston Public Schools Office of Research, Assessment and Evaluation. We also used the State and District reports of the recently implemented MCAS (Massachusetts Comprehensive Assessment System).

Qualitative data were gathered on site during data collection visits and from documentation supplied by each of the schools. In the reports that follow for each of the schools selected for analysis we will describe, in detail, the nature of the test data we use.

KING MIDDLE SCHOOL SUMMARY OF RESULTS

Comparing the performance of 8th grade students on the same type of standardized test over the years prior to ELOB implementation to recent years of active ELOB implementation, we see a steady increase in all discipline areas in student attainment levels at the King Middle School. These advances in academic achievement levels coincided with a population shift at the school. Over the period we are considering here, the population of LEP students coming from low income, modestly educated parents from all over the world increased by 20%. Both LEP and low income, low parent-education populations of students systematically score lower on standardized achievement tests than middle class, native English speaking, white students.¹ Therefore, the achievement levels *per se* and the gains in achievement over time of the King Middle School students over the period of ELOB implementation would indicate that ELOB has had a very positive impact on the academic achievement of this school.

Specifically, we found that:

Comparing the changes of King Middle School students' cumulative average scaled test scores with the two other District Middle Schools as well as the State averages over time, we see that the King consistently improved its scores vis-à-vis its

¹ For a full, recent treatment of the inequities that surround standardized testing, see Linda McNeil, "Contradictions of School Reform: Educational Costs of Standardized Testing. (Routledge, 2000).

comparison groups in individual disciplines. In terms of gains over time, in four out of seven subject areas, the King demonstrated the highest gains in scores of any of the District middle schools. In two out of seven disciplines, its gains were as high as the other top gaining middle school in the District. In other words, in only one discipline area, Writing, did the King students not improve more than the other District middle schools.

In the final year of MEA testing (1998-99), which instituted a new test format and process, the King students scored better than the State average in all of the disciplines except Health, in which it scored the same as the State average. Instead of scoring below the two other District middle schools as it had on previous MEAs, the King scored higher than either comparison school in one discipline--Writing--and never scored lower than both of the comparison schools in any discipline area (in several areas it scored the same as the other lower-scoring middle school). In all disciplines, however, the differences amongst the scores of Portland's middle schools was minimal.

Student writing in all three grades of the King Middle School has also improved over time as assessed by a District-wide Writing Assessment process. Students generally performed better on Portfolio assessments, which allowed for revision of final products than on on-demand writing assessments. The gains of the King Middle School students over time on the District Writing Assessment fell between the two other middle schools in Portland. More King students, in all three grades, demonstrated gains in Writing Assessments in recent years than in earlier years of the Writing Assessment.

RAFAEL HERNANDEZ SCHOOL SUMMARY OF RESULTS

As with the King Middle School, if we consider the population of the Hernandez as having more bilingual Hispanic students--statistically the lowest performing group on standardized tests--than the District or the State, the fact that the Hernandez generally performs as well as or outperforms the District in elementary grades on both Math and Reading/ELA, means that the Hernandez shows a strong academic achievement profile in elementary grades. This achievement diminishes in the middle school years, but the trend can be explained by several factors. First, the numbers of students taking the test in the 7th and 8th grades, the grades that show the lowest performance, are very small for the Hernandez, making any fluctuation in percentages or average score seem greater than it actually is. Second, the lower numbers are a result of the exodus of high academic performers to the District exam schools or to other non-public alternatives. This leaves a less prepared middle school population on the whole, but even at that we see general improvement from the 7th to the 8th grade.

There is considerable variability in Stanford 9 scores from grade to grade in the later years of ELOB implementation which is difficult to explain, but MCAS scores for these same years indicate that the students at the Hernandez, at least in Grades 4 and 8, are maintaining a high achievement profile, in keeping with that demonstrated in the early years of active ELOB implementation.

Finally, when we look at the Hernandez compared to the Garrison, a school sharing similar contextual and demographic characteristics, we see that the Hernandez generally significantly outperforms the Garrison in both Reading/ELA and Math, at all grade levels. The fact that the Hernandez implements ELOB principles and practices, while the Garrison doesn't, allows us to infer that at least one of the reasons for the Hernandez' greater achievement is due to the implementation of ELOB as an instructional program.

Specifically, we found that:

Pre ELOB implementation, the Hernandez students scored better than the District average in the early elementary grades but scored less well in upper elementary and middle school grades than the District in Reading and Math.

In the first years of active ELOB implementation--1993-1995--the Hernandez students scored higher than the District in more grades than prior to ELOB implementation. The improvement was seen in both Reading and Math scores, with more elementary grades outperforming the District.

In 1996, the first benchmark year of a different standardized assessment adopted by the Boston Public Schools (BPS)--(the Stanford 9 replaced the Metropolitan Achievement Tests)--the Hernandez continued to score as well or better than the District in the elementary grades in both Reading and Math, but fell slightly below District comparisons starting in Grade 6 and increased the gap in Grades 7 and 8.

Starting in 1996 we also compare the Hernandez to a matched school in the District that shares similar characteristics (both are pre K--8 two-way Spanish bilingual schools) but do not share in ELOB implementation. The Hernandez significantly outperformed the Garrison² school in Reading and Math in all grades tested, both elementary and middle school grades. In subsequent years the performance gap between the two schools lessened, but the Hernandez, on the whole, outperformed the Garrison in both Reading and Math.

In later years of ELOB implementation (1997-1999) the Stanford 9 tests were administered inconsistently by grades and by format (multiple choice or open-ended) since the assessment procedure for the District was undergoing sweeping changes from 1996 to 1999 (the last available test score year). In these later years of ELOB implementation, the Hernandez did not maintain its comparative profile. Compared to the District it nearly always scored lower and showed less overall improvement. Compared to the Garrison it was more similar in scores, meaning that the Garrison improved overall at a greater rate than the Hernandez. Score differences change, however, from grade to grade and from measure to measure, increasing the

² Pseudonyms are used for comparison schools but not for the ELOB demonstration schools.

variability we see in the way the Hernandez performed on versions of the Stanford 9 in the later years of ELOB implementation, 1997-1999.

The final measures we compare are the English Language Arts (ELA) and Math portions of the MCAS from 1998 and 1999, later years of ELOB implementation. These tests only apply to 4th and 8th graders. The messy picture presented by Stanford 9 results was not reproduced in MCAS scores. Over the same period of time, the Hernandez outperformed both the District and the Garrison in Grades 4 and 8 in both ELA and Math, overall. It showed the greatest gains in preventing failure, i.e., low percentages of students scoring at Level 1, and in the overall average scores for ELA and Math at both grade levels. The improvement between 1998 and 1999 was not as great as the comparison groups in BPS on the whole. It consistently performed lower than the State averages on the MCAS, but more often higher than the average District score.

INTRODUCTION

Expeditionary Learning Outward Bound (ELOB) is a comprehensive school reform design that was originally funded by a grant from the New American Schools Development Corporation, a non-profit organization in Washington, DC that identifies and supports school reform designs that have the potential to break the mold of traditional public schooling. 1992-93 was the design year that developed the principles and practices that would then be introduced into schools that competed for grants to become ELOB demonstration sites. Several schools across the country were identified as demonstration sites in 1993 and, with extensive financial, technical and professional support from ELOB, these schools helped refine the practices that identify ELOB as a comprehensive school reform design. Since 1996, ELOB implementation has continued at many of the original demonstration sites, as well as in new sites across the county, but without the extensive financial and technical support from the parent organization. Schools have continued to develop the design in their own contexts, with some professional development support from ELOB.

ELOB is characterized by Learning Expeditions--long term, multi-disciplinary, open-ended inquiries that require collaborative efforts among students and involve academic, service and physical components. Teams of teachers, supported by extensive professional development during the school year and in summer institutes, design their own expeditions, track student performance through a variety of measures (normally involving rubrics, portfolio assessments and exhibitions of learning), and revise and elaborate them over time. ELOB schools are also expected to modify traditional organizational structures in order to base the curriculum in Learning Expeditions. These adjustments include eliminating tracking, regular planning of blocks of time for expeditionary learning, and professional development time for expeditionary team teachers to meet during the school day as well as in specific ELOB Institutes.

Ten design principles express the philosophy of ELOB and were informed by the work of Kurt Hahn, the founder of Outward Bound, Paul Ylvisaker, former professor and Dean of the Harvard Graduate School of Education, and Eleanor Duckworth, professor at

the Harvard Graduate School of Education. The ten design principles that teachers use to plan and develop expeditions are the following:³

1. *The Primacy of Self-Discovery* holds that learning happens best in situations that offer adventure and the unexpected, that require people to accept responsibility and discover their abilities, passions and values.
2. *The Having of Wonderful Ideas* captures the notion that teaching should build on students' curiosity, allow time for thinking, experimentation and making sense of what is observed. Respect for everyone's ideas should be fostered.
3. *The Responsibility for Learning* principle holds that learning is both discovery and a social activity. Learning is therefore about increasingly taking responsibility for one's personal as well as the collective's learning.
4. *Intimacy and Caring* asserts that learning happens best in situations where there is trust, mutual respect and sustained caring among all members of the learning community.
5. *Success and Failure* are believed to both be critical for students. They need to feel successful in order to build confidence and to take risks, but they also need to experience failure to learn to deal with adversity and to overcome negative inclinations.
6. *Collaboration and Competition* join personal and group development. The values of friendship and trust support group development and students are encouraged to compete, not against each other, but against their personal best and high standards of excellence.
7. *Diversity and Inclusivity* are desirable elements of any group. The greater the diversity, the richer and more creative are the problem solving possibilities and learning opportunities of the group.
8. *The Natural World* is at the core of our continued existence as a species. Respect and appreciation of it are necessary for the human spirit. It also reveals important lessons of recurrent cycles of cause and effect.
9. *Solitude and Reflection* and silence replenish our energies and open our minds.

³ Synthesized from Doug Kilmister et al., *Are We There Yet? Benchmarks and Tools for Schools Implementing Expeditionary Learning Outward Bound* (Expeditionary Learning Outward Bound, 122 Mt. Auburn Street, Cambridge, MA 02138, 1999).

10. *Service and Compassion* are important values to cultivate as members of the human community.

In addition to the ten design principles, there are five core practices supporting ELOB. A list of these practices and benchmarks for them follows.⁴

1. Learning Expeditions
 - Planning and teaching learning expeditions
 - Developing character
 - Expecting and attaining high achievement
 - Fostering strong literacy
 - Using community resources
2. Reflection and Critique
 - Examining and assessing student work
 - Examining and assessing teacher work
 - Using portfolio assessment
3. School Culture
 - Incorporating the design principles
 - Fostering service
 - Experiencing Outward Bound
 - Ensuring equity
 - Fostering family participation
4. School Structures
 - Building shared leadership
 - Instituting block scheduling
 - Using teams effectively
 - Providing and using common planning time
 - Fostering multi-year relationships
5. School Review
 - Assessing school progress
 - Collecting and analyzing evidence

⁴ Ibid.

- Creating an action plan

Over the seven years of implementation of ELOB in more than 60 schools across the country, the implementation of principles and practices has helped ELOB further develop as a school reform design. The self-discovery and self-directed nature of ELOB implementation means that each school develops differently. Since each school is embedded in its own unique history and context which influence the beliefs and practices of its staff and students, schools implement ELOB differently and along a developmental continuum.

This report was commissioned by ELOB to examine primarily the academic achievement of two ELOB demonstration schools in Northeastern U.S. that have been engaged in implementing ELOB since the 1993-94 academic year, the King Middle School in Portland, Maine and the Rafael Hernandez School in Boston, Massachusetts. To better understand the results of the quantitative analyses of various test results over the time period of ELOB implementation, we also include some information about the context of the school prior to ELOB implementation as well as current conditions in the school.

METHODOLOGY

The main question this report seeks to answer is: What is the record of academic achievement, as revealed by standardized tests and District-wide assessments, over the years of ELOB implementation in each of the schools chosen for this study? While we can not prove a cause and effect relationship, a second question that we attempt to answer is: What has been the impact of ELOB on the schools' academic performance? The answer to this second question is inferred from the achievement data and interpreted within the context of demographic data as well as qualitative data gathered on site.

The data used for this report come primarily from secondary sources. For the King Middle School we have used the Maine State Department of Education's yearly reports on standardized tests (Maine Educational Assessment--MEA), and Portland School District reports of District-wide Writing Assessments results. The history and context of the school were established by means of documents provided by the King Middle School which included outside evaluation reports commissioned by external partners (University of Maine

Partnership, ELOB), school documents--letters, reports, memos, workshop agendas, etc.--and by parameters of socio-economic status provided by the State Department of Education.

Primary source data were collected during two visits to the school in Spring 2000. Informal conversations with teachers, administrators and students presenting exhibitions of their Learning Expeditions, provide much of the data we use to establish the extent of current ELOB implementation.

For the Rafael Hernandez School, a pre K--8 two-way Spanish bilingual school in Boston, we have used the Office of Research, Assessment and Evaluation of the Boston Public Schools published data on standardized assessment for the District. A variety of tests have been used over the seven-year period we are considering here and will be described in detail below. In addition to the District reports, we also use the State Department of Education's reports on the yearly MCAS exam--Massachusetts Comprehensive Assessment System--which has been administered in grades 4, 8 and 10 from 1998 onward. Other data on test performance, demographics and school organization come from the Hernandez School's reports to the District, to ELOB, and to a non-profit organization that provides funding for the school to engage in systemic reform--the Boston Plan for Excellence in the Boston Public Schools. In the 1997-98 school year, the Hernandez undertook an extensive self-study by District mandate and presented their findings to a panel of evaluators. This study, the In-Depth Review (IDR), was relied on for information about scheduling, curriculum organization, professional development, and school-wide reform initiatives.

Qualitative data were gathered during one day of data gathering and informal interviews with administrators in the school. These conversations, along with the IDR, helped form the current context of ELOB implementation at the Hernandez.

KING MIDDLE SCHOOL, PORTLAND MAINE

BACKGROUND

The King Middle School In Portland Maine, under the leadership of Principal Mike McCarthy, had been involved in several years of strategic planning before the introduction of ELOB to the school. Beginning in the 1988-89 school year, McCarthy and the staff were involved in a complex and comprehensive process that resulted in an evolving School Improvement Plan (SIP). The next four years brought significant changes to the school that aligned well with the introduction of ELOB in the 1993-94 school year.

Briefly, the process that the school undertook to create their SIP involved the staff in a process of study and discussion of the Effective Schools research and at the same time asked them to clarify their attitudes, beliefs about the present and future hopes for the school. Comprehensive quantitative data were collected and analyzed by the staff to identify critical issues that would be addressed to improve the school. Additionally, the entire staff was trained in problem solving to enable them to work collectively to identify improvement strategies and to implement them.

Four Action Teams were formed that worked on the following areas defined as critical to improving student achievement at King:

1. Flexible scheduling
2. Parent involvement
3. Learning effective middle school instructional practices
4. Alternative programming and strategies

The next few years saw the following changes, which were proposed and voted on by staff, implemented in the school:

- Schedules were teacher determined; a vertical House structure was implemented which placed students and faculty into smaller within school units that were stable for the three years of middle school; schedules and rules were developed by the individual Houses to respond to their needs.
- Tracking was eliminated; special education pullouts were significantly reduced and an inclusion model was adopted for all but a small portion of the students.
- Instructional strategies to support middle school students were introduced and supported through ongoing professional development in the areas of cooperative learning, multiple intelligences and thematic curriculum design.

- Parent involvement increased as a result of multiple activities designed to learn about and invite parents into the school; Community agencies and university programs affiliated with the school as partners.

The following chart compares structural changes from 1988 to 1992, the year prior to implementing ELOB at the King.

CHAPTER I

1988-89

1992-93

24% Pull-out

0% Pull-out

(Results were an increase of 6 NCE on the California Achievement Tests in 1991.)

SPECIAL EDUCATION

1988-89

1992-93

256 P.E.T.s⁵

141 P.E.T.s

26% Pull-out

4% Pull-out

104 students identified as special needs
4 leaving at end of year

76 students identified as special needs
18 leaving at end of year
Learning Center created for both special
education and regular education students

ABILITY GROUPING

1988-89

1992-93

7 Ability Groups

2 Ability Groups

Gifted
Accelerated
Middle
Low
Resource
Functional Academics
Intensive Education

Regular
Self Contained (6 students)

HOUSE STRUCTURE

1988-89

1992-93

⁵ Pupil Evaluation Team, a team convened to assess the special needs of a particular student.

<p>House structure included:</p> <ul style="list-style-type: none"> English Math Science Social Studies <p>Excluded:</p> <ul style="list-style-type: none"> Computer Art Health Tech. Ed. Music Reading Foreign Language Phys. Ed. 	<p>Included:</p> <ul style="list-style-type: none"> English Math Science Social Studies Related Arts (Art, Music, Home Ed. Health, Computer, Tech. Ed.) Special Ed. Reading <p>Excluded:</p> <ul style="list-style-type: none"> Phys. Ed. Foreign Language
--	---

COMMUNITY

1988-89

1992-93

1% Parent participation
0 Community Partnerships

27% Parent participation
3 Community Partnerships

From reports written just prior to the King becoming an official demonstration site of ELOB through the New American Schools grant in 1993, the King Middle School was well on its way to reinventing the middle school experience for most of the students and teachers. Mike McCarthy is generally credited with spearheading this process which was accomplished by bringing in resources, information and experts to guide the process and requiring teachers to work together to devise and implement the SIP. Teachers, parents, and administrators who were interviewed by outside consultants and evaluators⁶ gave generally favorable descriptions of the school improvement process. Teachers, in particular, were, for the most part, pleased with the democratic nature of the process, although some veterans were skeptical that they really would be in charge of change as evident in this comment, "I think it's great that change is coming from within, but still some teachers are skeptical about whether they have power for real." Another teacher stated, "I enjoyed it because you could participate without fear of having your opinions degraded or put down." One teacher compared the current process with former initiatives when she said, "I've always felt that in-

service workshops were nothing. This is the first time that I've actually seen something happen."

A second positive attitude emerged around the learnings that staff engaged in by being taken through a reflective process that asked them to reflect on themselves as teachers as well as on the school as a place for learning. This reflective process generated a richer understanding of the school as an institution and of one's functioning within it.

As with all large scale efforts to change, there were also frustrations. Most notable among the comments were concerns about the time that it takes to actually change and improve student learning. This was mentioned by faculty and administration. Mike McCarthy indicated in an interview with an external evaluator that, "as principal, he fluctuates between outrage at what is wrong and how much needs to change quickly, on the one hand, and knowledge that collaboration and some consensus is necessary to create lasting change, on the other. His question always is 'How fast and how far can we/I go?'"

Informal conversations with McCarthy and several teachers during my visits in May and June, 2000 indicated that over the course of ten years there have been deliberate adjustments in the staffing at the King. Once the school reached consensus on its direction, there was no going back or standing still. McCarthy said, during one of my visits, that "teachers were told what they were expected to do, and all of the training and resources they needed to make modifications were supplied to them through professional development opportunities made available at the school. If they didn't want to get on board, they transferred to another school, were not rehired, or were removed from a teaming situation." The complexity of "steering" a democratic process of school reform is captured in the following quote from an article co-authored by McCarthy and Carolyn Chaplin, another Portland principal:

For the reflective principal concerned with improving a school, each day is filled with being at moving points on a continuum of conflicting polarities: outrage vs. collaboration, nurturing vs. demanding quality, maintaining the present vs. delivering the future, negotiating consensus vs. implementing on the basis of data, and enjoying being a principal vs. constantly worrying that it's all not enough. There are conflicting expectations coming from all levels as members of our learning

⁶ Sources for these data come from Patrick J. McQuillan, "An Assessment of the School Improvement Process at King Middle School: Some Issues to Consider," August, 1990, and Donna E. Muncey, "Demonstration School Project Report for King Middle School," August, 1993.

communities (parents, students, teachers, school committee, central office, unions, business partners) embrace change and restructuring efforts at various degrees and levels of understanding. There is the sense that we are in the midst of making history, but there is the nagging question of, "Are we making any progress towards improving the essence of the enterprise: that of teaching and learning?"⁷

As we will see from the quantitative data presented in this report, there is evidence of progress in student achievement over time.

THE KING MIDDLE SCHOOL TODAY

Shireen and Katie are narrating the slide show of the 6th grade Outward Bound experience of last fall to visitors to their exhibition in the library. They tell us what challenges they undertook with the help of their friends. When we ask them about Learning Expeditions as a way of learning Shireen says she's done them since the 1st grade and finds learning what you decide you want to learn is a lot more interesting than doing assigned homework. Katie is less sure. She is new to public schools, having spent her elementary years in a Catholic school. She thinks expeditions have pros and cons. "They teach you about diversity and to respect differences in people that you wouldn't normally talk to, but I'm not sure I'm learning all that I should be learning [if the teacher doesn't tell me what to read and what assignment to do]."

This comment mirrors the initial attitude of parents towards Expeditionary Learning according to McCarthy. At first the parents of successful middle school students were unenthusiastic about heterogeneous grouping and student-directed inquiries. They thought their children would be brought to a lower level. Within a year of Expeditionary Learning, however, they came to see the value of this instructional environment. They are now more of Charlotte's opinion (another 6th grader), "We learn things that really stick with you. It's a great school."

When asked how the whole school managed to buy into the reform design, McCarthy said it was a combination of things. First of all, the whole school is headed in one major direction--that of ELOB and what that encompasses in terms of project-based learning, teaming, heterogeneous grouping and portfolio assessments. The staff are not

⁷ From Carolyn Chaplin & Michael McCarthy, "Winter of Despair, Spring of Hope" in *The Journal of Maine*

pulled in several different directions produced by competing agendas for reform. Anything that comes into the school fits under the umbrella of ELOB.

Secondly, what the staff are asked to do, they are given adequate time and professional development to accomplish. The school has common planning time each day for teams and a five day paid Summer Institute for planning and revising expeditions as well as ten professional days during the school year. There is a staff member, a former teacher in the school, assigned to facilitating the work of teams in planning and carrying out expeditions. Since she has been successful in her own classroom using expeditions, she has credibility among the staff and can help them plan adequately and avoid pitfalls based on her own expertise. She also helps them gather materials for the expeditions they plan. Professional development is organized around the ELOB focus of the school. University partners have helped the staff recognize that they needed to better understand what they were asking students to do as well as understand whether or not any given student had the necessary materials and skills to carry out the work. Since nearly 70% of the students qualify for the federally subsidized lunch program and over 20% of the students are second language speakers of English, teachers could not take for granted that there would be materials in the home required to complete homework assignments. Carefully constructing product descriptors that identified what students were to do and how they might accomplish it was a major step in addressing the heterogeneous needs of mixed groups of students.

Thirdly, according to several staff members, the ELOB design is respectful of teachers. It honors their work and recognizes that professional time is required to do it. McCarthy supports that view by his philosophy of providing all of his staff with what, according to him, constitutes the three essentials for a professional: time (common planning time and professional development days); money (the discretionary budget is given to teams to purchase materials for their expeditions); and a telephone (there is one in every classroom).

Finally, the staff saw the results of the first EL team and wanted to get the same results with their students. The first EL team had 50% fewer documented discipline problems; students were engaged in their work and their attendance improved; the quality of student work improved and they were proud of their products; the team of teachers worked

harder at planning but were less taxed during instruction time with students because they were not "in charge of a classroom of reluctant learners."

On the day of student exhibitions, both the library and the cafeteria were filled with projects, displays, video presentations, and a scaled model of a Medieval village. Students were on hand each class period to explain to visitors what the project entailed and what they had learned. They were articulate, enthusiastic and knowledgeable enough about their particular display to answer a wide variety of questions from the parents, community members and the press who attended the exhibitions. There was very little teacher involvement visible in the work of the day except to move students in and out of the cafeteria and library at the change of classes.

The video project chronicling the entire Medieval Life Expedition included footage from constructing a model village, writing and enacting a drama of the time, learning about and playing music of the time, and students answering the questions "What was life like then?" and "Would you rather have lived then than now?" The quality of the production was professional. It was even shown on local cable television and shown to other EL schools around the country. I wondered if there had been a professional editor who had put together this excellent product. The students and teachers assured me that it was the work of the students. The teacher in the Gifted and Talented program (which is not a high track program in the King's version) is very knowledgeable in video production. He describes the project he will guide and then allows any student who wishes, to apply for the project. He selects a small heterogeneous group of students who commit to doing a lot of work and coaches them in all aspects of professional video production. The student guide to the video spoke of all of the elements that went into the making of it--communication, investigation, storytelling, documentation, collaboration, media literacy, quality production, and complex thinking. He only lamented the necessity to constantly revise the work to make it better--an unusual practice for his age group. But, he added, "when you see how good it makes the final product, you don't mind so much." McCarthy noted that the high school teachers appreciate the King students' ability to edit their work. They also comment that students from the King ask them what the product of their learning will be.

Along with the structural changes that have been made over the seven years of ELOB at the King are important cultural changes in the school. Discipline issues are greatly

reduced and staffing is very stable compared to other schools in the District. In the words of the EL support person, "In general, the school works. People get along. It respects and honors kids."

WHAT THE TESTS SHOW

The student achievement scores that will be reported here are from two tests: the Maine Educational Assessment (MEA) and the District-wide Middle School Writing Assessment. MEA, a State-wide assessment administered yearly in October to the 4th, 8th and 11th grades reports the school's average scaled scores for each of the discipline areas as well as percentages of students scoring at determined performance levels. The data used in our report include MEA results from 1988-89 onwards, with cumulative scores reported for three-year intervals. We use the cumulative scores for comparative purposes, as does the State of Maine, since they more accurately reflect the performance of instructional programs because they are less sensitive to year-to-year differences in classes of students. Thus, we will present comparisons marking pre and post ELOB implementation years. In the last year for which the MEA data are available, 1998-1999, the test was completely revised, making the scores non-comparable across all the years we are considering in this report.

In 1996 the State of Maine adopted the Learning Results--Maine's version of content area Standards--which inform curriculum and instruction across the state and are measured by a standardized assessment (MEA). Updated versions of the MEA were administered to 4th and 8th graders in October, 1998. The major differences between the two versions are the following:

Old MEA	Revised MEA
<p>DESIGN</p> <p>Content was determined annually by teacher advisory committees representing the State's schools</p>	<p>DESIGN</p> <p>Content is determined by an external testing company and aligned with Maine's Learning Results content standards, using performance indicators for guidance.</p>
<p>TEST FORMAT</p>	<p>TEST FORMAT</p>

Open/constructed response questions, on which students had 10-15 minutes to develop their response. A single writing prompt in which students developed a narrative essay.

TIME

7 to 8 hours of student testing time.

SCORING

Advanced Systems staff score response items, using a 0-4 scale based on rubrics and question specific scoring guides.

Maine teachers score writing prompts.

Scaled scores range from 100-400.

Performance levels are divided into either

Basic and above or Advanced and above, or presented as quartile scores.

A variety of item types: approximately 60% open/constructed response questions of various lengths; 40% multiple choice questions. Multiple writing prompts which require different modes of writing.

TIME

10 to 13.5 hours of student time.

SCORING

Advanced Systems staff score response items based on rubrics and question

specific scoring guides, as well as multiple choice items . Maine teachers score

writing prompts using an analytic rubric for content development and writing

mechanics. The scaled scores range from 501-580. Performance levels correspond to expectations for meeting or exceeding the

Learning Results standards.

Given this change, the MEA scores will be reported as follows:

1. *Time 0*. Three-year cumulative average scores from school years 1989-90/1990-91/1991-92 will form the pre ELOB baseline.
2. *Time 1*. Cumulative scores from school years 1993-94/1994-95/1995-96 will provide a comparison for the first three years of ELOB implementation.
3. *Time 2*. Cumulative scores from school years 1995-96/1996-97/1997-98 will provide comparative data for the last years of the first version of the MEA, which coincides with the 3rd, 4th, and 5th years of ELOB implementation.
4. Single year scores from the 1998-99 school year will provide cross-sectional data on how King Middle School 8th graders perform in the 6th year of ELOB implementation compared to the other Portland middle schools as well as the State, since it is the first and only year available of scores on the new version of the MEA.

These data will provide a baseline for future evaluation of student achievement.

Demographic data, which are necessary to draw accurate inferences about the quantitative data, will be presented to show the increase in non-English speaking students over the years of ELOB implementation, as well as the socio-economic status of King Middle School families compared to those of other Portland middle schools, against which we are comparing King achievement scores. These data were reported along by the State along with MEA scores up until 1999. We do not have these data for the last year reported on here, but there was not a major change in demographics in 1999. In fact, the number of low income LEP students continues to increase in the District and specifically at the King.

District level Writing Assessment and Portfolio evaluations will also be reported and compared. A detailed explanation of these procedures and scores is presented in a later section.

MEA Data

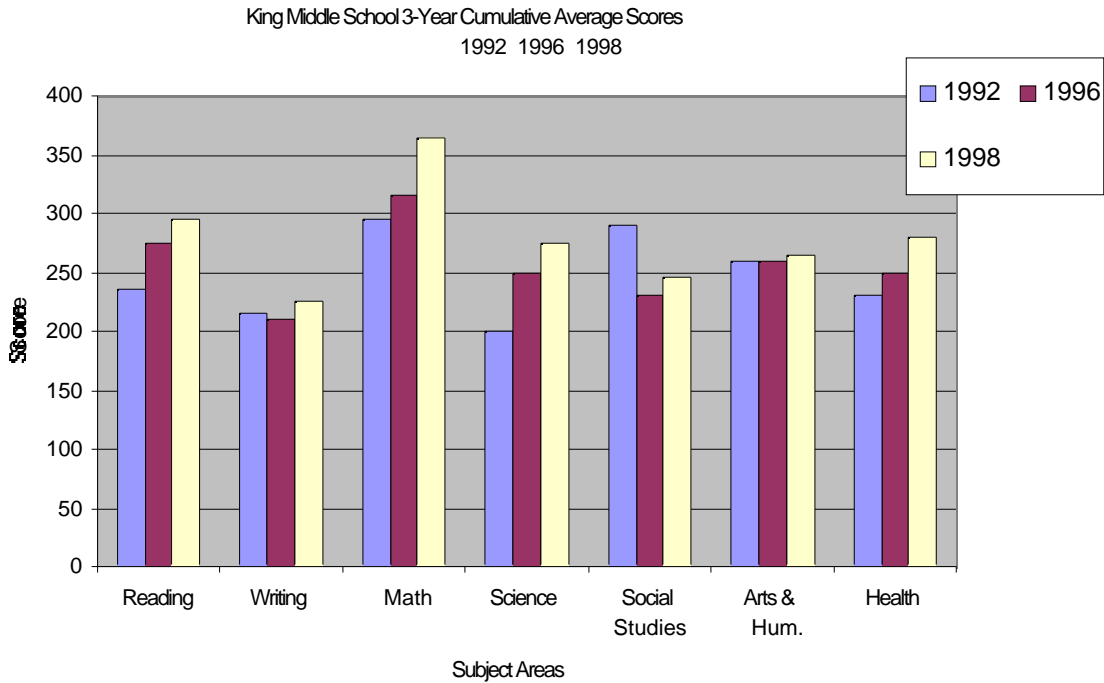
The following tables and chart show MEA three-year cumulative averaged scaled scores from 8th graders at the King Middle School for three time periods: Time 0--Pre ELOB (1989-92); Time 1--Early ELOB Implementation (1993-96); Time 2--Full ELOB Implementation (1995-98). Scores are reported for each of the subject areas that were included in the MEA. The results show that there was overall improvement in each successive period in all areas except for Social Studies at Time 1. Social Studies scores improve at Time 2 but not to the level of Time 0. There was also a slight dip in Writing at Time 1, but the subsequent score at Time 2 was a significant improvement over both Time 0 and Time 1.

School	Reading Cum Av. Score	Writing Cum Av. Score	Math Cum Av. Score
King MS 91-92	235	215	295
King MS 95- 96	275	210	315

King MS 97-98	295	225	365
------------------	-----	-----	-----

School	Science Cum Av. Score	Social Studies Cum Av. Score	Art & Hum. Cum Av. Score	Health Cum Av. Score
King MS 91-92	200	290	260	230
King MS 95-96	250	230	260	250
King MS 97-98	275	245	265	280

The following chart visually represents the above data, showing the increase in performance scores over time across subject areas.



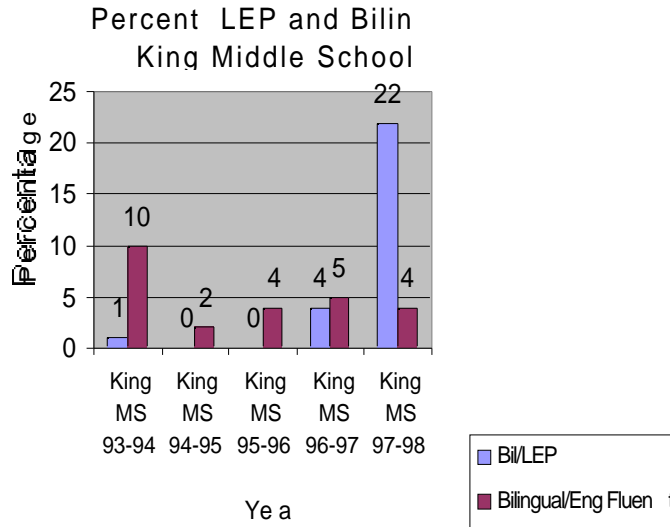
The King Middle School demonstrated consistent improvement in all areas of the MEA during the five years of ELOB implementation accounted for here. At the same time, the school's demography was changing rapidly. An influx of foreign-born immigration to the area dramatically increased the number of non-native speakers of English during the years we are considering here. The number of students who speak languages other than English has more than tripled in the Portland District over the past seven years. More than 40 different languages are spoken by these recent arrivals, with most of them speaking Khmer, Somali, Vietnamese, Serbo-Croatian, Spanish and Russian.⁸ Thus far, the King is the magnet middle school for multilingual students, and hence is assigned nearly all of the incoming non or limited English speaking students. In the 1993-94 school year, 98% of the students at the King were monolingual English speakers, 2% were bilingual but considered fluent in English and no students were officially considered Limited English Proficient

⁸ Languages are listed in the order in which they are numerically represented in the District.

(LEP).⁹ According to McCarthy, the 1999-2000 school year saw about 22% of the student body coming from homes where English is not the first language. The State Department of Education estimated that approximately 45 middle school students came through the Multilingual Intake Center, and 4 out of 5 of these students were assigned to multilingual classes (ESOL). The remaining 8 or so students were included in the mainstream. In spite of this increase in bilingual and LEP students, the King showed steady progress in all subject areas. Moreover, each year a larger percentage of LEP students are taking the MEA. The relatively smaller gains in Writing and Social Studies compared to Math, Reading, and Science may be due to the number of non-native English speaking students taking the test. Correct on-demand writing, as required by the MEA, proves to be particularly challenging for non-native speakers of the language, and Social Studies undoubtedly requires a background knowledge of culture and history of the US that may be at rudimentary levels for new arrivals. On the other hand, the high gains in areas like Reading, Science and Math are even more remarkable given the King's large immigrant population. Issues of language comprehension and background knowledge also impact these subject areas and students' performance, across the board, speaks to an instructional environment that is rich in content and language use.

The following charts indicate the growth of King's non-native English speaking population and compare it to the State and the District. The table and graph below show the large increase in LEP student enrollment at the King over the years of ELOB implementation.

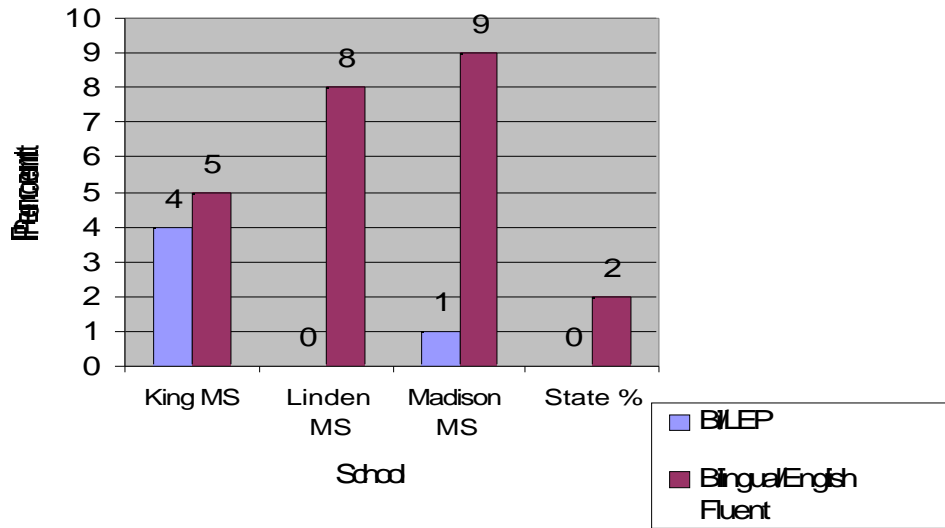
⁹ Data are derived from Maine's State Department of Education publications of yearly MEA and school demographic data. As mentioned above, these data are not available for the 1998-99 or 1999-2000 school years.



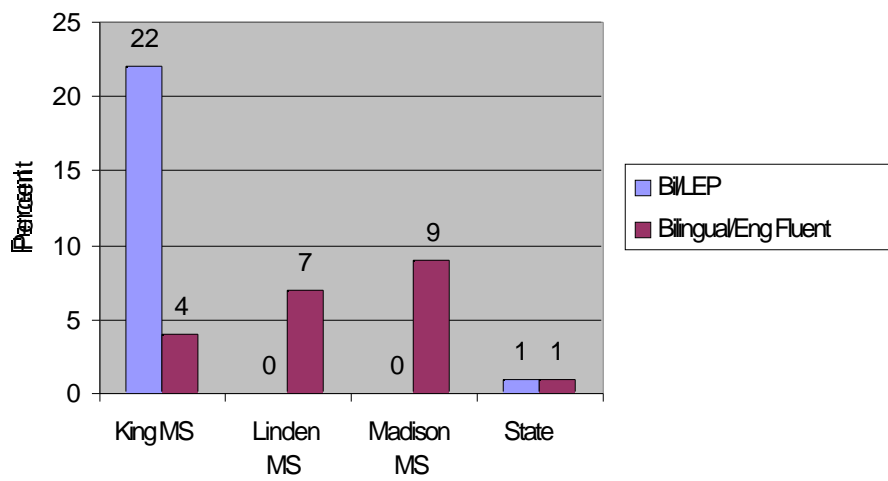
Year	Bilingual/ LEP	Bilingual/ English Fluent
93-94	1	10
94-95	0	2
95-96	0	4
96-97	4	5
97-98	22	4

The next two charts compare King's enrollment of LEP and Bilingual/English Fluent students with the other Portland Middle Schools and the State for the 96-97 and 97-98 school years.

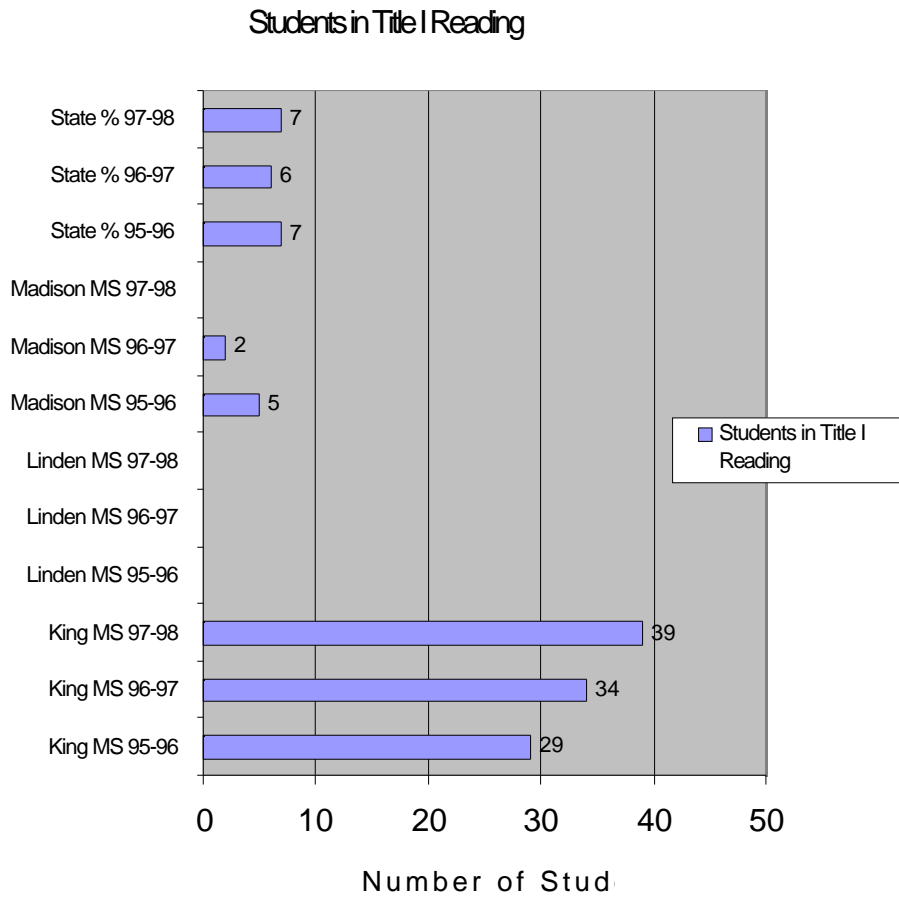
LEP & Bilingual Enrollment Compared
96-97



LEP & Bilingual/English Fluent Enrollment
Compared 97-98



We see from these charts that the King Middle School is very different in its composition of language backgrounds compared to either District middle schools or State-wide middle school populations. Similarly, we see a larger number of King students receiving special services, such as Chapter 1, when compared to the District middle schools as well as to the rest of the State. The following chart compares three school years--1995-96, 1996-97 and 1997-98. Over this time period we see the not-surprising trend of increasing numbers of students at the King Middle School needing special reading services, given the rise in LEP enrollment.



One final set of demographic data charts helps tell the socio-economic status story of the King Middle school. Portland, Maine, while not a large urban center compared to those in other states, nevertheless is this state's most urban area. As such, it contains a heterogeneous population in terms of class and economic indicators. Unlike most urban districts, however, Portland has not seen a large exodus of middle and upper middle class students from its public schools. The District, therefore, has a wide range of family backgrounds in its schools. This does not mean, however, that all of the social categories are equally represented in each school. The middle schools of Portland, as we will illustrate, differentiate along socio-economic indicators with the King enrolling a substantially poorer student body with less parent education and white collar employment than the two other schools. The Linden Middle School, by contrast, enrolls the largest number of middle and upper-middle class students. The Madison Middle School falls somewhere in between these two. The State average on these indicators, predictably falls somewhere in between the King and the Linden levels on socio-economic indicators.

The following data, collected from survey information filled out by students and principals as part of the MEA administration procedures, is reported along with MEA scores for comparative purposes. The three indicators that are used to represent socio-economic status are the following:

1. **Parent Education Index (PEI).** This index is measured on a range of 1-3, according to how students answered the question about Parent Education Level:
 - 1 I don't know
 - 1 Did not finish high school
 - 1 Graduated from high school
 - 2 Some education after high school
 - 3 Graduated from college

2. **Community Occupations Index (COI).** This index is computed by subtracting the points for the unskilled workers from the points for the white-collar workers, making the range for the COI -4 to +4, with higher values indicating a higher proportion of white collar workers.

3. **Free or Reduced Price Lunch Index (FLI).** This index was derived from responses on the Principal Questionnaire administered with the yearly MEA. They were asked to provide the percentage of 8th grade students eligible for the Federally subsidized lunch program. Possible responses and the points associated with them are as follows:

Points	% of 8th graders qualifying for Free or Reduced Lunch
1	Over 50%
2	41 to 50%
3	31 to 40%
4	21 to 30%
5	under 20%

The following chart compares the Portland Middle schools for socio-economic indicators for the 1997-98 school year (the last year for which these indicators are officially reported by the State)¹⁰. A higher number on the PEI Index indicates a higher level of education for the parent population of the school. King scores lower than the other middle schools on this measure. Given that there is only a three point range for this index, the small difference in column heights for the three schools is deceptive. In fact, the Parent Education Index for King indicates that most parents of the school's population barely have, or have little more than a high school education, whereas both Linden and Madison PEI levels indicate that the majority of their parents have some level of higher education beyond high school.

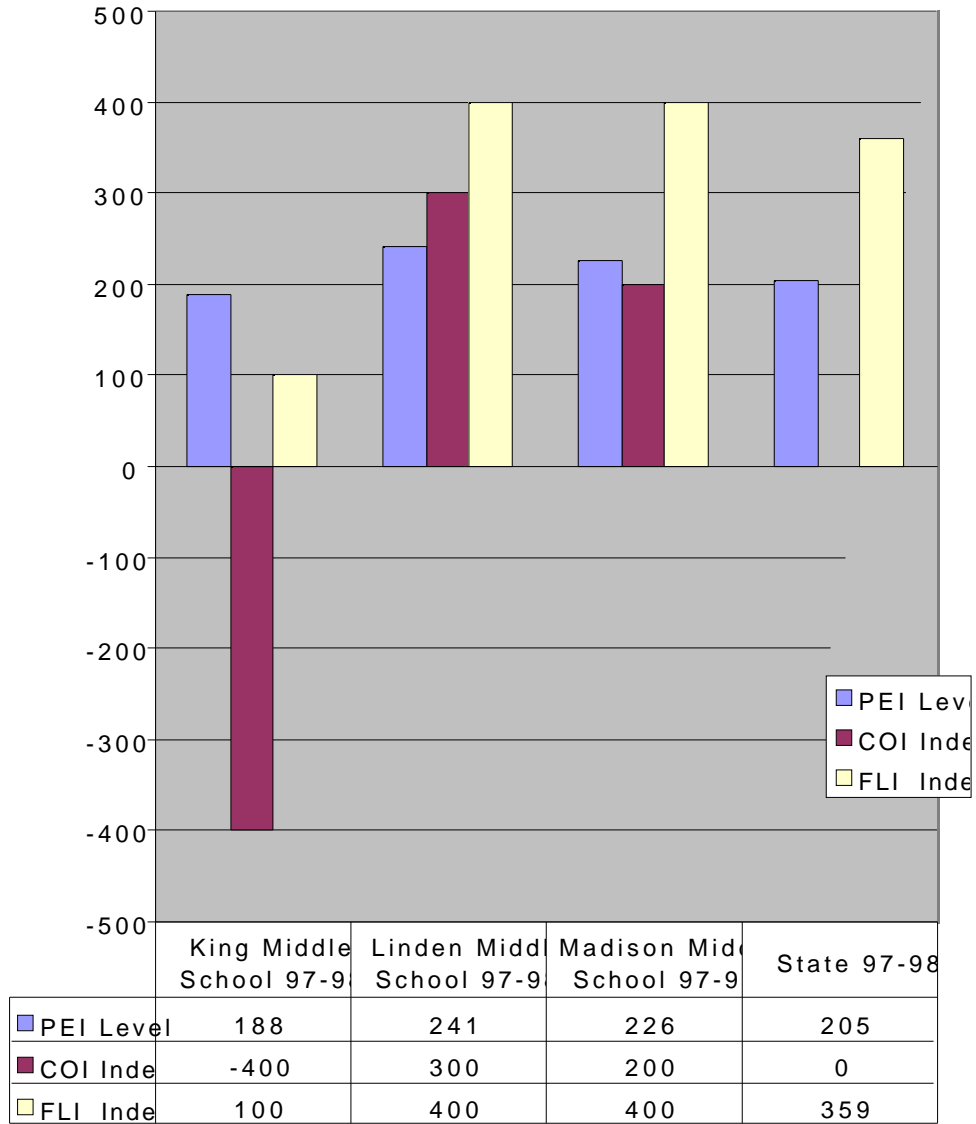
The COI--the index representing community occupations--shows the greatest fluctuation amongst the schools. King scores at the lowest possible level for this indicator, no doubt a reflection of the low-skill jobs filled by less educated and immigrant, non-English speaking parents. Linden and Madison, on the other hand, both have relatively high scores in this area, indicating a much larger percentage of white collar workers in their populations.

¹⁰ N.B. that the Index scores have been multiplied by 100 on the graphs to facilitate visual representation of the compared scores.

The FLI measures the percentage of students participating in federally funded reduced or free lunch programs. Counter-intuitively, the lower the number assigned to the school the higher is the percentage of students eligible for this benefit. The King scores significantly lower, meaning that it has a much higher percentage of eligible students, than the two other middle schools.

In sum, the *higher* the bar for the PEI, the more educated the parents; the *higher* the bar for the COI, the more white collar jobs held in the community; the *higher* the bar for the FLI, the fewer students qualify for free or reduced lunch.

SES Indicators 1997-



The chart of numbers below shows that this comparison holds constant for the time we are reporting standardized test scores in this report. The State and the District have similar averages on all of these measures, while the King Middle Schools indices show that its parent population is much poorer, less educated, and holding far more low skilled jobs than the State at large as well as the District.¹¹

Name of School & Year	PEI	COI	FLI
King Middle School 95-96	1.95	-4	1
District 95-96	2.15	0.66	3.39
State 95-96	2.04	-0.43	3.43
King Middle School 96-97	2.20	-4	1
District 96-97	2.21	0.17	3.79
State 96-97	2.05	-0.32	3.52
King Middle School 97-98	1.88	-4	1
District 97-98	2.20	0.59	3.11
State 97-98	2.05	0.03	3.59

Comparing the population of students attending the King Middle School to their District and State counterparts, the steady achievement on the MEA over the years of ELOB implementation in all areas indicates that the instructional program at the King is highly successful, and particularly successful with a challenging, normally low-achieving population.¹² The next few tables and charts show students' increasing academic achievement as measured by standardized tests at the King Middle School. We present the tables of cumulative average scores along with the chart comparing gains for the 2

¹¹ The District comparison indices includes the King scores as well, which means that the Linden MS and the Madison MS score at higher comprehensive SES levels than the State average.

¹² LEP students tend to score lower than their monolingual English counterparts on standardized tests, particularly in those subject areas that require subtle language discrimination, according to numerous reports on urban schools, standardized achievement tests, and on second language populations. Moreover, since the early Coleman reports of the 1960s, class differences in student bodies have also been evident in standardized test scores in a number of large scale assessment studies. The higher the students class, based

three-year cumulative scores reported in 1995-96 and 1997-98 for all three Portland middle schools as well as the State. The tables show that the King average scores are lower than the two other middle schools in Portland, but on the high end of the three-year Comparative Score Band which is calculated for each school based on socio-economic indicators. In other words, the Comparative Score Band for King (which we present below each of the tables) sets the predicted average cumulative scaled score range for a school with similar demographic characteristics. The tables also show that the gains made by King 8th graders over the five-year period we are considering are substantial. The gains meet or exceed the other middle schools in the District, in some cases by a significant amount. Only Writing demonstrates modest gains compared to the Linden Middle School.

Reading

	State Av. Cum Score	King Av. Cum Score	King Change	Madison Av. Cum Score	Madison Change	Linden Av. Cum Score	Linden Change
1995-96	295	275		310		335	
1997-98	300	295	20	330	20	350	15

Comparative Score Band 265-275

Writing

	State Av. Cum Score	King Av. Cum Score	King Change	Madison Av. Cum Score	Madison Change	Linden Av. Cum Score	Linden Change
1995-96	265	210		245		255	
1997-98	265	225	15	250	5	285	30

Comparative Score Band 230-260

Math

	State Av. Cum Score	King Av. Cum Score	King Change	Madison Av. Cum Score	Madison Change	Linden Av. Cum Score	Linden Change
1995-96	335	315		330		355	
1997-98	360	365	50	365	35	390	35

Comparative Score Band 325-350

on socio-economic indicators, and the higher the level of parental education, the more likely the student is to score highly on standardized tests of achievement.

Science

	State Av. Cum Score	King Av. Cum Score	King Change	Madison Av. Cum Score	Madison Change	Linden Av. Cum Score	Linden Change
1995-96	285	250		280		280	
1997-98	280	275	25	285	5	305	25

Comparative Score Band 260-290

Social Studies

	State Av. Cum Score	King Av. Cum Score	King Change	Madison Av. Cum Score	Madison Change	Linden Av. Cum Score	Linden Change
1995-96	250	230		270		275	
1997-98	240	245	15	270	0	285	10

Comparative Score Band 210-240

Arts & Humanities

	State Av. Cum Score	King Av. Cum Score	King Change	Madison Av. Cum Score	Madison Change	Linden Av. Cum Score	Linden Change
1995-96	275	260		290		310	
1997-98	270	265	5	285	-5	295	-15

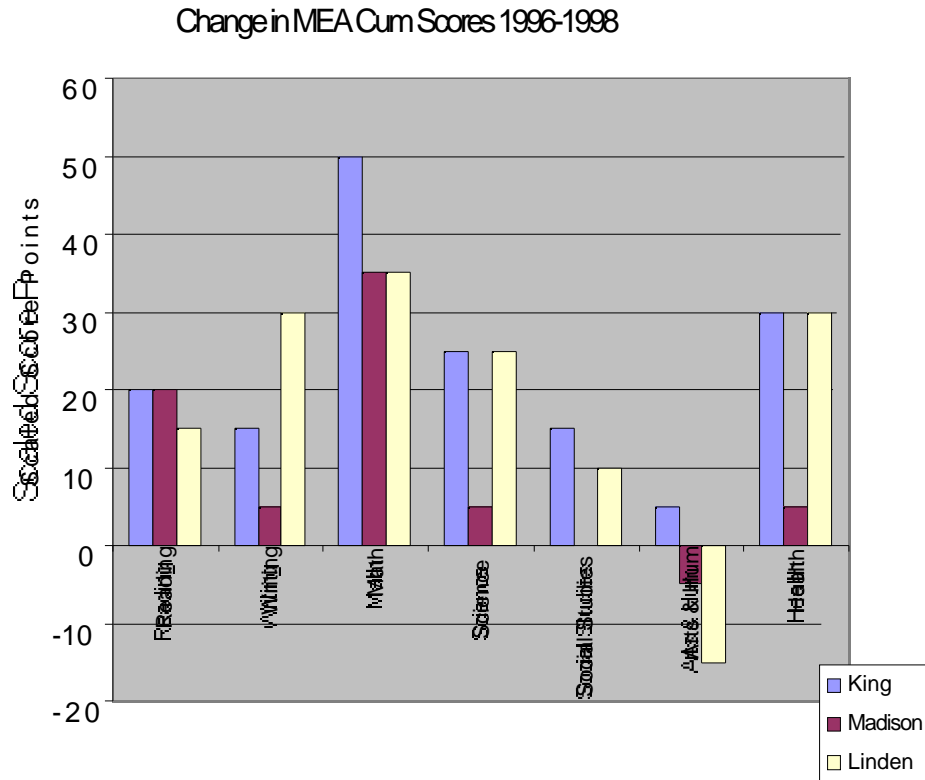
Comparative Score Band 240-270

Health

	State Av. Cum Score	King Av. Cum Score	King Change	Madison Av. Cum Score	Madison Change	Linden Av. Cum Score	Linden Change
1995-96	265	250		295		280	
1997-98	275	280	30	300	5	310	30

Comparative Score Band 250-280

The following bar graph graphically represents the gains in each of the subject areas for all three Portland Middle schools.



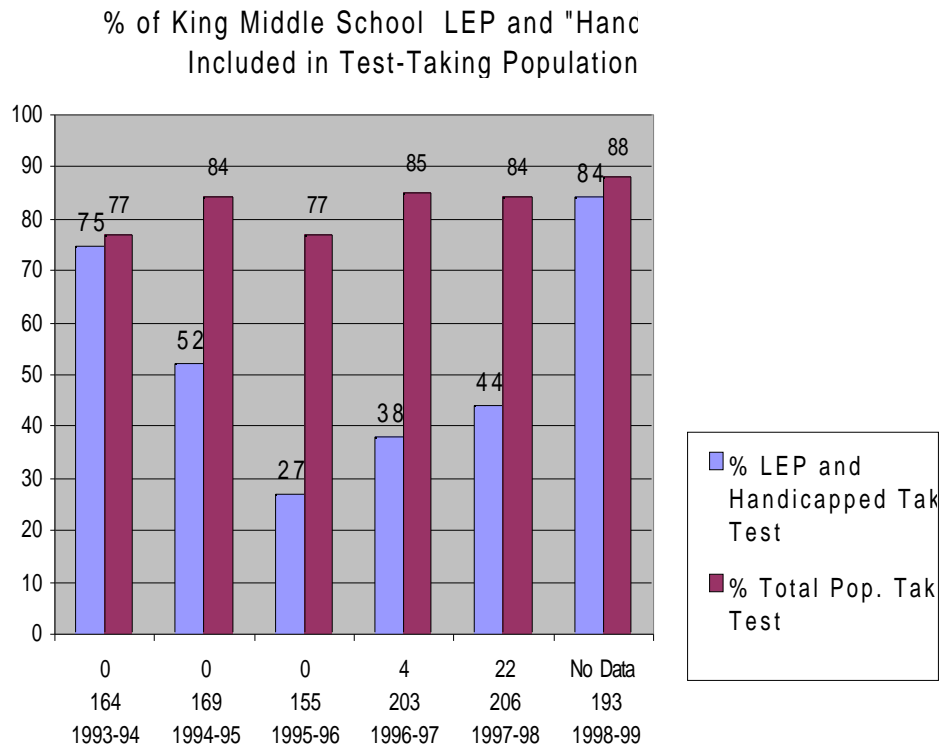
These gains were achieved at the same time that the King was enrolling more LEP students and as it included more of them in the testing procedure. Only students who complete all parts of the test are included in the report of MEA scores. This is also true for considering the percentage of LEP and "Handicapped" students included in the test.¹³ The data for all but the last year of active implementation of ELOB show a steady increase of students included in the tests.¹⁴ The three labels at the bottom of each bar correspond to:

1. Number of LEP students taking test¹⁵
2. Number of students taking test
3. Year

¹³ The State defines any student who is exempt from taking the MEA because of lack of English proficiency, a physical, cognitive or learning disability as Handicapped in the reporting of test-taking students.

¹⁴ 1999-2000 data were not available at the time of this report.

¹⁵ No breakout data of numbers of LEP test takers were provided by the State for 1998-99



These data--which show the progressive increase in students typically eligible for exclusion from testing within the total test taking population--together with the significant increases in average scaled scores over the years, allow us to infer that the inclusion of special needs and LEP students in learning expeditions and the elimination of tracking at the King has benefited the student body as a whole.

Revised-MEA Data

As stated above, the 1998-99 MEA breaks with the former MEA tests in content, format and scoring procedures. The 1998-99 results to show that the trend we have seen with the previous MEA scores is maintained with the new system. In fact, there is very

little difference between the Portland Middle School Scaled Scores in each subject area on the new MEA. The King performs better in every subject than the Linden Middle School, which has the highest SES according to reported indices. It also performs at a level better than or at least equal to the State average in all subject areas.

School & Year	Reading	Writing	Math	Science	S Studies	Arts	Health
King 98-99	541	536	532	529	537	534	541
Linden 98-99	539	533	532	529	537	534	540
Madison 98-99	542	534	531	531	538	537	543
State 98-99	539	533	529	528	537	532	541

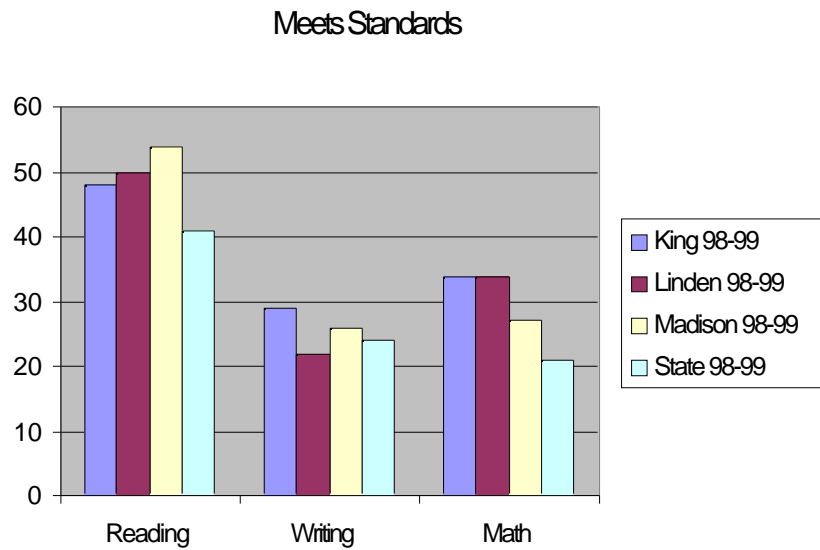
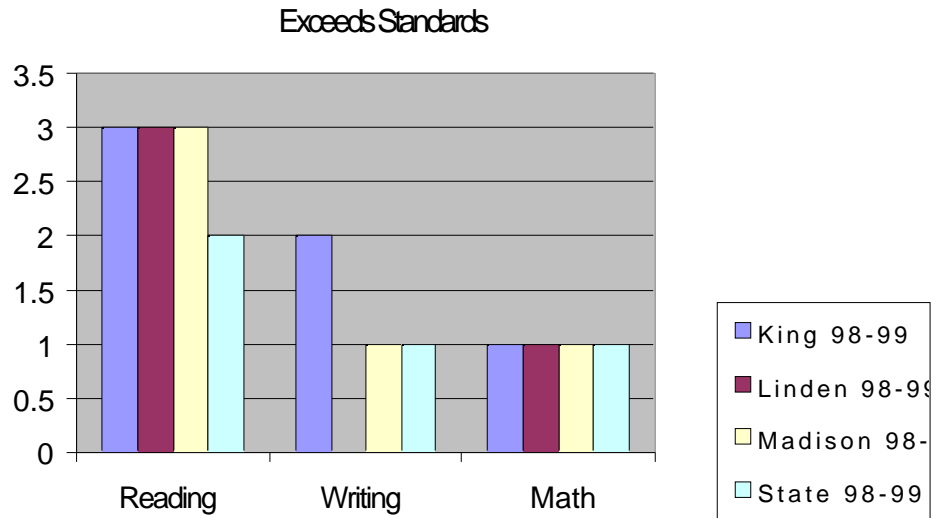
The 1998-99 MEA scores for each subject are reported in percentages of students in each of 4 Performance Levels, which are based on the Standards expressed in the Maine Learning Results. The four levels are:

1. **Exceeds Standards**
2. **Meets Standards**
3. **Partially Meets Standards**
4. **Does Not Meet Standards.**

As the following graphs show for Reading, Writing and Math, the King compares very favorably to the other District middle schools as well as to the State average percentages.

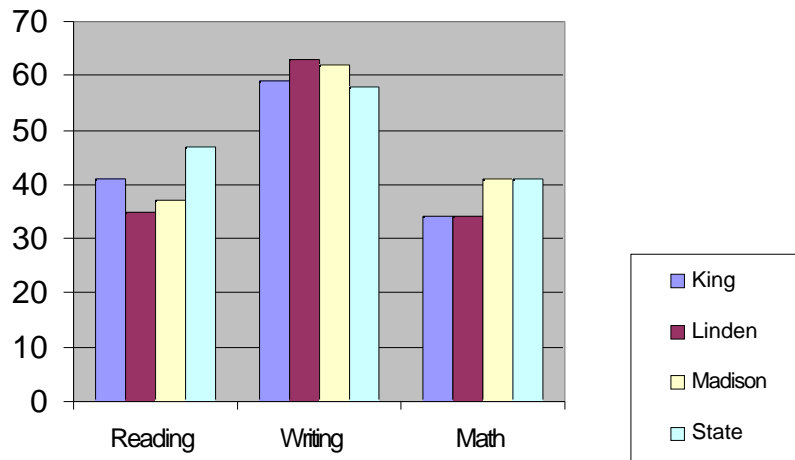
King Middle School has the largest percentage of students in the top category when compared to the other Portland middle schools and the State. Moreover, the generally poorer performance in Writing that we saw from King 8th graders in earlier versions of the MEA has been reversed. King 8th graders now significantly outperforms all of the comparative groups in the top performance category.

King 8th graders are also comparatively well represented in the second performance level. Combining the first two performance level percentages, we see that **51%** of students in Reading, **31%** in Writing, and **35%** in Math score at or above the standard set by the State for these subject areas. This compares favorably with the District and State middle schools.

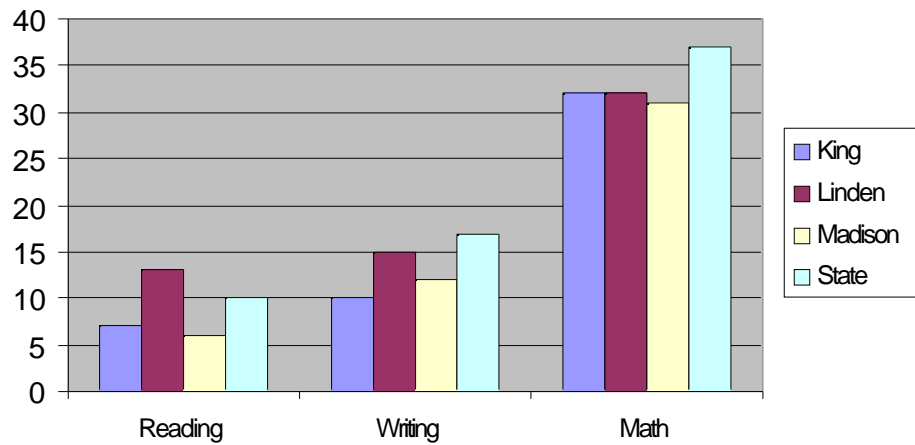


The next two performance levels, Partially Meets Standards and Does Not Meet Standards, are generally less well represented by King 8th graders than their comparison groups. This means that fewer King students are not performing at the expected level of high performance in these subjects.

Partially Meets Standards



Does Not Meet Standards



The percentages for the other subject areas compare similarly and so are not presented here.

A final comment on the MEA scores of 1998-99 regarding LEP student performance at the King: While the performance levels for LEP students are lower than for their Monolingual English counterparts, the difference is not dramatic. In fact, in the minor subject areas of Social Studies, Arts and Health, the LEP students are performing at closer levels to their native English speaking classmates. This may indicate a more rigorous set of standards for the main subject areas of Reading, Writing, Math and Science or less reliance on subtleties of language in the tests for the minor subject areas.

Reading	Exceeds	Meets	Partially Meets	Does Not Meet
Monolingual English	3	54	39	3
Biling/ LEP	3	14	55	28

Writing	Exceeds	Meets	Partially Meets	Does Not Meet
Monolingual English	3	36	54	7
Biling/ LEP	0	5	73	23

Math	Exceeds	Meets	Partially Meets	Does Not Meet
Monolingual English	0	38	32	30
Biling/ LEP	0	10	38	52

Science	Exceeds	Meets	Partially Meets	Does Not Meet
Monolingual English	1	21	55	23
Biling/ LEP	0	10	24	66

Social Studies	Exceeds	Meets	Partially Meets	Does Not Meet
Monolingual English	1	38	51	10
Biling/ LEP	3	21	48	28

Arts	Exceeds	Meets	Partially Meets	Does Not Meet
Monolingual English	1	35	45	18
Biling/ LEP	0	17	59	24

Health	Exceeds	Meets	Partially Meets	Does Not Meet
Monolingual English	0	50	48	3
Biling/ LEP	0	17	69	14

Writing Assessments

The second assessment we report on here is the Portland Public Schools Middle School Writing Assessment. This assessment practice began in the 1994-95 school year and all of the data reported to date are included in this report.

The Middle School Writing Assessment complements the MEA, but is District mandated and managed. Each middle school administers the same two writing prompts (A and B) randomly to all students in the 6th, 7th and 8th grades. This is done in the Fall and, exchanging the A and B prompts for the students, again in the Spring. A random sample of tests from each grade in each middle school are scored by a team of approximately 10 middle school teachers and reading consultants representing all the middle schools in the District.

The scoring of the prompts follows this procedure: Each school picks a random sample of approximately 30 writings, 10 from each grade. The team of scorers (different from those who pick the samples from each school) then rate the writings using the MEA writing rubric which allows a maximum of 6 points. Scorers are not told which grade level or which school the writing sample is from. Each writing sample is scored by two individuals and their scores are combined, allowing for a possible score range from 2-12. Any sample that was scored with a full point difference between scorers was arbitrated by a third reader. The scoring is remarkably consistent. In the 1994-95 year, for example, only five out of 360 papers needed arbitration.

The following chart shows the average scores on each of the two prompts for each grade and each school for the Fall and Spring. It also shows the amount of change from Time 1 to Time 2 for the random sample of essays. These scores do not indicate that there is a great difference between writers in the three middle schools. While the King scores are somewhat lower than the other schools, and include more LEP students, the variation in average scores is not significant. The King falls in the middle of the three Portland middle schools in terms of gains recorded from Fall to Spring each year. Madison does slightly better than the King and Linden does slightly worse.

	<i>Madison</i>	<i>King</i>	<i>Linden</i>
<i>Number of times school showed positive change from Fall to Spring.¹⁶</i>	27	23	21
<i>Number of times school showed highest gains from Fall to Spring</i>	13	11	7
<i>Number of times school showed no gain or negative gain from Fall to Spring</i>	3	6	8

The King shows greater improvement in writing scores in the last two years of this assessment than in the first two, however.

1997-98	5 out of 6 groups ¹⁷ showed gains
1998-99	4 out of 6 groups showed gains
1994-95	3 out of 6 showed gains
1995-96	1 out of 6 showed gains

The complete table of scores and changes is presented below. Red scores indicate that the King's score is higher than the District average. Blue scores indicate it is the same as the District average.

Prompt A 1994-95					Prompt B 1994-95		
Grade	School	Fall	Spring	Change	Fall	Spring	Change
6	King	4.4	5.7	1.3	4	5.4	1.4
6	Linden	5.2	6.1	0.9	5.4	5.5	0.1
6	Madison	5.5	6	0.5	5.1	6.2	1.1
Average		5	5.9	0.9	4.8	5.7	0.9

¹⁶ This number is out of a possible 30. three scores per school (6th, 7th, and 8th grade), two times a year, for five years (3 x 2 x 5).

¹⁷ Groups are defined as Prompt A & B for each of the three years tested (6th, 7th and 8th grade) in each school year. Thus, the six groups total for each year derive from two sets of prompts in each grade level per year.

7	King	5.1	5.2	0.1	5.8	5.7	-0.1
7	Linden	5.7	5.7	0	5.3	6.6	1.3
7	Madison	5.6	6.1	0.5	6.4	6.8	0.4
Average		5.5	5.7	0.2	5.8	6.4	0.6

8	King	7.6	6.7	-0.9	5.7	6.8	1.1
8	Linden	6.1	7.8	1.7	7.5	6.7	-0.8
8	Madison	6.7	7.1	0.4	6.3	6.2	-0.1
Average		6.8	7.2	0.4	6.5	6.5	0

Prompt A 1995-96					Prompt B 1995-96			
Grade	School	Fall	Spring	Change	Fall	Spring	Change	
6	King		5	5.8	0.8	4.5	5.5	1
6	Linden		4.8	5.2	0.4	5.3	5.9	0.6
6	Madison		4.8	6.6	1.8	4	6	2
Average			4.9	5.9	1	4.6	5.8	1.2

7	King	4.2	5.1	0.9	5.6	5.3	-0.3
7	Linden	5.5	7	1.5	6.2	5.8	-0.4
7	Madison	5.3	6.6	1.3	4.8	5.3	0.5
Average		5	6.2	1.2	5.5	5.5	-0.1

8	King	4.9	8.1	3.2	7.3	6.2	-1.1
8	Linden	6.6	7	0.4	6.8	7.2	0.4
8	Madison	6.7	6.5	0.2	5.9	6.1	0.2
Average		6	7.2	1.1	6.7	6.5	-0.2

Prompt A 1996-97					Prompt B 1996-97		
Grade	School	Fall	Spring	Change	Fall	Spring	Change
6	King	4.1	4.8	0.7	4	5.8	1.8
6	Linden	4	4.6	0.6	4	5.7	1.7
6	Madison	4.6	5.3	0.7	4.1	5.9	1.9
Average		4.2	4.9	0.7	4	5.8	1.8

7	King	4.1	5.1	1	4	5.8	1.8
7	Linden	4.7	5.4	0.7	3.6	5.8	2.2
7	Madison	4.5	5.6	1.1	5	5.9	0.9
Average		4.4	5.4	0.9	4.2	5.8	1.6

8	King	5.6	6.5	0.9	6	5.8	-0.2
8	Linden	5.5	6.5	1	7	7	0
8	Madison	5.3	6.5	1.2	5.3	5	-0.3
Average		5.6	6.5	1.1	6.1	5.9	-0.3

Prompt A 1997-98					Prompt B 1997-98		
Grade	School	Fall	Spring	Change	Fall	Spring	Change

6	King	4.8	5.7	0.9	4.7	5.7	1
6	Linden	5.4	5.9	0.5	5.4	5.3	-0.1
6	Madison	5	5.1	0.1	6	6.9	0.9
Average		5.1	5.6	0.5	5.4	6	0.6
7	King	5.8	6.5	0.7	4.9	6	1.1
7	Linden	6.5	5.8	-0.7	5	5.1	0.1
7	Madison	6.8	5.8	1	5	6.2	1.2
Average		6.4	6	.06	5	5.8	0.8
8	King	5.6	6.3	0.7	5.9	5.1	-0.8
8	Linden	6.2	6	-0.2	6.8	6.4	-0.4
8	Madison	6.2	6.9	0.7	5.4	7.8	2.4
Average		6	6.4	0.4	6	6.4	0.4

Prompt A 1998-99					Prompt B 1998-99		
Grade	School	Fall	Spring	Change	Fall	Spring	Change
6	King	4.5	5.5	1	3.7	5.8	2.1
6	Linden	5.6	6.2	0.6	5.3	5.9	0.6
6	Madison	5	5.4	0.4	4.5	6.1	1.6
Average		5	5.7	0.7	4.5	5.9	1.4
7	King	5.2	6.1	0.9	5.6	6.6	1
7	Linden	5	5.5	0.5	4.7	6.3	1.6
7	Madison	5.8	6.4	0.6	5.8	6.6	0.8
Average		5.3	6	0.7	5.4	6.5	1.1
8	King	5.9	5.6	-0.3	4.7	5.4	0.7
8	Linden	6.5	6.6	0.1	6.9	7.5	0.6
8	Madison	6.2	8	1.8	7	6.9	-0.1
Average		6.2	6.7	0.4	6.2	6.6	0.4

Learning Expeditions, which include lengthy collaborative investigations and products that are revised to meet high expectations of the students and their teachers would be likely to help improve the thinking and writing skills of students who engage in them regularly. This may account for the changes we see in writing performance over time at the King school. It appears that as the years of ELOB implementation increase, the number of students who register a positive change in their writing also increases. A second measure supports this inference. Spring writing sample scores were compared to Portfolio writing assessments for the same sample of students in each of the years that the

Writing Assessment was given. Portfolios were evaluated by the team of middle school teachers and reading specialist scorers, again using the MEA writing rubric. Three pieces of writing were evaluated for each portfolio and, on that basis, a Portfolio Assessment Score was assigned. As with the Writing Assessment, two scorers rated each portfolio and scores that differed--which occurred rarely--were arbitrated. The same scale of possible points from 2-12 was, therefore, used for both Writing Sample and Portfolio ratings. The grade averages for each school were then compared and a score indicating the difference between the on-demand writing score and the portfolio score were tabulated. Over the period from 1994 to 1999, the King students showed a significantly higher difference (improvement) in scores on their portfolios. Knowing how to revise and make a better product is part of Expeditionary Learning instruction; having the opportunity to do it helps all students present better work, but particularly those whose first language is not Standard English.

	<i>King</i>	<i>Madison</i>	<i>Linden</i>
<i>Sum of differences in scores between Spring Writing Samples and Portfolios from 1995 to 1999 (Positive score indicates higher assessment on Portfolio).</i>	16.22	11.58	9.3

The table of scores is reproduced here. Red scores indicate that the difference in King portfolio and writing sample scores was highest for that grade that year.

Grade	School	Average Portfolio Score	Writing Sample Score	Difference
6	King	7.28	7.13	0.15
6	Linden	*	*	*
6	Madison	6.93	6.13	0.8
Average		7.11	6.63	0.48

1994-95				
Grade	School	Average Portfolio Score	Writing Sample Score	Difference
7	King	6.33	5.33	1
7	Linden	6.2	6.9	-0.7
7	Madison	7.23	6.38	0.85
Average		6.59	6.2	0.39

1994-95				
Grade	School	Average Portfolio Score	Writing Sample Score	Difference
8	King	7.2	6.38	0.82
8	Linden	7.8	7.6	0.2
8	Madison	6.08	6.25	-0.17
Average		7.03	6.74	0.29

1995-96				
Grade	School	Average Portfolio Score	Writing Sample Score	Difference
6	King	7.2	4.25	2.95
6	Linden	5.6	6	-0.4
6	Madison	8.5	7.1	1.4
Average		7.1	5.78	1.31

1995-96				
Grade	School	Average Portfolio Score	Writing Sample Score	Difference
7	King	7.6	6	1.6
7	Linden	5.7	6.3	0.6
7	Madison	6.7	6	0.7
Average		6.7	6.1	0.6

1995-96				
Grade	School	Average Portfolio Score	Writing Sample Score	Difference
8	King	7	6.8	0.2
8	Linden	8.1	7.7	0.4
8	Madison	6.7	7	-0.3
Average		7.26	7.16	-0.12

1996-97				
Grade	School	Average Portfolio Score	Writing Sample Score	Difference
6	King	8.6	6.4	2.6
6	Linden	8.3	5	3.3
6	Madison	8	5.4	2.6
Average		8.3	5.6	2.8

1996-97				
Grade	School	Average Portfolio Score	Writing Sample Score	Difference
7	King	6.9	6.1	0.8
7	Linden	8.3	6.2	2.1
7	Madison	7.3	6	1.3
Average		7.5	6.1	1.4

1996-97				
Grade	School	Average Portfolio Score	Writing Sample Score	Difference
8	King	7.4	7.7	-0.3
8	Linden	8.2	7	1.2
8	Madison	7.6	7.4	0.2
Average		7.7	7.4	0.4

1997-98				
Grade	School	Average Portfolio Score	Writing Sample Score	Difference
6	King	6.1	5.4	0.7
6	Linden	7.4	6.3	1.1
6	Madison	7.7	5.6	2.1
Average		7.1	5.8	1.3

1997-98				
Grade	School	Average Portfolio Score	Writing Sample Score	Difference
7	King	7.2	5.4	1.8
7	Linden	7.2	5.7	1.5
7	Madison	6.9	6.3	0.3
Average		7.1	5.8	1.3

1997-98				
Grade	School	Average Portfolio Score	Writing Sample Score	Difference
8	King	7.5	5.4	2.1
8	Linden	7.2	5.8	1.4
8	Madison	7.9	7.3	0.6
Average		7.5	6.2	1.3

1998-99				
Grade	School	Average Portfolio Score	Writing Sample Score	Difference
6	King	6.8	5.9	0.9
6	Linden	4.75	5.9	-1.2
6	Madison	6.5	5.6	0.9
Average		6	5.8	0.2

1998-99				
Grade	School	Average Portfolio Score	Writing Sample Score	Difference
7	King	7	6.9	0.1
7	Linden	5.4	6.1	-0.7
7	Madison	6.7	6	0.7
Average		6.4	6.3	0.1

Grade	School	1998-99		
		Average Portfolio Score	Writing Sample Score	Difference
8	King	5.4	4.6	0.8
8	Linden	6.9	6.4	0.5
8	Madison	6.4	6.8	-0.4
Average		6.4	5.9	0.3

In addition to demonstrating a greater ability in Portfolio (i.e., edited) writing, we also notice from these tables that the average scores for both Portfolios and the Writing Sample do not vary greatly among the three schools. This is noteworthy since the King enrolls a very high percentage of ESL students compared to the other middle schools, a fact that would tend to cause King students, as a whole, to score consistently lower than their counterparts in other schools in the District.

CONCLUSIONS

The two questions we posed for this report--What was the academic achievement as measured by standardized tests for the King Middle School over the years of ELOB implementation? and, What impact did ELOB have on the academic achievement of the King Middle School students?--can be summarized as follows:

Comparing the performance of 8th grade students on the same type of standardized test over the years prior to ELOB implementation to recent years of active ELOB implementation, we have seen a steady increase in all discipline areas in student attainment levels.

Comparing the changes over time of King Middle School students' cumulative average scaled test scores with the two other District Middle Schools as well as the State averages, we see that the King generally scored lower than its comparison groups in individual disciplines. In terms of gains over time, however, in four out of seven subject areas, the King demonstrated the highest gains in scores of any of the District middle schools. In two out of seven disciplines, its gains were as high as the other top gaining middle school in the District. In other words, in only one

discipline area, Writing, did the King students not improve more than the other district middle schools.

In the final year of MEA testing, which instituted a new test format and process, the King students scored better than the State average in all of the disciplines except Health, in which it scored the same as the State average. Instead of scoring below the two other District middle schools in Writing, as it had on previous MEAs, the King scored higher than either comparison school. It never scored lower than both of the comparison schools in any discipline area (in several areas it scored the same as the other lower-scoring middle school). In all disciplines, however, the differences amongst the scores of Portland's middle schools was minimal.

Student writing in all three grades of the King Middle School has also improved over time as assessed by a District-wide Writing Assessment process. Students generally performed better on Portfolio assessments, which allowed for revision of final products than on on-demand writing assessments. The gains of the King Middle School students over time on the District Writing Assessment fell between the two other middle schools in Portland. More King students, in all three grades, demonstrated gains in writing assessments in recent years than in earlier years of the Writing Assessment.

These indicators of academic achievement show a school that is consistently improving its performance over the years of ELOB implementation, often at a greater rate than the comparison schools in the District, which are not ELOB schools. While we cannot prove a cause and effect relationship between ELOB and these test scores, the comparative data with other non-ELOB schools as well as the qualitative information that attests to the embeddedness of ELOB as a philosophy and practice in all aspects of the school curriculum, structure and culture allow us to infer that ELOB has had a positive impact on student academic achievement, as measured by standardized tests, at the King Middle School.

The comparative gains of the King over the years of ELOB implementation are even more impressive when we take into account the shifting population of the school. Over the years we have looked at here, the King's student population has transformed from a basically monolingual English one to a 40% multilingual, multicultural population. Not only have the students scored better over the years and achieved higher gains from year to year, more of the student population has consistently been assessed over time. This speaks to a high level of academic achievement across the school as a whole, and particularly high for traditionally lower scoring students--LEP and students receiving special services such as Chapter 1.

These gains in achievement may be attributed to the rich language and content experiences that students engage in during Expeditionary Learning. It may be that this type of learning is particularly effective for students who are motivated to improve their circumstances, a common finding in studies of recent immigrant students,¹⁸ and who have a limited ability in English due to it being their new second language. A rich, collaborative exploration of content that is interesting to students because they can choose their paths of inquiry, and a collaboratively produced, high quality product of the learning may provide the kind of environment that aids motivated LEP students in learning both content and language with and from their peers.

The embeddedness of ELOB within the King School provides a consistency in conducting expeditions as a way of learning. The fact that both Houses team in the development and revision of expeditions, that all of the professional development opportunities for staff are tied into EL, and that there is sufficient support and teacher autonomy in creating the learning environment seem to facilitate the growth of this initiative. Given the performance indicators of its students over time, the implementation of ELOB at the King Middle School is succeeding well as a comprehensive school reform design.

¹⁸ See John Ogbu, "Variability in Minority School Performance: A Problem in Search of an Explanation," Margaret Gibson, "The School Performance of Immigrant Minorities: A Comparative View" and Marcelo Suarez-Orozco, "'Becoming Somebody': Central American Immigrants in U.S. Inner-City Schools" in Evelyn Jacob & Cathie Jordan (Eds.), *Minority Education: Anthropological Perspectives*, (Norwood, NJ: Ablex, 1993) for a discussion of immigrant motivations and school success.

RAFAEL HERNANDEZ SCHOOL, BOSTON MASSACHUSETTS

BACKGROUND

The Rafael Hernandez School is a small pre-K--8 two-way Spanish bilingual school in Boston. It is a district-wide magnet school, which means that students can choose to enroll in the Hernandez from any of the designated areas that make up the Boston Public School Zones. Non-magnet schools enroll students from within their Zone, limiting the choice available to parents, but keeping the schools relatively comparable in racial and ethnic composition.¹⁹ The Hernandez, as a two-way bilingual school, differs in enrollment from other Boston schools in that it generally enrolls a higher number of Hispanic students and does not enroll Asian or other language students. The proportion of White and Black students is slightly less than other area schools, but the combined Black and Hispanic population is comparable to other BPS schools' total population of color.²⁰

Over the period of time we are considering in this report (1992-1999), the approximate average population characteristics of the Raphael Hernandez indicate that it is a relatively small school, enrolling about 385 students per year in grades pre-K--8. There are two homerooms of about 20 to 25 students each in each grade, making a total of approximately 40 students per grade in Grades 1-6. One homeroom serves students whose primary language is Spanish, the other English. Both of the grade level teachers are bilingual but one teacher teaches in Spanish, the other in English, with students switching between the languages. Curriculum and subject areas are coordinated by the team of teachers in each grade for first and second language instruction in both Spanish and English.

The racial distribution of the school is approximately 25% Black, 15% White, and 60% Hispanic. The percentage of bilingual/LEP students averages about 45%, which is significantly higher than the District average of approximately 20%.²¹ The following

¹⁹ School assignment that controls for race has been in effect since 1974 when it was mandated by Court Order. It will be abolished as of the 2000-2001 school year.

²⁰ Population of color refers to Black, Hispanic, Asian and Native Americans. Native Americans are minimally represented in the Boston Public Schools.

²¹ The Hernandez considers all of its students to be bilingual, since instruction is carried out in both Spanish and English. The District, however, denotes bilingual students as those who receive some portion of their

chart provides a comparative view of the Hernandez population compared to the District percentages for K-12 enrollment (which includes three exam schools encompassing grades 7-12). The numbers are an average of the years we are considering in this report, 1992-1999.

	Hernandez	District
<i>Enrollment</i>	385	
<i>Bilingual/LEP</i>	45%	20%
<i>Black</i>	25%	49%
<i>White</i>	15%	17%
<i>Hispanic</i>	60%	25%
<i>Other</i>	0%	9%

At the 7th and 8th grade level, the total class population drops by nearly 40% at the Hernandez. In the 6th grade in BPS schools, students may take an exam to qualify for one of the three exam schools in the city. Non-exam middle schools typically lose students at this juncture, either to the exam schools or to other options such as private or parochial schools. The demographic profile of the Hernandez changes in the 7th Grade over the years we are examining in the following way. The percentage of White students drops from 15-20% to 0-4%. At the same time there is an increase of Hispanic students from approximately 55% to 75%. Black student attendance at the Hernandez does not vary significantly.²²

classes in a native language other than English. For this report, the term "bilingual" is used in the District sense and refers to LEP students

²² Since Court ordered busing of Boston Public Schools began in 1974, there has been continued White flight from the city's schools while at the same time large numbers of immigrants have entered the schools primarily from Puerto Rico, the Dominican Republic, Guatemala, El Salvador, Columbia, Haiti, Jamaica, Somalia, Ethiopia, Vietnam, China, Cambodia, Laos, and Eastern Europe. Over the years there has been a steady return of White students in the elementary schools, where academically oriented students are offered placement in Advanced Work Classes available in some schools starting in the 4th grade. Parents with academic aspirations for their children who can afford non-public alternatives appear to opt for other choices if their children do not make it into one of the city's exam schools. Students of color may also be eligible for the METCO program, an open enrollment of inner-city minority students within accepting suburban school districts surrounding Boston. The overall effect is one of dramatically reducing the academically oriented students, particularly in upper middle school grades and high school within the magnet and non-exam district schools. The scores on standardized tests reflect this demographic change, as we will see below.

ELOB Instructional Program

The Rafael Hernandez School phased in ELOB implementation over two school years: 1993-94 and 1994-95. It began implementation in the middle school grades (6-8) and subsequently extended it to the elementary school. At active implementation, all students explore four different themes or topics for approximately 10 weeks, or one expedition per quarter. Expeditionary learning time at the Hernandez generally takes place daily in all grades during an afternoon block. The expeditions are more normally carried out in English, but some grade levels combine languages in exhibitions. One team has an entire expedition in Spanish as well as ones in English. Efforts are being made to include more Spanish in most of the expeditions, as well as increasing the amount of math and science. The expeditions are planned by team members, who keep students for two years in a looping cycle, and are interdisciplinary in nature.²³ ELOB Summer Institutes as well as three days of common planning time during the year are used to refine and elaborate expeditions based on a review of student work, and City and State mandated curriculum standards. The Hernandez teachers collect portfolios of student work to document their expeditionary learning.

Since the 1996-97 school year, Boston Public Schools have been engaged in systemic, Standards-based reform. The entire District was divided into four cohorts, each of which entered into the reform process in successive years. The Rafael Hernandez was a member of the first cohort, and so began systemic reform in the 1996-97 school year. They had previously, and independently, chosen ELOB as their model for reform in 1992 but, as of 1996, they were also required to address specific mandates of the District. They were required to choose an instructional focus, which the staff determined was literacy, and to adopt a literacy model throughout the school.

Initially, the Hernandez chose their ongoing literacy program, Voices of Love and Freedom, which promoted reading multicultural literature from a variety of authors which enhanced the values of love (trust, help, sharing) and freedom (self-respect, self-confidence, self-understanding and hope). Beginning with the 1997-98 school year, however, the Hernandez began implementation of an Early Literacy Instruction model

²³ Looping occurs within the following grade levels: K-1, 2-3, 4-5. The middle school teachers (6-8) meet regularly as a team but do not loop their students.

developed and supported with professional development training through Lesley College, and over the past year, 1999-2000, expanded it to the upper grades. This newly adopted literacy model is a balanced literacy approach which works on decoding skills, vocabulary, reading strategy instruction and higher level comprehension. With this literacy model, teachers are encouraged to use appropriate level reading texts with small groups of students at homogeneous reading levels. It has been a challenge to find materials for expeditions at the various reading levels of the students and there has, thus, been an increasing tendency for teacher teams to separate expeditions from the language arts block of instruction, reducing the scope of expeditions in some cases.

Additionally, teachers at the Hernandez perceive that the detailed Curriculum Guidelines produced by the BPS Central office which mandate performance evaluations in each of the traditional subject areas present a challenge for planning lengthy, project-based interdisciplinary expeditions. Teachers at the Hernandez remain committed to ELOB and are finding it necessary to revise their curriculum to realign it with City and State standards. Currently, Math is not as well integrated into the expeditions as previously because staff feel they don't have a choice in planning the curriculum. They are aligning their instruction with cumulative chapter tests that prepare students for the Stanford 9 Math tests, which are required for grade promotion under a new District policy. For similar reasons in the middle school grades the expeditions have been reduced in scope. The 7th and 8th grade teachers are presenting expeditions in which the guiding questions for their units are determined by what is required by the Standards rather than by student inquiry or interest. Although this is not a prerequisite for Expeditionary Learning, some of the teachers perceive that teacher-formulated questions stray from their view of a "true" expedition. Teachers are, nevertheless, maintaining ELOB principles with field work and service projects and revising their expeditions to better align them with new curriculum Standards..

The elementary grades continue to organize a portion of their instruction around expeditions, but they too may differ in terms of how much of the work is student directed and how much is guided by teacher formulated key questions. The variation depends on the teachers, their perception of what students can do independently and how experienced

the teachers are (some are recent hires) in the development and implementation of expeditions.

Data Sources

The data we used to compile this report were the following:

1. Qualitative data gathered during a visit to the Hernandez in Spring 2000.
2. Quantitative data gathered by the school for tracking student achievement.
3. Quantitative data gathered and analyzed by the District comprising various standardized assessments administered in all BPS schools.
4. Demographic data reported by the school, the District, and the State Department of Education.

We report on test data at three points in time.

- *Time 0* presents data from the 1992-93 school year, which is pre-ELOB implementation.²⁴
- *Time 1* data come from the 1994-95 school year, the first year of whole school implementation of ELOB as well as the last year that BPS used the Metropolitan Achievement Tests (MAT) as the standardized measure all schools required of their students.
- *Time 2* data come from the 1995-96 school year, the first year BPS switched to Stanford 9 (SAT 9) as the mandated standardized test used in all grades to assess student achievement.²⁵ It is not possible to compare the MAT scores with the SAT 9, so 1995-96 sets a new benchmark for comparative data with the District and a matched pre-K-8 comparison school.
- *Time 3* data, the final years reported, come from the most recent school years in which any given score for the Stanford 9 was reported, mostly 1997-98 or 1998-99. They allow for a comparison from early ELOB implementation to later years.

DRP (Degrees of Reading Power) test scores are also provided when available.

This test was administered, District-wide, in grades 5, 8 and 12. The test is constructed of

²⁴ The Hernandez staff voted to adopt ELOB in 1992.

reading passages that are ordered by difficulty. Students demonstrate reading comprehension by filling in blanks with the appropriate word from among a choice of items. A score is tabulated based on which passages the student can read with adequate comprehension. The District sets a minimal passing score for the DRP in grades 5, 8 and 12 and passing the test at those levels was linked to promotion criteria from grade to grade and in order to graduate from high school. Students are given a number score, not a comparative score such as a percentile rank, which is how the Metropolitan Achievement Tests results are reported.

In addition to the standardized achievement tests mentioned above, we will present data from the newly implemented MCAS exam--Massachusetts Comprehensive Assessment System. The MCAS was developed by an independent testing company (the same company that developed the MEA for Maine) following the State Department of Education Curriculum Frameworks. The MCAS exam was first administered in Spring 1998 to all 4th, 8th and 10th graders in the Boston Public Schools who were eligible. It has been repeated in Spring 1999 and 2000. The results for 2000 were not available at the writing of this report.

The MCAS exams test English Language Arts (ELA), Math, Social Studies and Science and Technology. It is a high stakes exam linked to grade promotion, and no student who has not passed the ELA and Math portions of the 10th grade MCAS will graduate from high school beginning in the year 2003.

WHAT THE TESTS SHOW

The various standardized test scores will be analyzed to show how the Hernandez fared prior to and over the time of ELOB implementation. We will compare the Hernandez scores against the District as a whole. In addition, we will compare the Hernandez at Times 2 and 3 as well as on the MCAS with a matched school in the District that shares similar qualities. The Garrison School is the only other two-way Spanish bilingual school in the district. It is pre-K--8 as is the Hernandez. Its demographics are roughly comparable, differing in the following ways: It is slightly larger than the

²⁵ The Stanford 9 was substituted for the MAT by the new Superintendent. Because it is a more rigorous test of student achievement and because it reports scores in Performance Levels rather than as a composite score, it allows schools to more effectively track student progress over time.

Hernandez and has a higher percentage of Black students and a lower percentage of White students than the Hernandez. It also ranks slightly higher in the percentage of students who qualify for free or reduced lunch. The following chart, which provides an average enrollment over the years considered here (1993-99) displays these demographic differences.

	<i>Hernandez</i>	<i>Garrison</i>
<i>Enrollment</i>	385	450
<i>Bilingual</i>	45%	43%
<i>Black</i>	25%	45%
<i>White</i>	15%	1%
<i>Hispanic</i>	60%	50%
<i>Other</i>	0%	0%

We also compare the Hernandez to State-wide performance on the MCAS, although these comparisons inevitably yield a huge disparity in performance due to the demographic and economic factors operating in the State's largest metropolitan area compared to the rest of the State's school districts, which are more middle class, White, and suburban, on the whole.

One final comment regarding standardized test data is necessary. Like most poor, predominantly minority, urban districts, Boston test data show a significant achievement gap between White students and students of color, particularly Black, Hispanic and Native American students. Black and Hispanic students are similarly represented at Level 1, the lowest level of performance, and at Levels 3&4 (the highest performance levels) in ELA and Math at all grades taking the MCAS in 1998 and 1999. They perform the poorest of any racial group in the city. Additionally, when Stanford 9 and MCAS scores were compared by educational program (Regular Education, Special Education or Bilingual Education), Bilingual Education students performed the lowest in 8 out of 13 District-wide cohort analyses (Math for grades 4-11, Reading/ELA for grades 5-11).²⁶ These District-wide trends put the data we will examine for the Hernandez and the Garrison in context. Average achievement vis-à-vis District levels actually represents much higher

²⁶ 1997-98 Stanford 9 data were analyzed for these data and Reading was not administered in Grade 3 of that year. Special Education rated lowest in 6 out of 13 cohort analysis, although Bilingual and Sp.Ed. students were nearly equal in the 7th grade cohort for Reading.

performance on the part of Hernandez and Garrison students, who are over-represented in the category of Hispanic bilingual/LEP students compared to the rest of the District.

Time 0 (1993-1994) Hernandez Metropolitan Achievement Test Scores

The Metropolitan Achievement Test is a nationally normed standardized test of Reading and Math. The format of the test is multiple choice and results are reported as percentile scores averaged by grade level. We report three-year cumulative MAT scores for Time 0 (1992, 1993, 1994) since cumulative scores lessen the variation that can appear from class to class. The MAT version used in the years we are considering here was Edition 6.

The Rafael Hernandez School performs at higher than the national average in Reading on the Metropolitan Achievement Tests in grades 1, 2, 3, 4, and 6. There is a tendency to score at a lower percentile as the grade level increases, however. Comparing Hernandez Reading scores to the rest of the District in each grade, the Hernandez performs higher than the District average in grades 1, 3, and 6 only. We see an expected dip in Hernandez scores in the 7th and 8th grades as students qualifying for the exam schools leave.

Reading Scores

<i>Grade</i>	<i>Hernandez</i>	<i>BPS</i>
1	70	63
2	55	57
3	57	50
4	53	55
5	43	50
6	53	49
7	36	47
8	39	48

Students in Boston generally scored higher on the Math portion of the Metropolitan Achievement Tests than Reading for the three years that comprise these cumulative average percentile scores (1992, 1993 and 1994). All grades scored above the national average (50%) for the District, and Hernandez 1st, 2nd and 3rd graders scored well above the median percentile, scoring higher than the District average for those grades. In

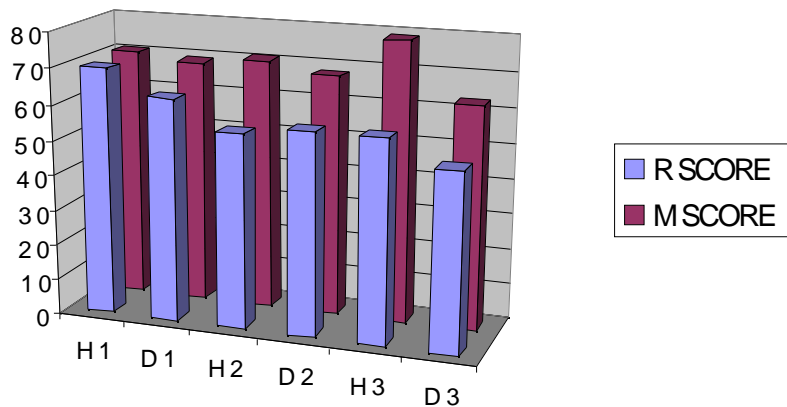
the 4th grade, however, the Hernandez students score slightly lower than their District counterparts and by the 5th grade they fall below the national average as well as below their District counterparts. Again there is a significant dip in scores in the 7th and 8th grades.

Math Scores

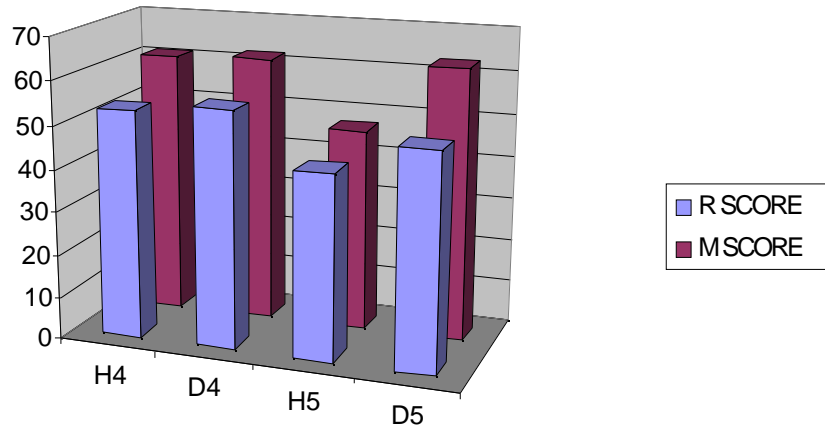
<i>Grade</i>	<i>Hernandez</i>	<i>BPS</i>
1	71	69
2	71	68
3	79	63
4	61	62
5	47	63
6	43	57
7	32	52
8	29	50

The following charts graphically represent the differences between Reading and Math Scores with each double column representing one grade for either the Hernandez (H) or the District (D). The first charts show elementary grade comparisons with Red and Blue columns designating Reading and Math scores respectively. The third shows Middle School comparisons.

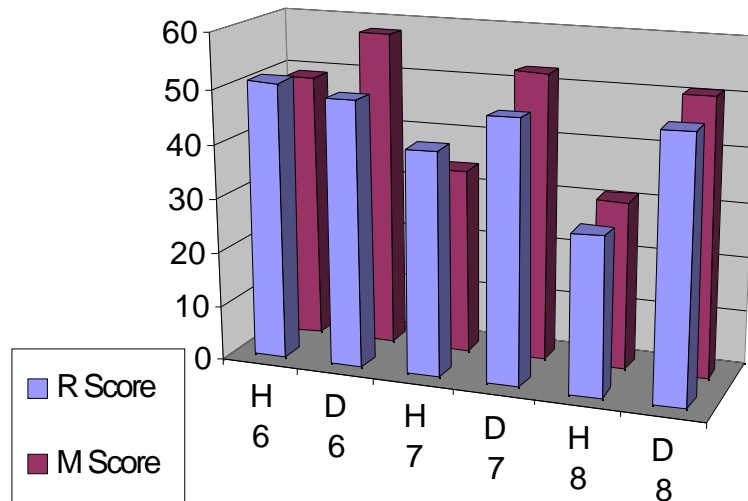
Grades 1 2 3 MAT Cumulative Scores
1993-94



Grades 4 & 5 MAT Cumulative Scores
1993-94



Middle School MAT Cumulative Scores 1993-94



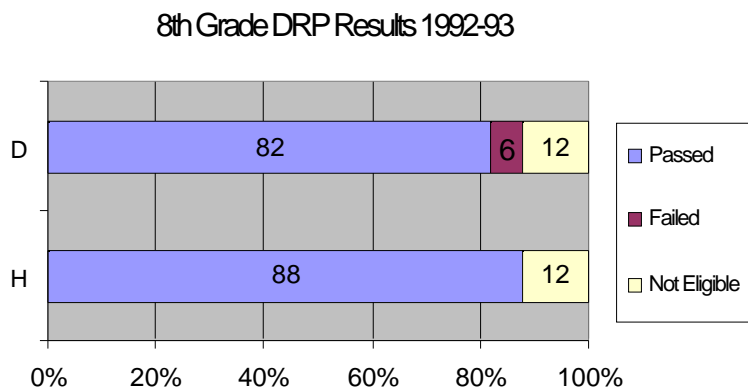
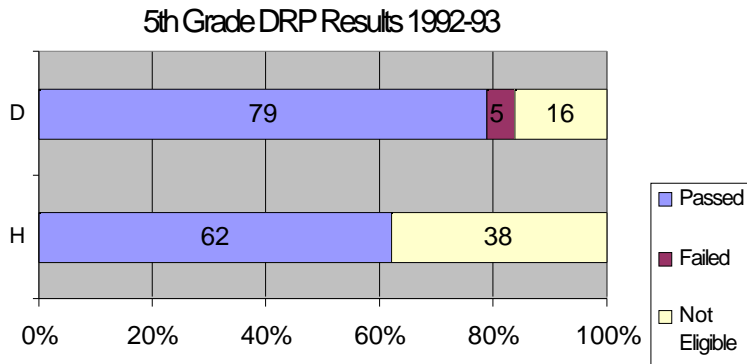
Time 0 DRP Scores

DRP scores for 1993 confirm the comparisons between the Hernandez and the District on the MAT for Grade 5, but not for Grade 8 (the test was administered annually in grades 5, 8 and 12). Hernandez 5th grade students scored, on average, in the 43rd percentile in 1993 on the MAT Reading test, while the District average for 5th graders was at the 50th percentile. Similarly, only 62% of Hernandez 5th graders passed the DRP, while 79% of the District's 5th graders passed.

In the 8th Grade, Hernandez students scored at the 39th percentile compared to the District average at the 48th percentile on the MAT. However, a larger percentage of Hernandez students passed the 8th Grade DRP than 8th graders in the District as a whole-- 88% vs. 82%.²⁷

The following charts illustrate the DRP comparisons between the District and the Hernandez for the 1992-93 school year.

²⁷ "Not eligible" refers to students who were exempted from the exam due to limited English proficiency or other learning or physical disability.



Time 1 (1994-95) MAT and DRP Scores

At Time 1, the first year of whole school ELOB implementation, the Hernandez shows improvement in its MAT scores when compared to the District. Different from the Time 0 cumulative score comparisons (1992, 1993, 1994), in which the Hernandez only scored higher than the District in the first three grades for Reading and Math, at the end of the 1994-95 school year, it outscored or scored the same as the District average in 5 out of

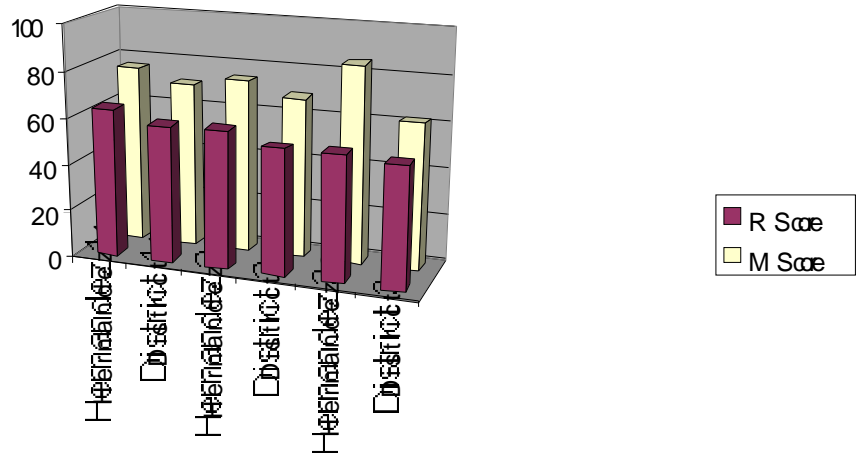
8 grades in Reading and 6 out of 8 grades in Math. Again, there is a large drop in Hernandez scores from the 7th grade onward when the population of the school changes significantly. Given that the Hispanic population, and particularly Hispanic students in bilingual education programs, score lowest on standardized measures at all grade levels, the gains demonstrated by the Hernandez after ELOB implementation are quite significant.

The following table displays the percentile scores for Reading and Math on the 1994-95 Metropolitan Achievement Test for grades 1-8 comparing the Hernandez and the District averages.

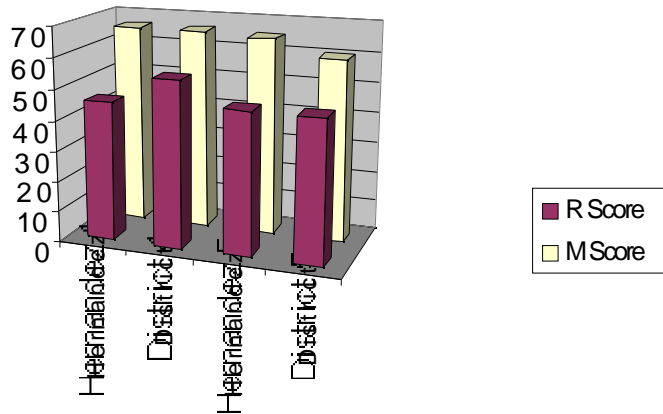
1994-95 MAT Scores		
	Reading	Math
Hernandez 1	64	77
District 1	59	71
Hernandez 2	59	75
District 2	54	69
Hernandez 3	54	85
District 3	52	63
Hernandez 4	46	66
District 4	55	66
Hernandez 5	47	65
District 5	47	60
Hernandez 6	62	74
District 6	48	57
Hernandez 7	22	28
District 7	41	39
Hernandez 8	29	26
District 8	39	43

This relationship is presented graphically below. Contiguous double columns represent the Hernandez and the District by grade level. The Red column indicates Reading MAT percentiles and the Yellow column represents Math. The first two graphs compare the elementary grades; the third compares middle school grades.

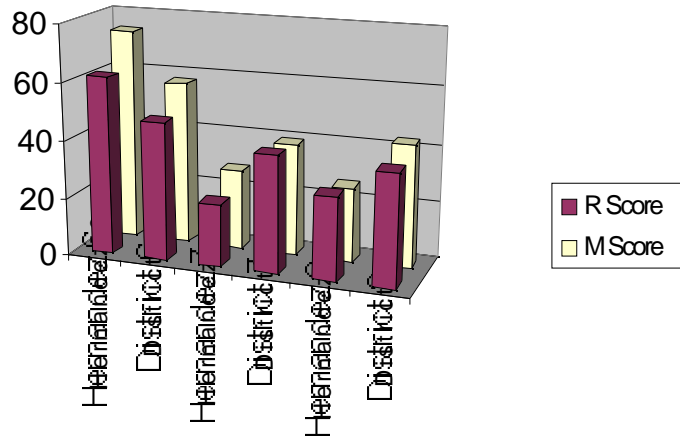
Grades 1 2 3 MAT Scores 1994-95



Grades 4 & 5 MAT Scores 1994-95

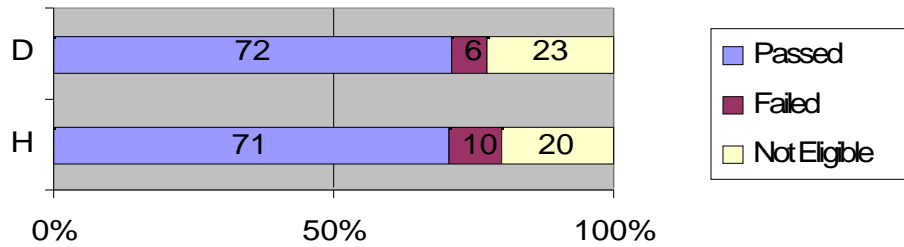


Middle School MAT Scores 1994-95
Hernandez and District

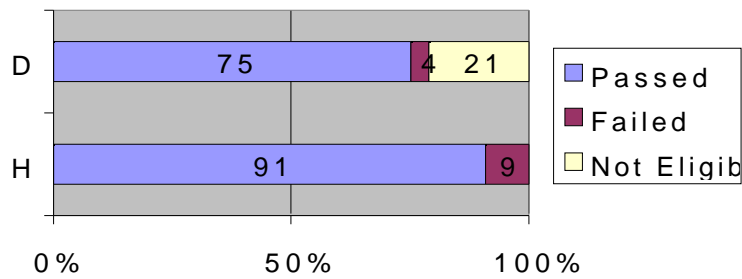


DRP scores match the MAT comparison between the Hernandez and the District in the 5th grade in 1994-95. Students scored the same in both groups on the MAT and the percent of students who passed the DRP in the 5th Grade in 1995 were very similar. 71% of Hernandez 5th graders passed compared to 72% District-wide. In the 8th grade, however, the DRP results differ greatly from the MAT Reading scores. The Hernandez scored 10 percentiles lower than the District in 8th Grade Reading on the MAT (29th compared to 39th). Yet, a significantly greater number of Hernandez students passed the DRP when compared to District-wide 8th graders (91% compared to 75%). The following bar graphs illustrate this comparison of DRP performance for the Hernandez and the District in grades 5 and 8. Note also that the Hernandez has 0 students not eligible for the test because of LEP or special needs compared to 21% in the District.

5th Grade DRP Results 1994-95



8th Grade DRP Results 1994-95



To summarize, between Time 0 (before ELOB implementation) and Time 1 (first year of whole school ELOB implementation) we see improvement in the Hernandez scores on the MAT and the DRP with respect to the District. This improvement is not in every grade, but more grades outscore the District average at active ELOB implementation than prior to it.

Time 2 (1995-96) Stanford 9 Achievement Tests

This year's data represent a new benchmark for standardized test scores in Boston since the system switched from the Metropolitan Achievement Tests to the Stanford 9 in an attempt to raise the standards and expectations for performance in BPS classrooms. It also represents the second year of active ELOB implementation at the Hernandez. BPS opted for the Stanford 9 as a replacement for the MAT6 for several reasons. Stanford 9 content reflected the State's Curriculum Frameworks and the newly emerging Citywide Learning Standards better than the MAT6. The multiple choice questions were constructed to tap higher order thinking skills in students more than other tests then available. And, finally, the results are reported in Performance Levels as well as in average scaled scores for each grade and subject. Performance Levels are determined by educational experts nationwide and reflect high expectations for student performance. This type of reporting interprets the results into a form that communicates what a child can do in a subject area, consequently allowing schools to maintain more accountable records of student progress as well as to set benchmarks for improving performance which can be more meaningful than those based on average scores alone.

The four Performance Levels assessed by the Stanford 9 for grades 1-8 are the following:

Level 1 - Little or no mastery of fundamental knowledge and skills;

Level 2 - Partial mastery of knowledge and skills that are fundamental for satisfactory work;

Level 3 - Solid academic performance indicating that students are prepared for the next grade level; and

Level 4 - Superior performance beyond grade-level mastery.

The determination of which Level a student scores at is based on the number of correct answers. The range of possible scores within each level varies considerably. For example, there is a much greater range of possible scores at Level 2 than at any other Performance Level.

The SAT9 was given in Grades 3, 5, 6, 7, 9 and 11 in Spring 1996. Grades 1 and 2 were excluded because multiple choice questions are not the most appropriate assessment

tool for young children. Grades 4, 8 and 10 were also excluded because those students took the Massachusetts Educational Assessment Program (MEAP), the statewide educational assessment test.

For the 1995-96 Stanford 9 results, we present a three-way comparison. We compare the Hernandez with the District as a whole as well as with the matched sample Middle School, the Garrison. The comparison bar graphs below show the relative distribution across the four Performance Levels for each grade in which the Stanford 9 was administered that year. (Fewer students at Level 1 shows higher academic achievement. More students at Levels 3&4 also indicates higher academic achievement.) We see from the 1995-96 data that the Hernandez has slightly fewer students scoring at Level 1 than the District for Grade 5 in Reading (-2%) and scores only 1% higher than the District at Grade 3. For these same grades (3 and 5) the Hernandez has more students scoring at Levels 3&4 combined, indicating solid or exceptional performance, 5% and 11% respectively. It greatly outscores the District on Level 4 alone for these grades, 17% in Grade 3 and 15% in Grade 5.

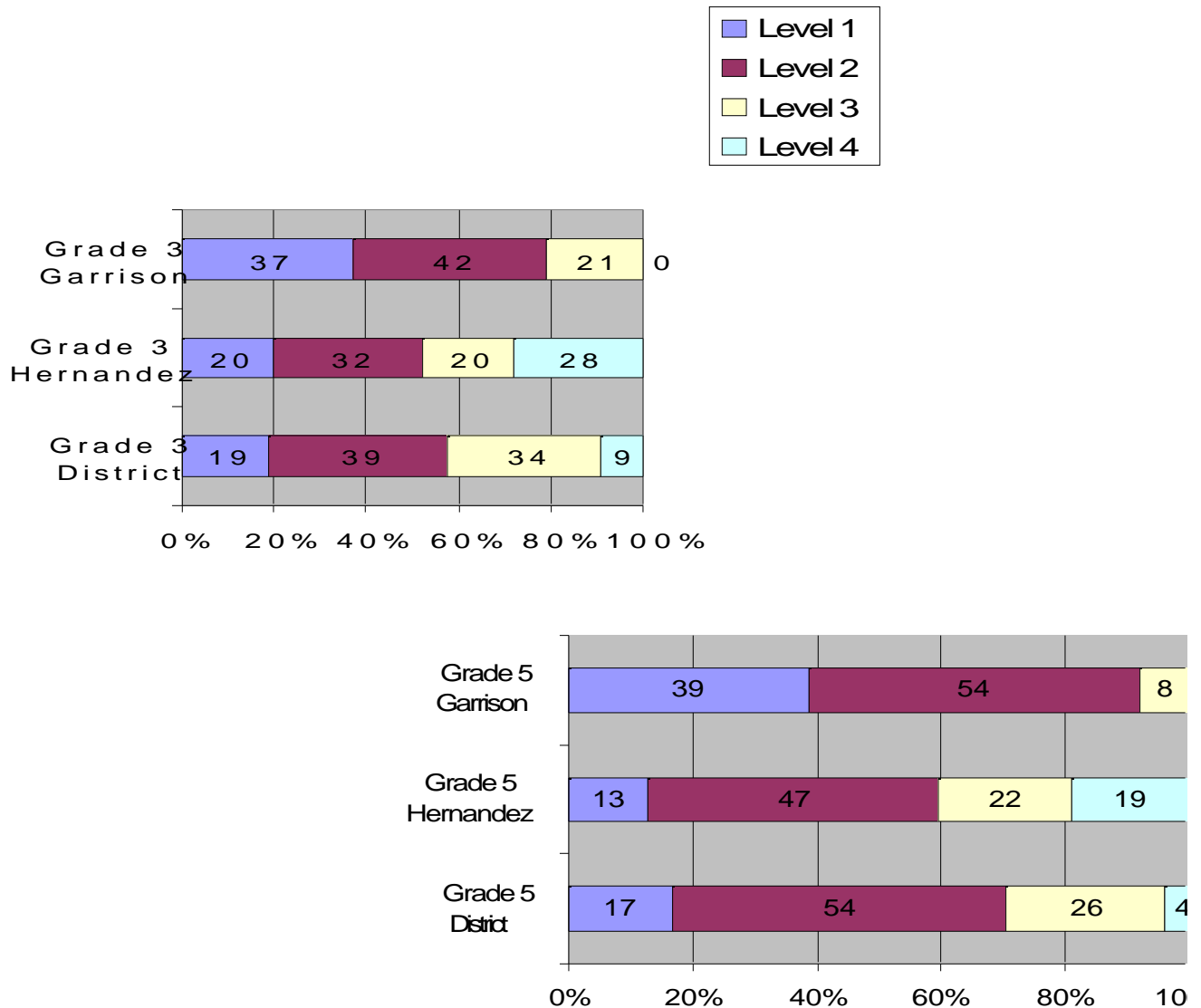
Reading scores for the Hernandez were slightly lower than the District in grades 6 and 7. There were 12% more students scoring at Level 1 at the Hernandez in Grade 6 and 24% more at Level 1 in Grade 7 compared to the District as a whole. In Grade 6 the combined percentages of students at Levels 3&4 was only 1% lower for the Hernandez than the District, but in Grade 7 that difference dipped by 10% for the Hernandez. In sum, elementary grade students at the Hernandez perform better than the District average in Reading but middle school grade students perform worse than the rest of the District on average.

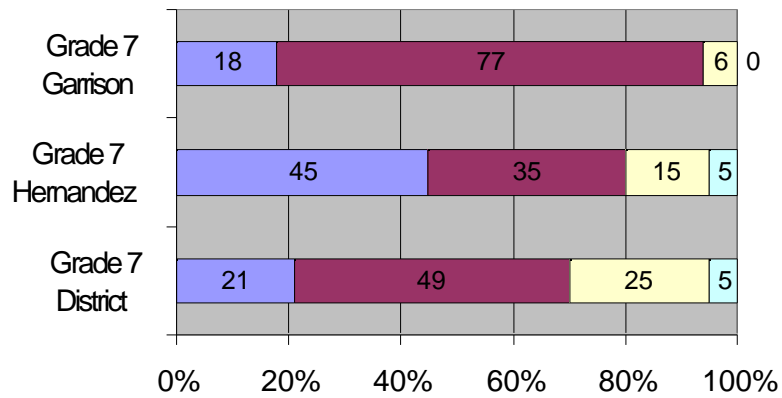
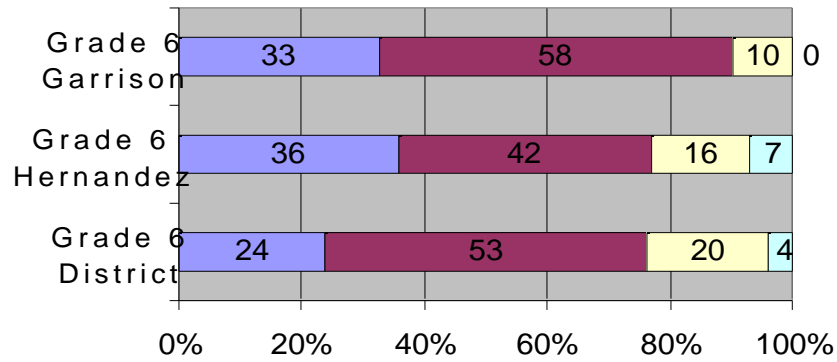
The Hernandez displays a far greater achievement level at all levels in Reading when we compare it to the Garrison, a similarly populated pre-K-8 two way Spanish bilingual school. The Hernandez has 17% fewer students scoring at Level 1 in Grade 3 compared to the Garrison and 26% fewer at Level 1 in Grade 5. Hernandez 6th and 7th graders outscore the comparable Garrison students in combined scores for Levels 3&4 by a large margin, 48% compared to 21% in Grade 3 and 41% compared to 8% in Grade 5. There are no students scoring at Level 4 in either the 3rd or the 5th grade at the Garrison,

while the Hernandez has 28% and 19% of its 3rd and 5th graders respectively scoring at the highest level. 6th and 7th graders at the Hernandez score at a slightly higher percentage in Level 1 compared the Garrison, however--36% at the Hernandez in Grade 6 compared to 33% at the Garrison; 45% at the Hernandez in Grade 7 compared to 18% at the Garrison.

The bar graphs below show the relative proportions of percents of students scoring at each of the Performance Levels in Reading for the Garrison, the Hernandez and the District, and graphically present the summary of the data reported above.

READING 1995-96 STANFORD 9 PERFORMANCE LEVELS





Math scores on the 1995-96 Stanford 9 show similar relationships between the Hernandez, the District and the Garrison. The Hernandez moderately outperforms the District in Grades 3 and 5, when comparing high scoring students, and greatly outperforms the Garrison at these grades. Level 1 students are comparable at the Hernandez and the Garrison and greater than the District in Grade 3, but in Grade 5 the Hernandez has significantly fewer Level 1 students than either the District or the Garrison. The following table provides the comparative percentages of students in each grade scoring at Level 1, Level 3 and Level 4.

	District Grade 3	Hernandez Grade 3	Garrison Grade 3
Level 1	21	39	40
Level 3	26	19	10
Level 4	7	19	0

	Grade 5	Grade 5	Grade 5
Level 1	36	18	67
Level 3	20	33	4
Level 4	6	12	0

The middle school grades scores in Math show a trend similar to that of Reading. A higher percentage of Hernandez 6th and 7th graders score at Level 1 than in the District as a whole, 68% at the Hernandez vs. 53% in the District in Grade 6 and 81% at the Hernandez vs. 58% in the District in Grade 7. The Hernandez students do not score nearly as well as the District as a whole in Grades 6 and 7 in the higher Performance Levels; only 10% of Hernandez 6th graders scored at Levels 3&4 while 18% of the District's 6th graders scored at the highest levels. In the 7th grade, where we noticed the greatest drop in student achievement on the Metropolitan Achievement Tests as well, the Hernandez only has 5% of its students at the Proficient or Advanced levels, while 17% of the District's 7th graders score at Levels 3&4 combined.

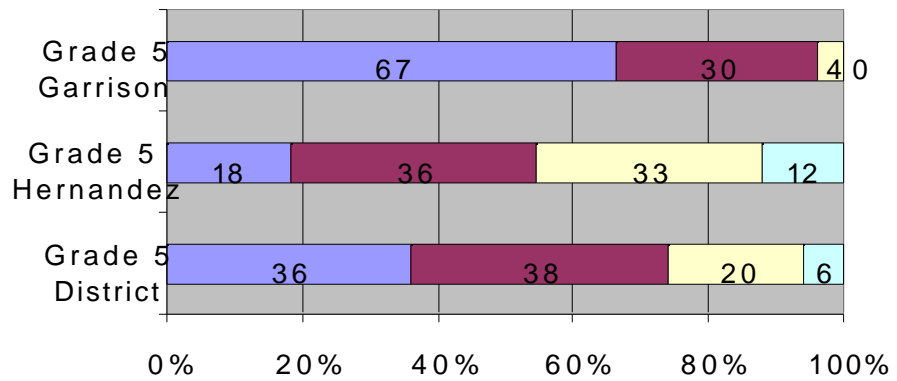
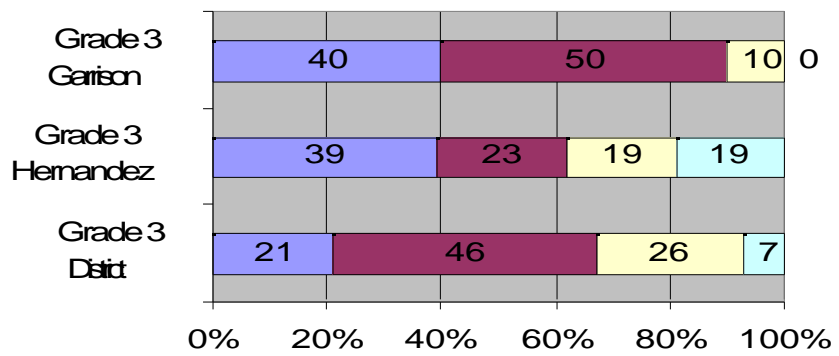
On the other hand, the Hernandez continues to score better in Math than the Garrison with fewer students scoring at Level 1 and more students scoring at Levels 3 and 4 at all grade levels. The following table indicates the compared percentages in Math Performance Levels for the District, Hernandez and the Garrison for Grades 6 and 7.

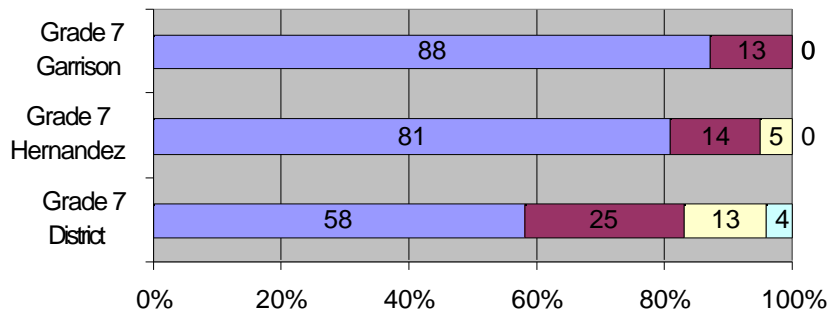
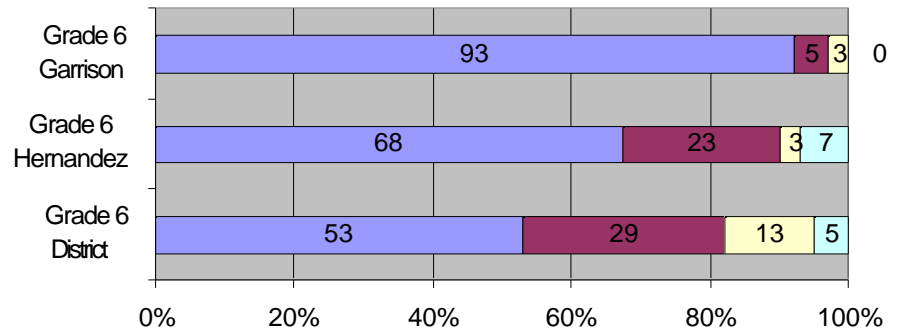
	District Grade 6	Hernandez Grade 6	Garrison Grade 6
Level 1	53	68	93
Level 2	29	23	5
Level 3	13	3	3
Level 4	5	7	0

	Grade 7	Grade 7	Grade 7
Level 1	58	81	88
Level 2	25	14	13
Level 3	13	5	0
Level 4	4	0	0

These relationships are graphically represented in bar graphs below. As above, Blue indicates Level 1, Red Level 2, Yellow Level 3 and Aqua Level 4.

MATH 1995-96 STANFORD 9 PERFORMANCE LEVELS





In sum, the Stanford 9 scores display similar trends when the Hernandez and the District are compared in Reading and Math. The Hernandez does slightly better than the District in the early grades in both Reading and Math. Whereas the District begins to outperform the Hernandez at Grade 7 on the MATs, they outperform the Hernandez beginning at Grade 6 on the 1995-96 Stanford 9 tests. At the same time, however, the matched school comparison, which we begin showing here in 1995-96²⁸, indicates that the Hernandez significantly outperforms the Garrison at all grade levels, but particularly in the early grades.

²⁸ The Garrison was not a two-way bilingual school and did not have middle school grades for the entire period we considered at Times 0 and 1 (1990-1994).

Time 3 (1997-1999) Stanford 9

The achievement levels of Hernandez students at both the elementary and middle school grades show less consistency when compared to the District and the Garrison in later years of ELOB implementation. On the whole, however, the Hernandez performs more inconsistently than it did by comparison in the early years of ELOB implementation. We will present a variety of Stanford 9 data breakdowns, summarize the comparisons, and conjecture about the causes of the unevenness or relatively poorer performance when it occurs.

Not all of the comparisons that follow are from the same years of testing. This is due to several factors in the way the tests were administered and reported. Not all grades were tested in both Reading and Math in all years. Some years students were given open-ended Reading and Math versions of the Stanford 9 instead of the multiple choice version. This was phased in gradually and, since we are reporting the early years of Stanford 9 testing results, the holes in scores for various grades and subjects are numerous. In addition, we have relied on secondary data to analyze these scores, which means that we are dependent on the reports issued by the Boston Public Schools. There have been changes in the way scores have been reported over the years of Stanford 9 testing, and not all breakdown categories are available in each year of the BPS reports. We will explain the data sources for each comparison we make below.

Median Percentile Stanford 9 Scores

We begin the Time 3 comparison with the widest lens on the Stanford 9 data: three-year median percentile scores for both Reading and Math covering the Spring testing in the years 1996-97-98. These scores apply to the multiple choice version of the Stanford 9 only, and are available for grades 4 through 8.

In general, the Hernandez continues to perform better, relative to the District in the elementary grades than in the middle school grades. While it never scores higher than the District on the three-year cumulative average median percentile scores for Reading, the gap is less at the 4th and 5th grade levels. On the other hand, as regards the comparison with the Garrison (the matched sample school from the District), the Hernandez performs less well in Reading, albeit slightly, in Grades 4 and 8. Compared to earlier years, the

Hernandez is not performing as well as it had, or, alternatively, the Garrison shows improvement at a far greater rate in Reading. The following table presents the Reading comparisons.

Stanford 9 Reading Cumulative Median Percentiles 1996-1999

Grade	<i>District</i>	<i>Hernandez</i>	<i>Garrison</i>
4	42	38	41
5	46	45	28
6	43	36	33
7	43	30	27
8	45	32	39

In Math, however, the Hernandez maintains its progress. Math three-year cumulative median percentile scores at the Hernandez are higher than the District's in the 4th and 5th grades, slightly lower in Grade 6 and somewhat lower in Grades 7 and 8. The greatest difference in scores occurs at the 7th grade, when the school population shifts significantly, but 8th grade scores begin to close this gap. We will see this trend in other data comparisons that follow. Compared to its matched school, the Hernandez outperforms the Garrison in Math by a significant amount in all but one grade--Grade 4. .

Stanford 9 Math Cumulative Median Percentiles 1996-1999

Grade	<i>District</i>	<i>Hernandez</i>	<i>Garrison</i>
4	47	51	52
5	47	60	26
6	44	42	27
7	43	33	19
8	47	42	17

An advantage of the Stanford 9 scoring is that percentages of students scoring at the various Performance Levels is reported. This allows a more detailed breakdown of student performance by grade and subject area as well as for more sensitive tracking of changes in student performance over time. Since the tests were not given in all grades each year we have data comparing the schools as follows:

- No Grade 3 Reading comparison since ST 9 was not administered in this grade after 1996, the first year.
- Grade 3 Math comparison from 1996 to 1999.
- Grade 4 Reading and Math comparisons from years 1997 and 1998 only since it was not administered in other years.
- Grades 5, 6, and 7 Reading and Math comparisons from 1996 to 1999.
- Grade 8 Reading comparisons from 1997 to 1999.
- Grade 8 Math comparisons from 1998 to 1999.

Stanford 9 Reading Performance Levels

We will begin with the analysis of change over time in Reading Performance Levels for the grades and years available. As with the earlier Stanford 9 data analysis for Time 2, we will chart the changes in the percentage of students scoring at Level 1 and at Levels 3&4 combined over time. Improved performance is measured by a reduction of students scoring at Level 1 over time, and/or an increase of students scoring at Levels 3&4 combined over time. All numbers refer to percentages of students, and since the numbers of students per grade at the Hernandez and the Garrison are small, the changes in percentage reflect small numbers of actual students scoring at different levels. In some cases we are seeing the difference of one or two students. Nevertheless, the changes over time provide some indication of the effectiveness of the educational program at these schools in meeting the demands of the Stanford 9 achievement tests.

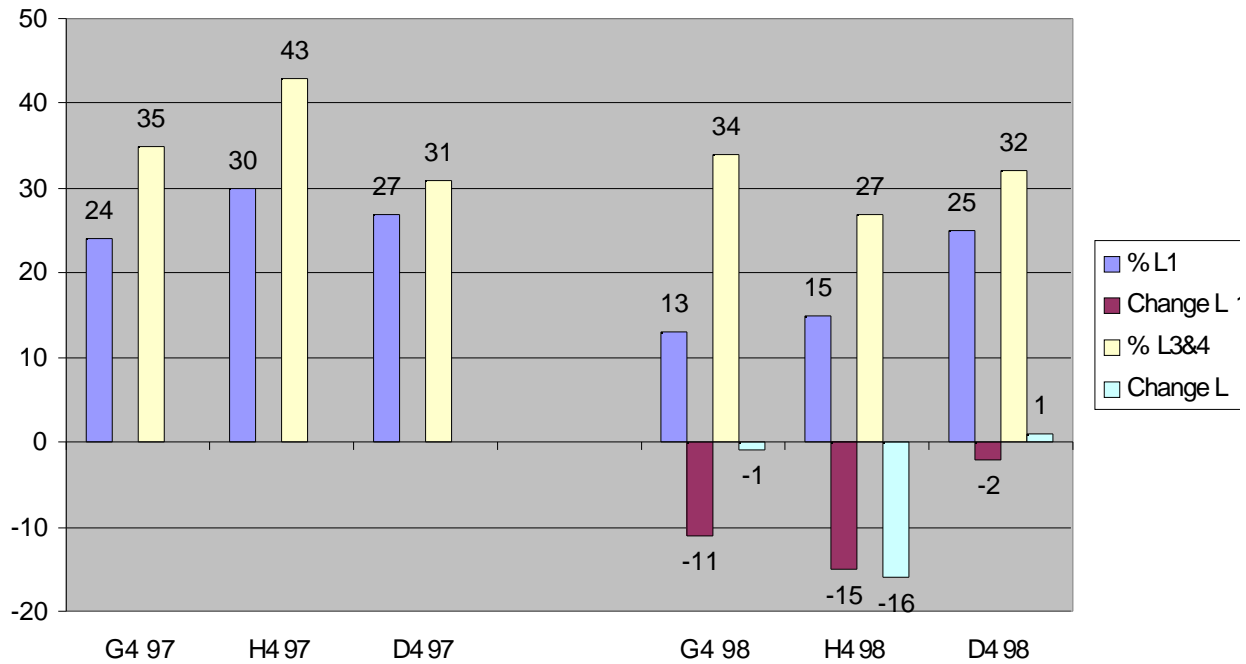
4th Grade Reading comparative data are available for only two contiguous years, 1996-1997 and 1997-1998. While the District scores stayed relatively stable, the Hernandez and the Garrison showed greater differences in 4th grade Reading Performance Levels. The Hernandez had the greatest decrease in Level 1 performance: -15%. The Garrison had -11% and the District only

-2%.²⁹ At combined Levels 3&4 we see the opposite trend. The Hernandez loses 16% in the high performance levels over time compared to 1% at the Garrison. The District increases its scores at these levels by 1% showing that there was greater improvement at the District-wide level than at either the Garrison or the Hernandez. The Hernandez reduced its numbers of failing students in Reading, but although it had the highest percentage of high performing students in 1997, it did not show improvement in increasing the number of proficient or advanced readers in 1998. In fact, it had fewer students in Levels 3&4 than either the District or the Garrison in 1998.

The following chart represents these data. The Blue bars represents the percentages of Level 1 readers in each comparison group for each of the two years compared. The Yellow bars indicate the percentages of students scoring in combined Levels 3&4 for each of the years compared. In addition, the Red and Aqua bars in the second year portion of the graph indicate the changes in Level 1 and Levels 3&4 respectively over the two year period. Thus, a larger negative Red bar indicates improvement in Level 1 readers while a taller positive Aqua bar indicates improvement in proficient and advanced level readers.

²⁹ N.B. The larger differences in the individual schools represent a very small number of students. The greater the number of students in any sample, the more likely the scores will show small differences over time, but on the other hand, these differences are more indicative of overall improvement or lack thereof than with smaller samples.

Grade 4 ST9 Reading Changes
1997 to 1998



In Grade 5 the Hernandez shows the poorest performance overall in Readings scores compared to the District and the Garrison, as well as to other grades in the school. Comparing the Stanford 9 Reading scores for Levels 1 and 3&4 combined from 1996 to 1999, we see an increase in Level 1 readers of 4% and a decrease of Level 3&4 readers of 14%.³⁰ For this same period, the District scores remained steady, showing neither increases or decreases at these Performance Levels. The Garrison, however, improved on both counts. There was a 3% decrease of Level 1 readers and an 8% increase of Levels 3&4 combined .

Grade 6 comparisons for the same time period show a return to the overall pattern we have previously seen. The Hernandez decreases Level 1 readers by 18%, the largest decrease in Level 1 readers of any grade for any comparison group during the time we are considering. Levels 3&4 readers also increase by 7%. The District shows a slightly

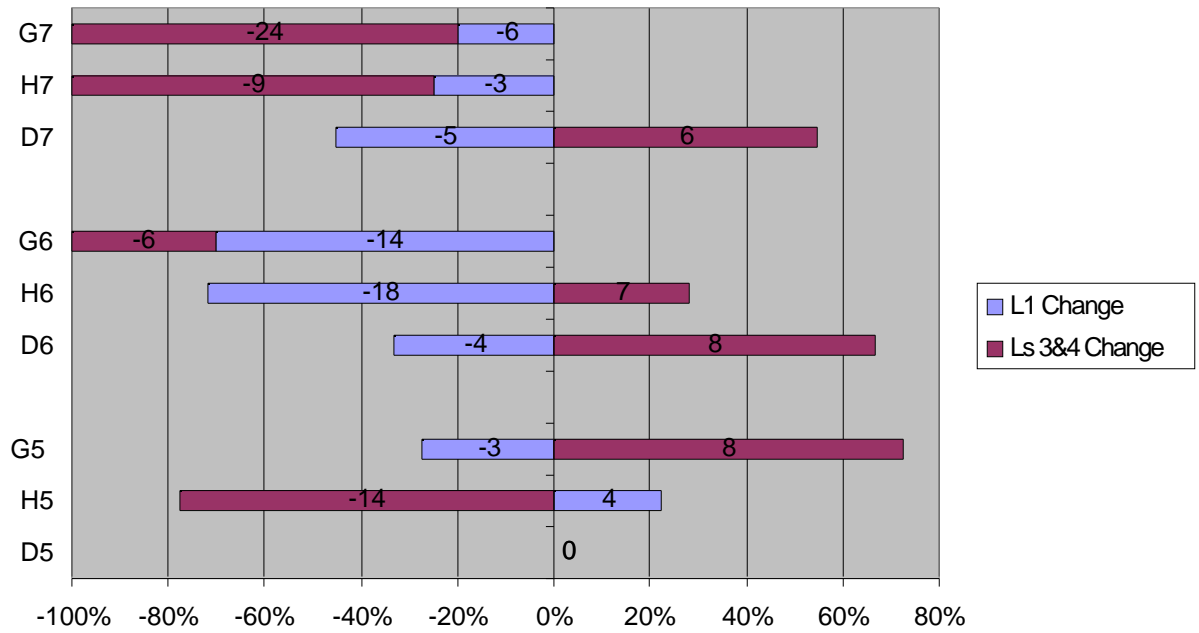
³⁰ This represents the comparison of the 1996 and 1999 scores. The interim years, 1997 and 1998 show even poorer performance at Grade 5.

higher gain in Levels 3&4--8%--but a lower decrease in Level 1 readers--4%. At the 6th Grade the Garrison improves in the area of Level 1 readers, decreasing by 14%, but it also decreases its Levels 3&4 readers by 6%.

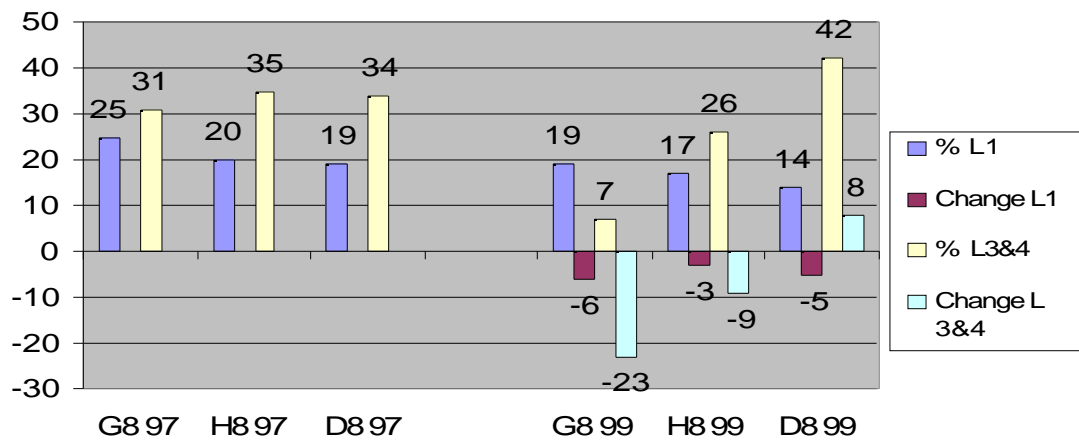
At Grade 7 all of the comparison groups improve on Level 1 Reading scores similarly. The Hernandez reduces its Level 1 readers by 3%, the District by 5% and the Garrison by 6%. Only the District increases its number of Level 3&4 readers however, by 6%. The Hernandez decreases highly proficient readers by 9% and the Garrison by 24%.

The following chart graphs the relationships of increased and decreased percentages for the comparison groups in Grades 5, 6 and 7. The Blue bar indicates Level 1 readers and so a negative number indicates improvement. The Red bar indicates Level 3&4 readers and a positive number indicates improvement.

Grades 5, 6, 7 ST 9 Reading Change 1996 to 1999



Grade 8 ST 9 Reading Changes 1997 to 1999



In sum, the Hernandez performs more poorly than the comparison groups in the 4th and 5th grades. It also shows little if any improvement in Grade 8, but performs better than

the Garrison in stemming the decrease of high performing readers. The Hernandez shows the greatest improvement among the comparison groups at the 6th grade and falls in between the two groups in Grade 7, scoring better than the Garrison, but showing less improvement than the District as a whole.

It is difficult to interpret these fluctuations in scores with any confidence. As mentioned above, large changes in percentages reflect small numbers of actual students in the school comparisons (not in the District where the N is far larger). In addition, the scores are not available for the same time period for each of the grades due to testing variations. Nevertheless, possible explanations for the variations we see from year to year may depend on individual classes, individual teams of teachers, more or less specific preparation for Stanford 9 multiple choice tests, or the instructional program adopted. We will return to these possibilities in our conclusion.

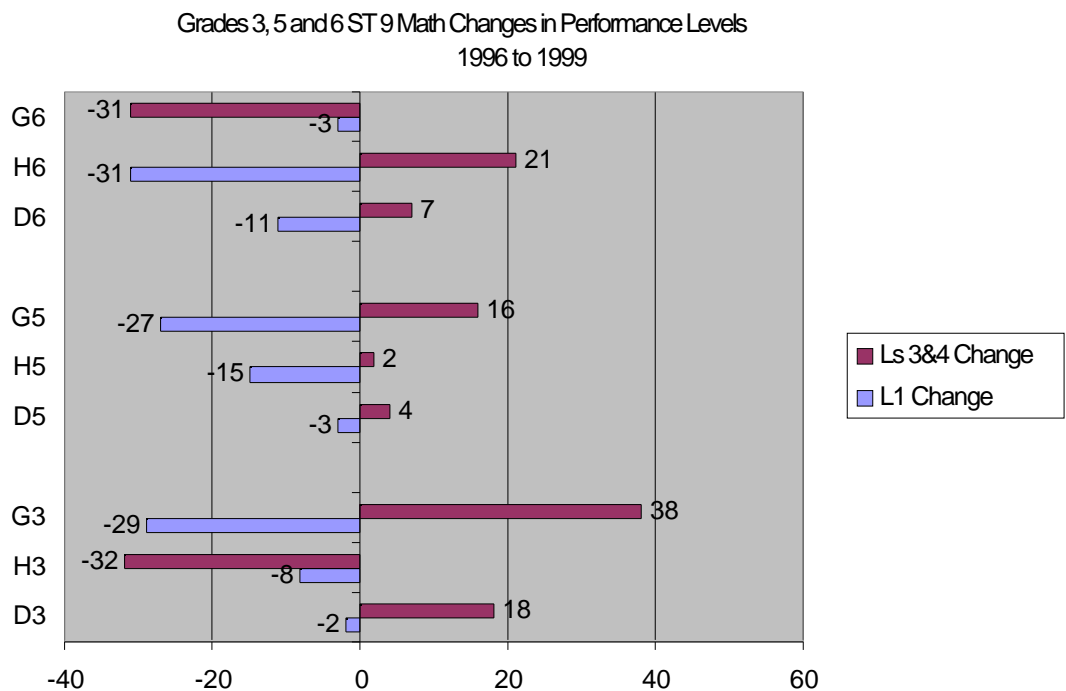
Stanford 9 Math Results

Stanford 9 Math Performance Level changes over time show more improvement for the Hernandez over the same period of time, 1996 to 1999. At Grade 3 the Hernandez shows the greatest decrease in Level 1 scorers compared to the District and the Garrison: -32% compared to -2% for the District and -29% for the Garrison. Similarly, at Grade 6 the Hernandez decreases its Level 1 students in Math by 31%, the same as the Garrison, but greater than the District's decrease of 11%. In the 5th Grade, however, the Hernandez increases its Level 1 students by 2% while the District and the Garrison both see decreases, i.e., improvements in their scores over the same period-- -3% and -27% respectively.

In Grades 3 and 5, the Hernandez decreases its percentage of high performing Math students, while both the District and the Garrison increase their percentages of these students. In the 6th grade, however, the Hernandez demonstrates the greatest achievement overall with a large decrease of Level 1 students (-31%) and the largest increase of Levels 3&4 students (21%) compared to the District and the Garrison. We see the poorest performance in Math, as we did with Reading, in the 5th grade students for these years at the Hernandez, with the only increase in Level 1 percentages of the three comparison

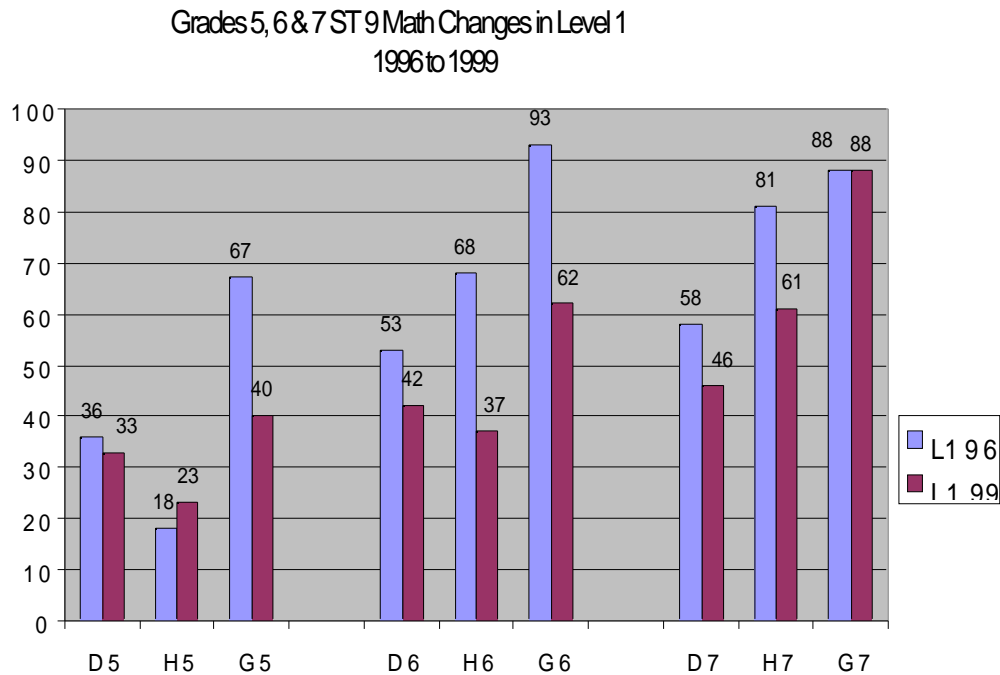
groups and a large decrease in Levels 3&4 percentages, again, the only one of the three groups for the 5th grade.

The following graph illustrates these numbers. For each comparison group the Blue part of the bar indicates changes in Level 1 percentages and the Red part of the bar indicates changes in Levels 3&4 percentages between 1996 and 1999. Improvement is seen as a larger negative Blue bar and a larger positive Red bar.



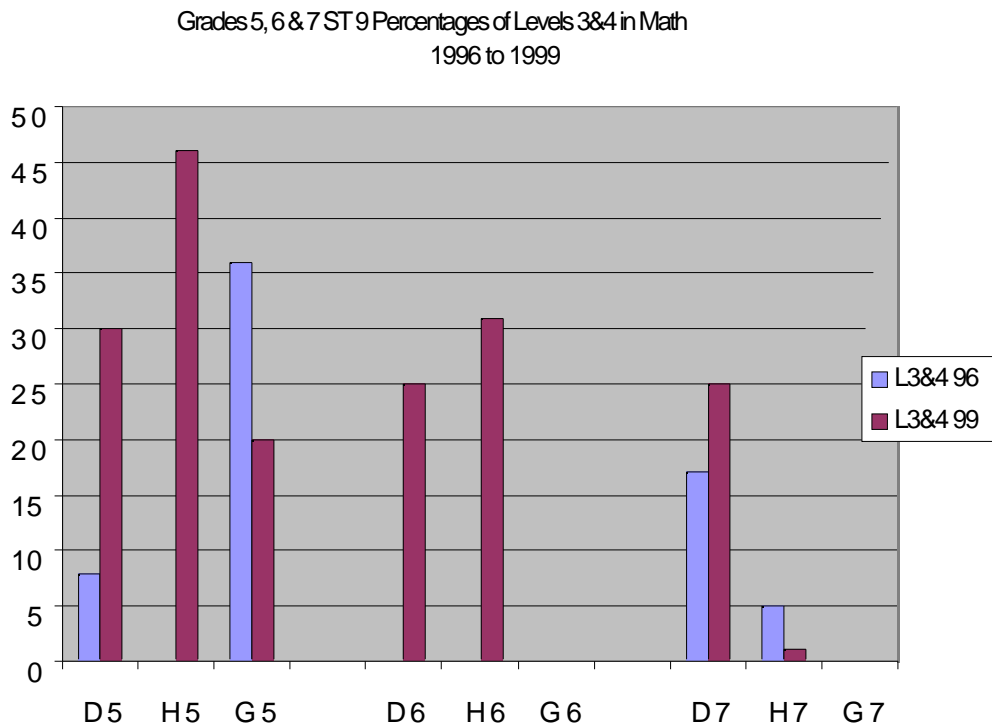
The following graph shows the actual percentages of students scoring at Level 1 in Math for the three years in which there are comparative data for Grades 5, 6 and 7. We see from this chart that although the improvement of the Hernandez students at Grade 5 was the least of the comparison groups, overall there were fewer students scoring at Level 1 in 1996 and 1999 than either the District or the Garrison. At Grade 6, a lower percentage of students at the Hernandez scored at Level 1 than either the District or the Garrison, demonstrating the most improvement of any year for raising students out of

Level 1 Math. In the 7th grade the percentage of Hernandez students was somewhat greater than the District in Level 1, but considerably less than at the Garrison. There was also a higher rate of improvement for the Hernandez on this measure than for the comparison groups. The Blue column indicates Level 1 percentages for each compared group in 1996 and the Red column indicates Level 1 percentages for each compared group in 1999. The lower the columns the better the attainment of students, in general. The greater the decrease between the Blue and the Red column, the Red being smaller, the greater the improvement over time.



The other measure of improvement consists of the increase of students in combined Levels 3&4 over time. As we see from the chart below, which shows the actual percentiles of students scoring at the highest levels in Math for the three grades we are considering, the Hernandez has the highest gains from 1996 to 1999 in Grades 5 and 6. In Grade 7, when the student population diminishes by half, the gain is minimal. The

District shows a moderate gain, but the Garrison has no gain at all in the 7th Grade at the highest levels of Math performance. The Blue columns represent the 1996 combined Levels 3&4 and the Red columns represent the 1999 percentages. The taller the column, the more students are performing at higher levels of Math proficiency. The greater the difference between the Blue and Red columns, with Red taller, the greater the improvement over time. The Hernandez shows the highest gains on this measure in Grades 5 and 6.



Open-Ended Stanford 9 Results

After the introduction of the Stanford 9 Assessment Tests to the BPS system in 1996, BPS began introducing the open-ended format of the test gradually. Open-ended Reading tests were first administered in 1997 to Grades 2, 3, 5, and 8 and in Grades 2 and 6 in 1998 and 1999. Open-ended Math tests were administered beginning in 1998 in Grades 5 and 7 and repeated in those grades in 1999. We will look at grades that have multiple years of test scores for change over time comparison purposes.

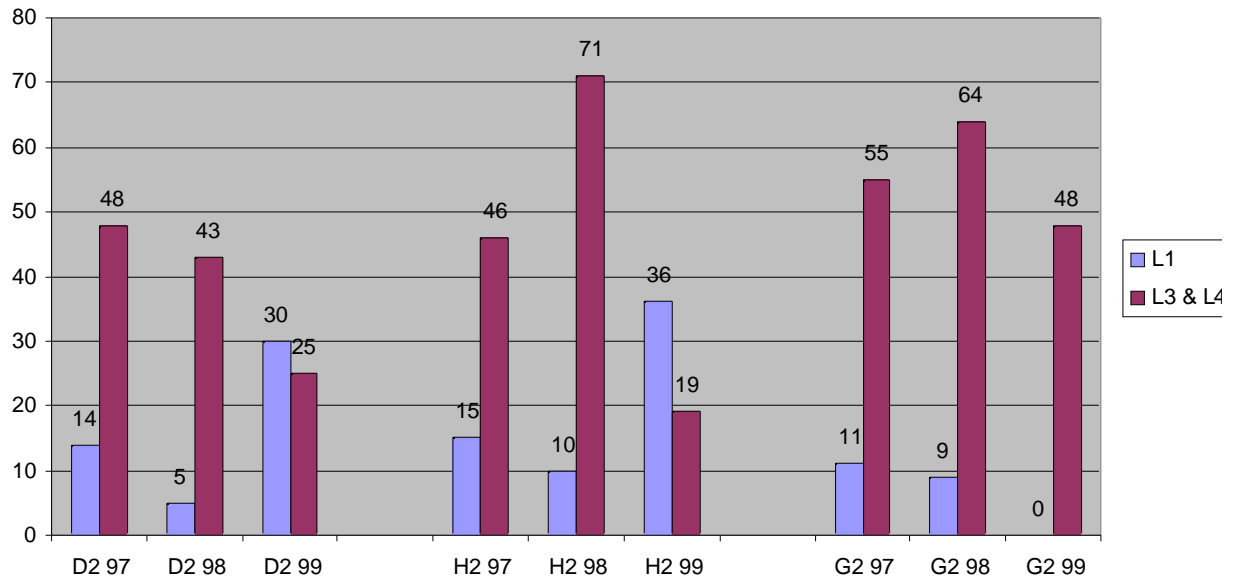
Open-ended assessments differ from the multiple choice Stanford 9 tests, in which they choose answers from predetermined alternatives, in the following ways:

- Each open-ended test consists of nine short answer questions on both the Reading and Math tests.
- Students are asked questions that seek both process and content knowledge in Reading and in Math.
- Students answers are scored on a 0-3 scale allowing for partial credit if students express some understanding in their written answers. This differs from multiple choice questions which are scored either 0 or 1, wrong or right.
- Scorers hired by the testing company use performance indicators for each question to determine partial, full, or no credit to open-ended questions.
- Students are not penalized for spelling or grammar errors in their answers.
- Numerical scores are then translated into Performance Levels according to expert criteria.

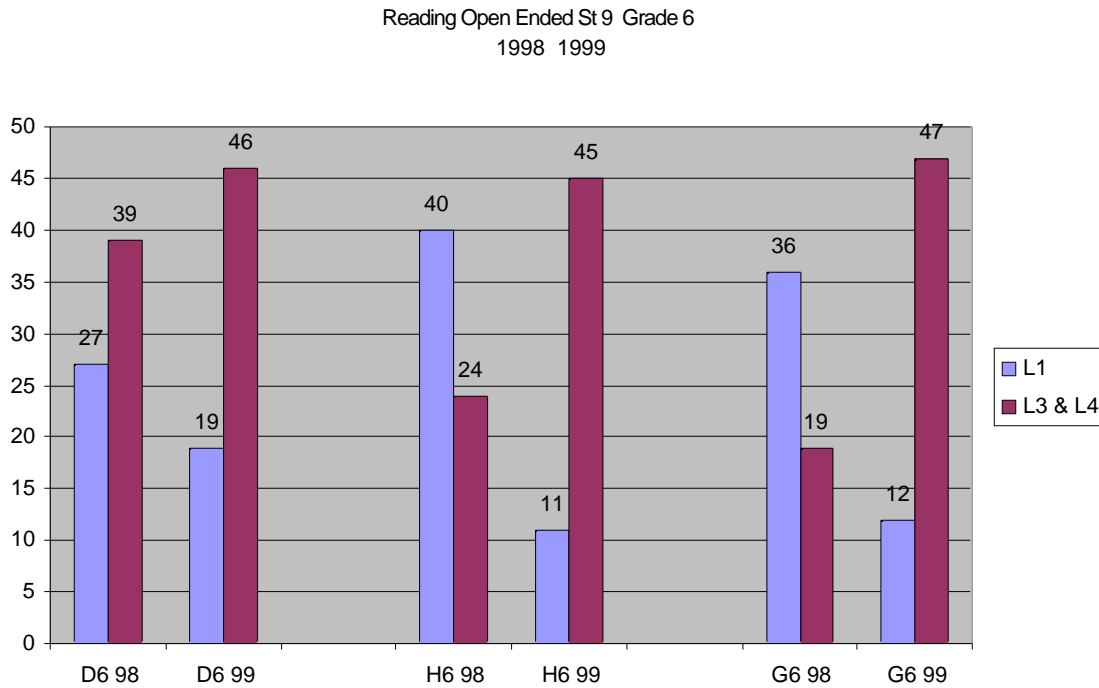
Given the nature of Learning Expeditions of student generated inquiries, we would expect the Hernandez students to demonstrate high performance levels on this form of assessment, even more than on forced choice answers. Expeditions require students to problem solve, evaluate information, categorize, relate and explain it--all areas that are purportedly tapped in the open-ended format of the Stanford 9 Assessments. This assumption is not borne out, however, as the data below indicate.

Grade 2 open-ended Reading scores from 1997, 1998 and 1999 show the Hernandez scoring higher in Level 1 in all three years than either the District or the Garrison. On combined Levels 3&4, the Hernandez outscores the District and the Garrison in 1998, but shows a lower percentage of high performing readers than the District and the Garrison in 1997 and 1999. Thus, the Hernandez, in Grade 2, does not compare particularly well against the District or its matched sample on open-ended Reading scores. The following graph illustrates these numbers. The Blue columns indicates Level 1 scores and, as such, higher achievement is seen in shorter Blue columns. Red columns designate percentages of students performing at Levels 3&4 combined. Taller Red columns indicate higher achievement.

Reading Open Ended ST 9 Grade 2
1997-98-99



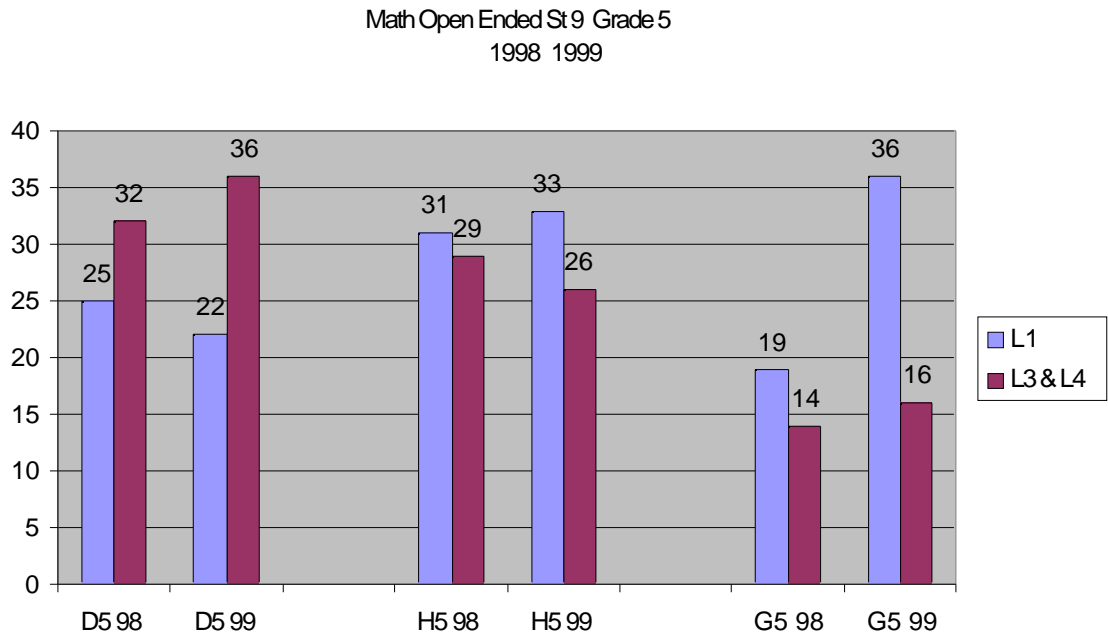
Open-ended Reading scores in Grade 6, on the other hand, show improved performance for the Hernandez over the two year period 1998-1999. The percentage of Level 1 students decreased by the largest amount of the three compared groups, 29%, while the District decreased 8% and the Garrison by 24%. At the same time, the Hernandez also increased the percentage of students scoring at the highest levels, as did the District and the Garrison, 21%, 7% and 28% respectively. As mentioned earlier, the low numbers of test takers in each grade at the Hernandez and the Garrison probably make the gains in all three groups here roughly comparable. 6th graders across the board improved in open-ended Reading tests in the two year interval we are considering. The graph below illustrates these gains.



Open-Ended Math Scores

Math open-ended scores for the 5th grade are presented in the graph below. We see that the District improves in lowering Level 1 percentages, albeit marginally, by 3%. The Hernandez increases its percentage of Level 1 students, minimally, by 2% indicating no improvement on this measure. The Garrison also increases its percentage of Level 1 students, by a greater margin, 17%.

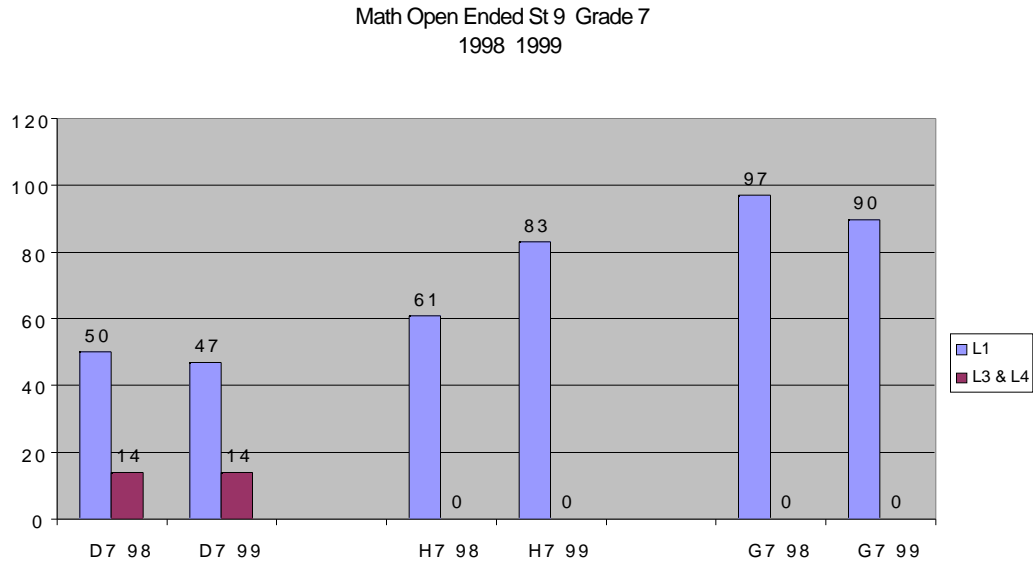
High scoring students increase minimally in the District and at the Garrison, 3% and 2% respectively, while the Hernandez shows a slight decrease in Levels 3&4 combined, -3%. Given the small numbers, this means that the Hernandez remained virtually the same over the two years that the Math open-ended Stanford 9 tests were administered.



The final open-ended Math score comparison comes from the 7th grade in 1998 and 1999. This is the grade in which we have seen the greatest dip in scores on previous measures and the same trend holds true here. While both the District and the Garrison demonstrate a small diminution of Level 1 scorers, -3% and -7% respectively, the Hernandez registers a sizeable increase in Level 1 scorers, 22%. All three comparison groups hold steady in the number of high performing students, with only the District registering any students scoring at these Levels, 14%, compared to 0% at both the Hernandez and the Garrison.

Initial open-ended results for the Hernandez do not show relatively high performance nor clear improvement in any of the grades except for Grade 6 Reading and even these scores appear to be comparable to gains across the District and at the Garrison. Thus, the open-ended scores on the Stanford 9 do not vary significantly from multiple choice versions of the test for the Hernandez. Both tests indicate that some grades improve and others fall behind and, in general, the Hernandez is more successful at

decreasing the number of students scoring at Level 1 than in increasing the number of high performing students when compared to the District and the other two-way bilingual school in the District.



SUMMARY OF STANDARDIZED MAT AND STANFORD 9 TEST DATA

It is difficult to make definitive statements about the progress or lack of it as measured by standardized tests over the years we are considering here (pre ELOB implementation--Time 0, early active ELOB implementation--Times 1 and 2, and later ELOB implementation--Time 3) because the tests administered to students in BPS differ during this time period. Moreover, due to the change in testing regimes, not all grades were tested similarly during the years considered here. In some years early grades were tested and then it was shown to be inappropriate to test young children on multiple choice tests before the 4th grade, thus, ending that form of testing. No comparative scores are

available. In very recent years the system began giving an alternative form of the Stanford 9, open-ended questions requiring written answers, in some grades. There is little comparative data to explore here given the recency of these initiatives. And finally, since the MCAS was begun in 1998, grades that had previously been given the Stanford 9 are sometimes exempted in order to lighten the testing load for some grades, notably Grades 4 and 8. When both tests are administered in the same grade, however, students spend a significant number of days in the final quarter of the school year preparing for and taking these tests. Some teachers and school administrators fear that students experience test burnout and do not give the tests their best performance.³¹ These observations about the test results and the conditions that prevail not only make it difficult to summarize trends in the data and may also offer some explanation for the variation in performance that we see in Time 3 reported results.

Very grossly outlined, however, we can describe the performance of the Hernandez on a variety of standardized measures over the course of the years examined here as follows:

- Pre- ELOB implementation, the Hernandez performed slightly lower than the District overall. It did better vis-à-vis the District in early grades, but began to lose this advantage in the late elementary years and continued to widen the gap during the middle school years.
- At the beginning of whole-school ELOB implementation (1994-95), the Hernandez showed considerable improvement over the District when compared on the same tests. The number of grades in which it outperformed the District increased to later elementary and early middle school grades and it maintained its advantage in the early grades as well. The gains were even greater for Math than for Reading.
- At active ELOB implementation in the early years, when BPS initiated a new standardized testing regime (1995-96), the Hernandez continued to fare well when

³¹ This concern was voiced throughout the BPS, not only at the Hernandez. It reveals practitioner perceptions about test taking conditions as a result of their contact and conversations with students, but there are no data to prove or disprove these perceptions.

- compared to the District, although not quite as well as it had on the earlier standardized test, a less demanding and less sensitive measure of academic achievement. When compared to the matched school within BPS that shares similar characteristics to the Hernandez but not engaged in ELOB, the Hernandez showed significantly higher achievement. It scored higher in all grades and in all subject area than the matched school.
- In later years of ELOB implementation, at most three years, but often one or two years later depending on the reporting of scores (1996 to 1997, 1998 or 1999), the Hernandez does not maintain its comparative profile. Compared to the District is nearly always scores lower and shows less overall improvement. Compared to the Garrison it is more similar in scores, meaning that the Garrison improves overall at a greater rate than the Hernandez. This general trend changes, however, from grade to grade and from measure to measure increasing the variability we see in the way the Hernandez performs on versions of the Stanford 9 in the later years of ELOB implementation.

MCAS Results

The State of Massachusetts requires all 4th, 8th and 10th graders to take the Massachusetts Comprehensive Assessment System (MCAS) in English Language Arts, Math, Science and Technology and Social Studies.³² These tests, like the newer version of the Maine MEA, consist of a combination of multiple choice and short answer questions, as well as a long written essay. The short answers and the essay are scored by trained raters from the testing company that designed the test. The content of the exams reflects the Curriculum Frameworks developed by the State Department of Education, which are also reflected in the newly drafted Citywide Curriculum Guidelines for the BPS. As mentioned earlier, the MCAS is part of the high stakes testing policy adopted by the Superintendent and School Board in Boston. Following the State-wide mandate, students

³² Science and Technology was tested in 1998 and 1999, Social Studies in 2000. We only report on the ELA and Math scores here.

must pass the 4th and 8th Grade versions of the ELA and Math components of the test to pass on to the next grade. 10th graders, as of 2003, will be required to pass the ELA and Math portions of the 10th grade MCAS in order to graduate from high school.

Scores for the MCAS are reported in Performance Levels as are the Stanford 9 results. Level 1 is **Failing**, Level 2 **Needs Improvement** but is considered passing, Level 3 indicates **Proficient** and Level 4 **Advanced** performance. Composite scores for each subject area of the exam are also reported. Experts determine which scores comprise the bands of the four Performance Levels.

The introduction of the MCAS in 1998 coincided with a major effort launched by the Superintendent with the aid of Annenberg Challenge Funds and other local corporate donations to implement comprehensive school reform in every school in the District. Boston, once a rather loosely run school system which generally allowed schools to teach how and what they deemed appropriate, has now embraced Standards-based reform, with the Central Office requiring specific curricula, products, assessments and coordinated literacy instruction at all levels of schooling.³³ All schools in Boston are now in the process of Standards-based whole school change. The Hernandez initiated the District-mandated reforms with the first cohort of schools in 1996-97. The Garrison began the process with the second cohort, in 1997-98. Curriculum and assessment has changed radically in the BPS over the past four years. Professional development has engaged teachers in literacy instruction across the curriculum as well as in performance assessments and regular collective planning and assessment of student work. Teachers in the Boston schools are feeling a pressure to change and to produce better results in their students that veterans of the system can not remember ever feeling to the same degree. The reactions among faculty are mixed. Those resistant to change attempt to thwart the reform process or at least be unaffected by it. Many more, however, are taking advantage of the resources made available to them to learn new methods of instruction and to collaborate with colleagues in learning to implement the new, more demanding curriculum and assess their students' progress. Nevertheless, nearly all teachers admit to being

³³ The following description of BPS reform, District-wide, comes from the author's observations and close association with the reform efforts since 1998.

overwhelmed by the new, higher expectations for student performance and the new and unfamiliar work it requires of them.

The MCAS results, because the test is more demanding than the Stanford 9 and more aligned with the specific requirements of the City and the State, are becoming the major benchmark of student achievement in the City and across the State. It will not replace the Stanford 9 in Boston, however, because it is only given in three grades and therefore does not allow for cohort analysis, nor does it assess all students each year.

The two years of MCAS results we report on here, from 1998 and 1999, show more consistent progress and achievement for the Hernandez relative to our comparison groups than the Stanford 9 have shown. It is not possible, unfortunately, to correlate the MCAS with the Stanford 9 results because the City exempted MCAS testees from taking the Stanford 9 whenever feasible. Thus, 4th and 8th graders do not have all the relevant Stanford 9 scores to compare with the MCAS.

MCAS ELA 4th Grade

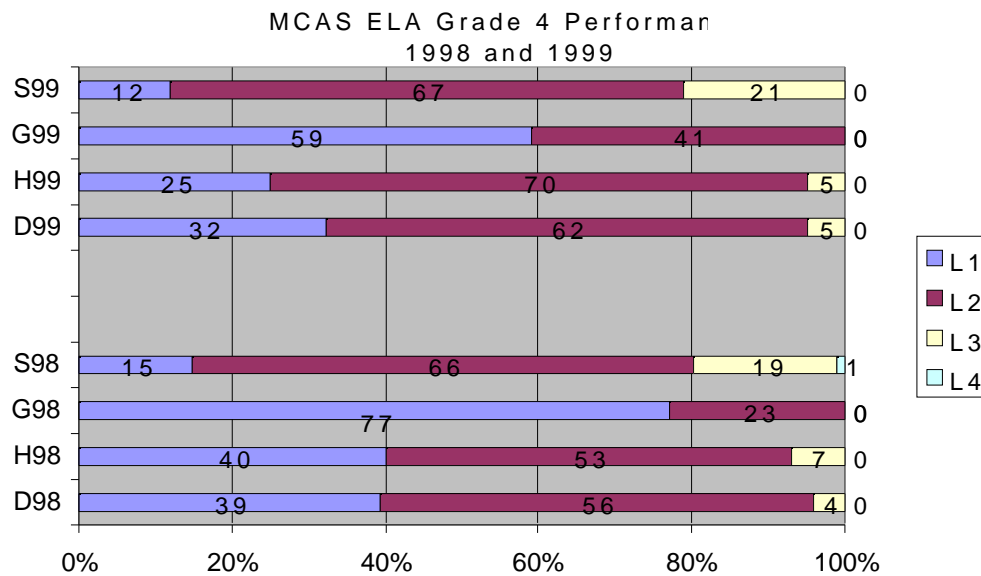
The first group of results we present are those for 4th graders on the English Language Arts portion of the exam. On all measures, the Hernandez performs well, as well or better than the District, much better than the Garrison, and only more poorly than the State at large. This is not surprising given the demographics of the BPS compared to the rest of the State.³⁴

In 1998 the Hernandez and the District had comparable numbers of students scoring at Level 1, 40% and 39% respectively. The Garrison had far more Level 1 students, 77% of its total 4th graders. At the same time, the Hernandez had the highest number of students scoring at combined Levels 3&4 of the three groups: 7% compared to 4% in the District and 0% at the Garrison. Only the State shows a completely different profile with only 13% of its 4th graders at Level 1 and 19% at combined Levels 3&4. Only 1% of 4th graders State-wide scored at Level 4 in 1998, none in 1999. No students in Boston scored at Level 4 in either year in the 4th grade on the ELA portion of the MCAS.

³⁴ While the State of Massachusetts has several urban districts, all of which score at or below BPS levels, the majority of students in the state attend suburban schools. These districts vary widely in their performance on the MCAS, but generally perform better than urban districts. The one exception to this is the scores of the Boston exam schools which rank among the highest performing secondary schools in the State.

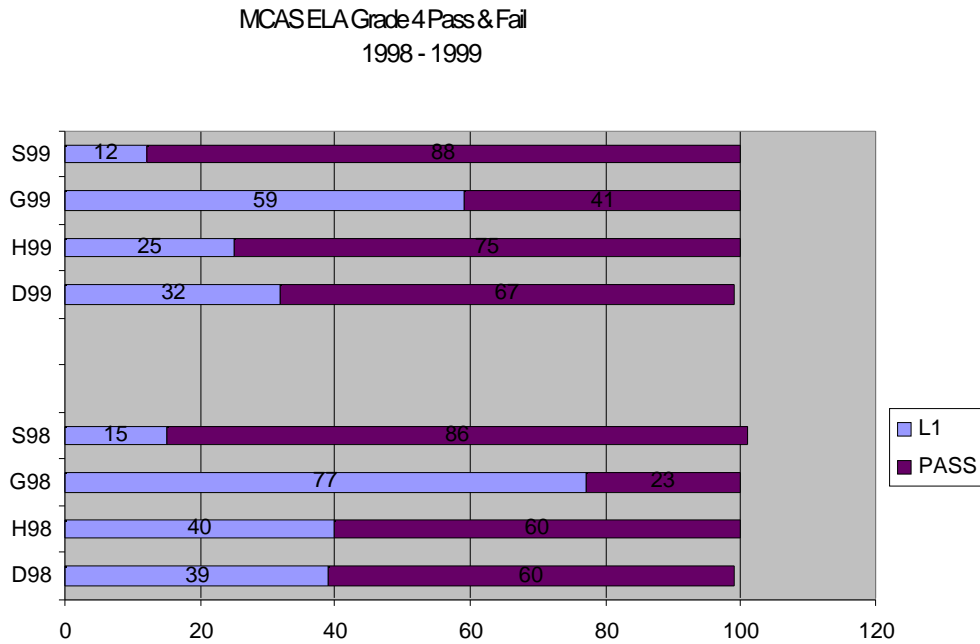
In 1999 all of the groups we are comparing improved in reducing the number of low performing students. The Hernandez had the lowest percentage of students in Level 1 of all the groups except the State, with only 25% compared to 32% District-wide and 59% at the Garrison. The State levels remain relatively constant throughout our comparisons, with 12% of 4th graders scoring at Level 1.

The following chart illustrates these numbers. Individual bars represent the comparison groups for 1998 and again in 1999. The Blue part of the bar represents Level 1 students, Red indicates Level 2, Yellow is Level 3 and Aqua shows students at Level 4.



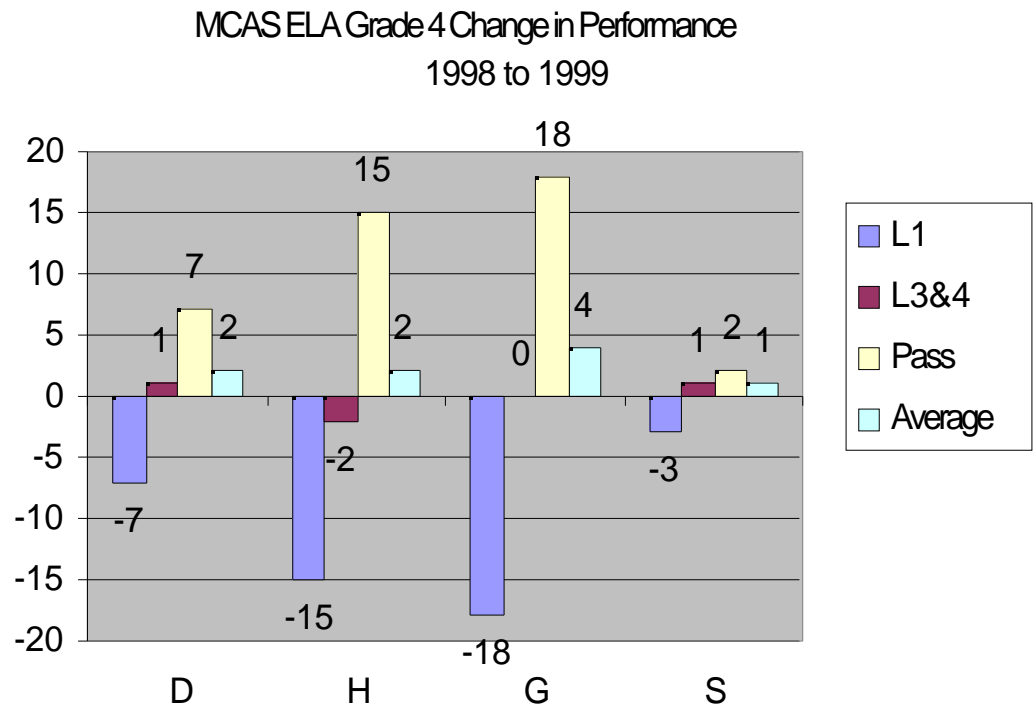
Since the MCAS is a high stakes test, we have illustrated these same data in terms of students passing vs. failing the test in the two years we analyze here. Failing applies to students scoring at Level 1 and passing includes students in Levels 2, 3 & 4. The graph

below shows this relationship and highlights not only the superior performance of the Hernandez in both years vis-à-vis the Garrison, but also how it improves at a rate greater than the District when we compare 1998 and 1999 scores.



The percentage of change from 1998 to 1999 is represented in the following graph. The Blue column represents Level 1 changes, the larger the bar on the negative portion of the graph, the greater the improvement in moving students out of the lowest performance level. The Hernandez improves more than the District, but not more than the Garrison, but the Garrison had far more students in Level 1 to begin with and far more than the Hernandez even after the improvement we see in 1999. The Red column indicates the change in Levels 3&4 between 1998 and 1999. A higher column on the positive side of the graph indicates greater improvement. We see on this measure that the Hernandez is the only group to decrease its number of high performing students. The District and the State increase their percentages and the Garrison stays steady with no students performing at either Level 3 or 4. The Yellow column indicates the change over time in the percentage of students who pass the exam. All groups show an increase in this category.

The Hernandez improves more than the District and the State but marginally less than the Garrison. The final column in Aqua indicates the change in the average number score for the entire ELA portion of the 4th grade MCAS between 1998 and 1999. All of the compared groups show improvement in the average score over time as indicated by the positive Aqua columns.



The table below shows the raw score for the 4th grade on the ELA portion of the 1999 MCAS for all the comparison groups. The 1999 raw scores show a slight improvement for all of the groups from the 1998 raw scores.

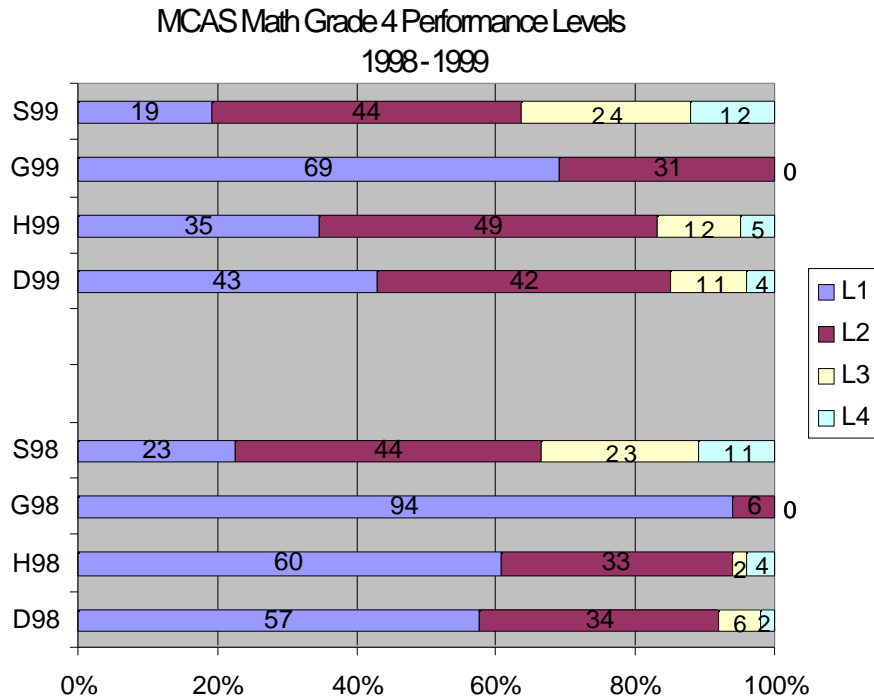
	98 Average	99 Average
District	222	224
Hernandez	222	224
Garrison	213	217
State	230	231

MCAS Math 4th Grade

Math MCAS scores for the 4th graders show the Hernandez only slightly below the District's performance in 1998 and outperforming both the District and the Garrison in 1999, showing greater improvement than the District from 1998 to 1999. The graph below, similar to the one showing ELA Performance Levels above, indicates the percentages of scorers in all four Performance Levels for both years. It shows the Hernandez having slightly more students scoring at Level 1 than the District in 1998, 60% compared to 57%, but many fewer than the Garrison which had 94% of its 4th graders scoring at Level 1. The Hernandez also had slightly fewer students scoring at Levels 3&4 than the District by a small margin in 1998, 6% compared to 8%. These differences, given the small numbers in the Hernandez sample, are negligible. The Garrison, however, does not have any students scoring at the higher levels in 1998.

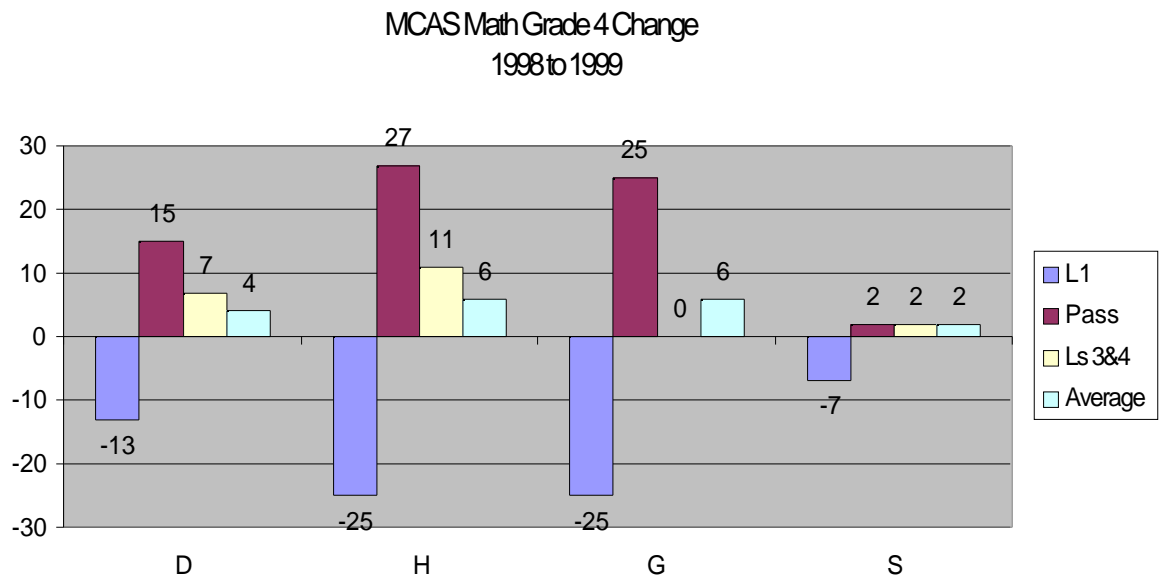
In 1999 the Hernandez outperforms both the District and the Garrison. It reduces its Level 1 students to 35% and increases its Levels 3&4 students to 17%. For the same year the District has 43% of its total 4th graders scoring at Level 1 in Math and 15% scoring at the higher levels. Again the Garrison does much worse than either the Hernandez or the District on Math scores in the 4th grade with 69% performing at Level 1. This, however, is a great improvement over the 1998 score. The Garrison still does not show any students scoring at either Level 3 or 4 in 1999.

As with the ELA comparison, the State outperforms the District and the Hernandez and Garrison significantly both in the lower number of Level 1 students, 23% and 19% in 1998 and 1999, respectively, and in the percentage of high performing students, 34% in 1998 and 36% in 1999.



The graph of changes between 1998 and 1999 below shows the percentage loss in Level 1 students in the negative Blue columns for all four comparison groups. The Hernandez and the Garrison show the same high percentage of improvement, 25%, but given the starting point of the two schools, the Hernandez had greater improvement in reducing its Level 1 students than any other comparison group. Similarly, the Hernandez shows the greatest increase in its percentage of students who pass the Math portion of the 4th Grade MCAS in 1999, 27%. The greater the improvement in numbers of students passing, the higher the Red bar on the chart below. Levels 3&4 also increase more for the Hernandez than any of the comparison groups: 11% compared to 7 % in the District, no increase for the Garrison, and 2% increase State-wide. The final measure recorded for all four groups on the following graph is the change in number score for Math between 1998 and 1999. We see the same amount of change, 6 points, at the Hernandez and the Garrison, although the Hernandez outscored the Garrison in both years: 220 compared to

207 in 1998 and 226 compared to 213 in 1999. The Hernandez also outscored the District in both years: 220 compared to 219 in 1998 and 226 compared to 223 in 1999. Moreover, the improvement at the Hernandez was greater than that of the District: 6 points compared to 4 points. Again we see the State Performance Levels more stable over time, and the raw score of State-wide 4th graders is significantly higher than any BPS score: 234 in 1998 and 235 in 1999. The table showing the raw scores on the Math portion of the 4th grade MCAS follows the chart.



	98 Average	99 Average
District	219	223
Hernandez	220	226
Garrison	207	213
State	234	235

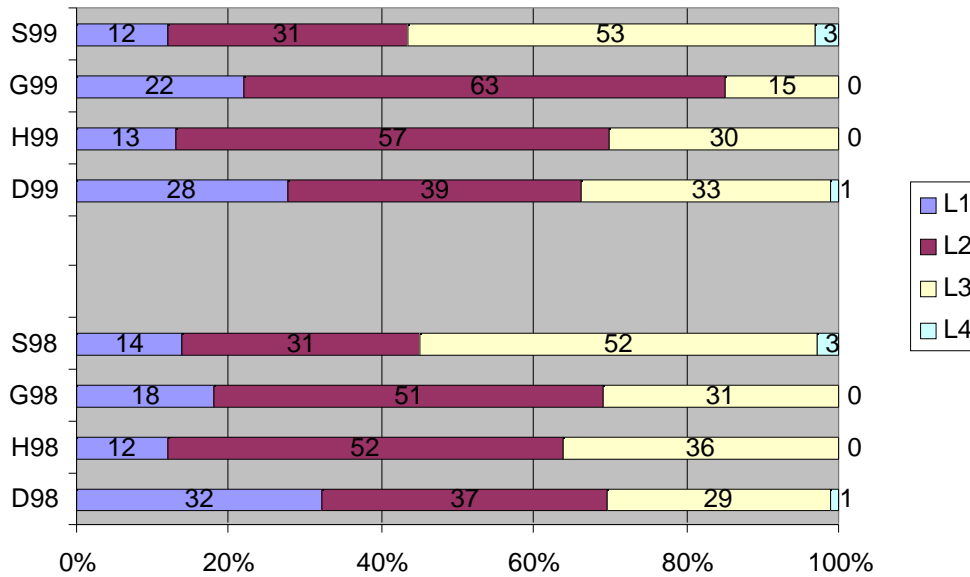
MCAS ELA 8th Grade³⁵

ELA MCAS scores for the Hernandez in the 8th grade show more proficiency than either the District or the Garrison in 1998. In fact, the Hernandez comes close to State-wide levels at the lowest performance level, Level 1. As the graph below shows, the Hernandez has only 12% of its 8th graders failing the MCAS ELA assessment at Level 1 in 1998 compared to 32% in the District and 18% at the Garrison. It comes closest to State-wide performance at Level 1, which is 14%. The Hernandez continues to outscore the District and the Garrison in 1998 at the highest performance levels, Levels 3&4. The Hernandez shows a combined percentage of 36%, while the District has 30% and the Garrison 31% in the highest levels. The State, on the other hand, records a total of 55% of its 8th graders scoring at the highest levels in ELA.

In 1999 the Hernandez shows a slight diminution in its performance but still outscores both the District and the Garrison at Level 1: 13% compared to 28% at the District and 22% at the Garrison. The State percentage of Level 1 performers is 12%, again similar to the Hernandez, but much better than the District or the Garrison. High scoring students at the Hernandez also decrease slightly in 1999, to 30%. The District scores slightly higher at 34% while the Garrison scores much lower--15%. The State outscores the District and our comparison schools by a wide margin at the higher levels of performance in ELA with 56% of 8th graders State-wide demonstrating proficient or advanced achievement levels.

³⁵ The numbers of students taking the MCAS in Grade 8 at the Hernandez decreases by 1/2, making any changes in percentages over time highly variable. The same is true for the Garrison, which also has small numbers in each of the grades, but more so at the middle school grades. The District and State numbers of test takers is sufficiently large to make changes in these scores from year to year significant.

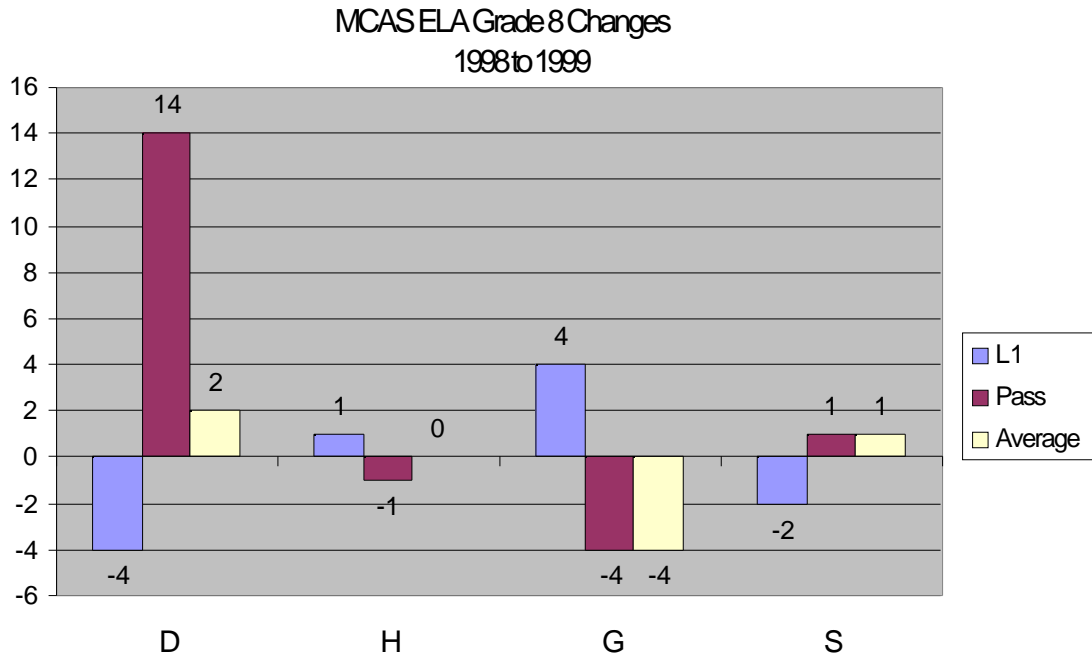
MCAS ELA Grade 8 Performance Levels
1998 and 1999



The next graph illustrates the changes in percentage scoring at Levels 1, and Levels 2, 3 & 4 (Passing) from 1998 to 1999 for each comparison group. It also shows the changes in average scores on the 8th Grade ELA portion of the MCAS for all four comparison groups. The aforementioned caveat referring to small numbers of students taking the test at this grade level should be kept in mind for these comparisons in particular.

In terms of improvement between 1998 and 1999, the Hernandez shows no gains, while the District decreases its percentage of Level 1 students by -4% and increases its high performing students by 14%. The Garrison fares more poorly than the Hernandez. It increases its Level 1 students by 4% and decreases its higher performing students by 4%. The State shows small gains between 1998 and 1999, with a decrease in Level 1 students in ELA of 2% and an increase of higher performing students by 1%. The average total score for 8th graders remains relatively constant, except for a decrease of 4 points at the Garrison in 1999. The District improves by 2 points, the State by 1 point and the

Hernandez remains the same, although the Hernandez' raw score is higher than either the District's or the Garrison's. The raw total ELA scores for 1998 and 1999 are contained in a table following the graph.



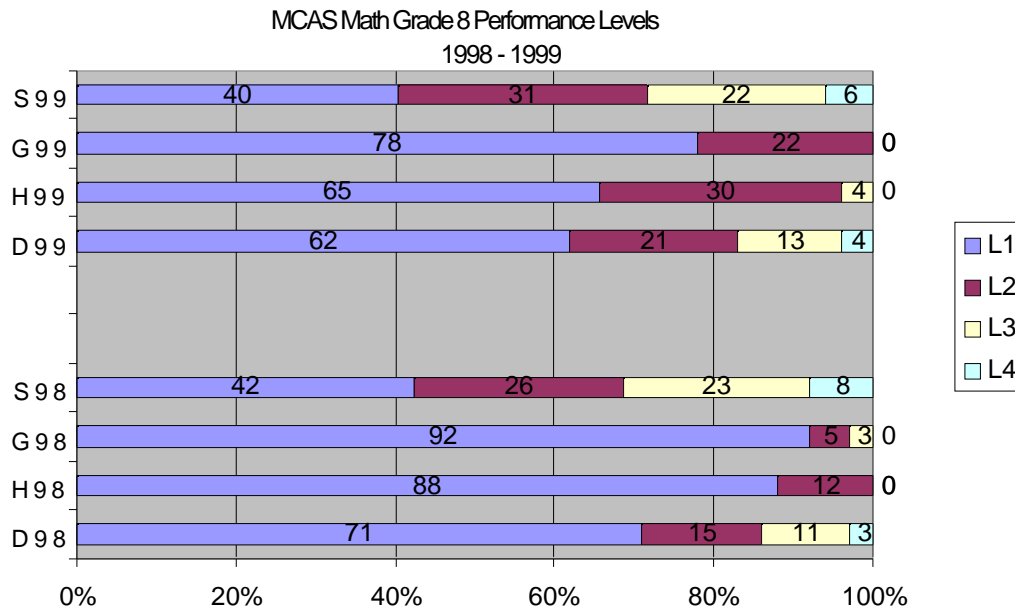
	98 Average	99 Average
District	228	230
Hernandez	232	232
Garrison	231	227
State	237	238

MCAS Math Grade 8

8th Grade Math scores on the MCAS show marked improvement for the Hernandez between 1998 and 1999, but the initial scores in 1998 are lower than the District and the State and only slightly higher than the Garrison. Level 1 students at Grade 8 in Math are highly represented in BPS in general. The Hernandez has 82% of its 8th graders scoring at

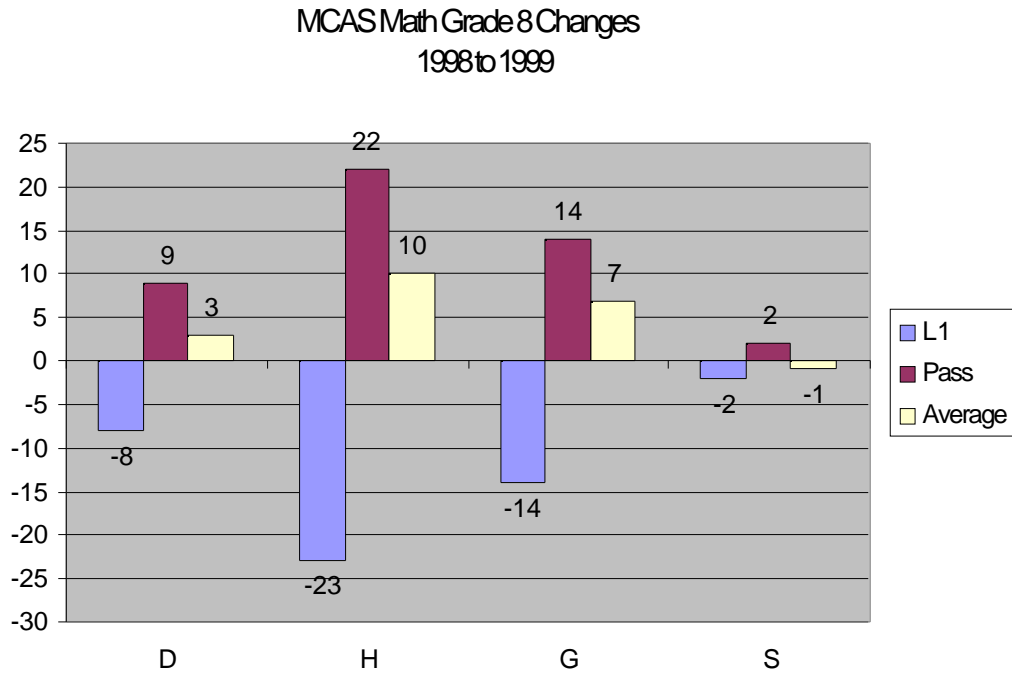
this level compared to 71% in the District and 92% at the Garrison. The State, on the other hand, has somewhat less than half of its students at this level, 42%.

At the higher scoring levels the District outperforms the Hernandez and the Garrison with scores of 14%, 12% and 8%, respectively, while 31% of 8th graders State-wide scored at Levels 3&4.



Given the small numbers of students in the Hernandez and Garrison samples, the changes over time, which are minimal anyway, are even less meaningful. Basically, the Hernandez and the Garrison maintained their 1998 distribution across Performance Levels in Math in 1999. Since the numbers for the District and the State are sufficiently large, the following graph shows that the District as a whole improved its performance in Math between 1998 and 1999 by reducing the number of Level 1 students by 10% and increasing Levels 3&4 students by 14%. There was also a 2 point overall gain in the average score on the Math portion of the 1999 MCAS. The State improved slightly: 2% fewer students scoring at Level 1 and 1% more students scoring at Levels 3&4 for the same time period. The point difference between 1998 and 1999 MCAS Math averages was 1. The total points scored on the Math portion of the MCAS by 8th graders in our

comparison groups, showing that the Hernandez maintained the highest overall score compared to the District and the Garrison, is presented in a table following the graph.



	98 Average	99 Average
District	214	217
Hernandez	208	218
Garrison	205	212
State	227	226

SUMMARY OF MCAS RESULTS

We see from the data presented above that the Hernandez outperforms both the District and the Garrison in Grades 4 and 8 in both ELA and Math, overall. It shows the greatest gains in preventing failure, i.e., low percentages of students scoring at Level 1, and in the overall scores for ELA and Math at both grade levels. The improvement

between 1998 and 1999 is not as great as the comparison groups in BPS on the whole. It consistently performs lower than the State averages on the MCAS, but more often higher than the average District score.

Given the fact that the Hernandez and the Garrison represent the "normal" BPS student compared to the District, whose averages include Advanced Work Class (high track) students in the 4th grade and selected exam school students in the 8th grade, the Hernandez shows high proficiency on all the reported MCAS assessments compared to the District as a whole and certainly as compared to the matched sample school, the Garrison. Considering, moreover, that 100% of the students at the Hernandez (and the Garrison) took the exam, compared to only 67% of LEP students District-wide and 58% of LEP students State-wide, the high scores of the Hernandez are the more remarkable. As a two-way bilingual school that enrolls more than half of its students as bilingual, Spanish-dominant speakers, the high performance of the Hernandez on the MCAS, the most rigorous of standardized tests given to BPS students, indicates that 4th and 8th graders excel in academic achievement as measured by the MCAS compared to other BPS students.

CONCLUSIONS

The two questions we posed for this report--What was the academic achievement as measured by standardized tests for the Rafael Hernandez School over the years of ELOB implementation? and, What impact did ELOB have on the academic achievement of the Rafael Hernandez School students?--can be summarized as follows:

In the first years of whole-school ELOB implementation, the Hernandez students scored higher than the District in more grades than prior to ELOB implementation. The improvement was seen in both Reading and Math scores, with more elementary grades outperforming the District.

On the first benchmark year of a different standardized assessment adopted by the system (1996), the Hernandez continues to score as well or better than the District

in the elementary grades in both Reading and Math, but falls slightly below District comparisons starting in Grade 6. This gap increases in Grades 7 and 8, but the population of the Hernandez reduces by 40% in these years because of the exodus to exam schools and non-public alternatives, making any fluctuations in scores seem greater than they actually are.

Starting in 1996 we also compare the Hernandez to a matched school in the District that shares a similar population and organizational characteristics but does not share in ELOB implementation. The Hernandez significantly outperforms the Garrison school in Reading and Math in all grades tested, both elementary and middle school grades.

In later years of ELOB implementation (1997-1999) the Stanford 9 tests are administered inconsistently by grades and by format (multiple choice or open-ended) since the assessment procedure for the District is undergoing sweeping changes from 1996 to 1999 (the last available test score year). In the later years of ELOB implementation (1997-1999), the Hernandez does not maintain its comparative profile. Compared to the District, it nearly always scores lower and shows less overall improvement. Compared to the Garrison it is more similar in scores, meaning that the Garrison improves overall at a greater rate than the Hernandez. We see variations in the comparisons, however, from grade to grade and from measure to measure, increasing the variability we see in the way the Hernandez performs on versions of the Stanford 9 in the later years of ELOB implementation (1997-1999).

The final measures we compare are the English Language Arts (ELA) and Math portions of the MCAS from 1998 and 1999, later years of ELOB implementation. These tests only apply to 4th and 8th graders. The messy picture presented by Stanford 9 results is not reproduced in MCAS scores. The Hernandez outperforms both the District and the Garrison in Grades 4 and 8 in both ELA and Math, overall. It shows the greatest gains in preventing failure, i.e., lowering percentages

of students scoring at Level 1, and in the overall scores for ELA and Math at both grade levels. It consistently performs lower than the State averages on the MCAS, as do all but the BPS exam schools, but more often the Hernandez scores higher than the average District score. This high comparative performance on the MCAS is all the more significant because the percentage of test-takers at the Hernandez exceeds those of the District and the State. Whereas only 67% of LEP students in the District and 58 % in the State are accounted for in MCAS scores, 100% of Hernandez LEP students are included in its scores.

There are clear indications from these data that the Hernandez--even with a very high percentage of Hispanic bilinguals, traditionally among the lowest performing students on standardized measures--shows relatively high academic achievement. It keeps pace with the District and often exceeds its performance in spite of the fact that the District scores include exam school students from middle school onwards and Advanced Work Class students (high track classes available in some BPS schools from the 4th to the 6th grade). The Hernandez, as a non-tracked school, provides the same high demanding curriculum of all its students, regardless of label (special education or bilingual) or past performance. This is especially evident in the newly instituted MCAS beginning in 1998. Although only the 4th and 8th graders are tested in this system, the performance at the Hernandez of 100% of their students is better in most cases than the District which only tested 67% of its LEP students, and sometimes approaches State levels where only 58% of LEP students were tested, and the population and economic conditions of the majority of State-wide schools coincides with traditionally higher performing students.

Thus, as we saw in the case of the King Middle School as well, ELOB implementation appears to be providing a strong academic curriculum that allows students from typically disadvantaged backgrounds to thrive. In the case of the Hernandez we see that a non-ELOB school that shares many similarities with the Hernandez does not score nearly as well on most measures as the Hernandez. This provides further support for our conclusion that ELOB is making a positive difference in the academic achievement of the Hernandez.

The unevenness of the Stanford 9 data we compared at Time 3 (a variety of time periods in the later years of ELOB implementation) must be explained, however. It appears from many of the measures that the Hernandez loses ground compared to the District and to the Garrison. It scores overall more poorly than the District in most areas and grades, and for the first time is outperformed by the Garrison on some measures. The Hernandez also demonstrates less improvement over these years than either the Garrison or the District.

As mentioned earlier, there are several things to bear in mind while attempting to explain this apparent drop in performance. First, the numbers of students at the Hernandez and the Garrison are sufficiently small, especially in the 7th and 8th grades, so as to make any comparisons of percentages over time show greater fluctuations than actually exist. Second, the measures presented vary considerably due to differing administration and reporting of the Stanford 9 test scores. Third, the downward trend in the Hernandez Stanford 9 data for later years of ELOB implementation is not mirrored in the MCAS for those same years, although the MCAS only tests 4th and 8th graders. This raises some questions as to what might cause differences in outcomes on the two tests. Are they testing the same things? Were they treated with similar seriousness by students and staff? Which, if any, of the two tests is a more accurate profile of actual student achievement?

Some of the contextual information regarding the later years of ELOB implementation at the Hernandez may be relevant here. As the BPS initiated its own comprehensive reform across the system, a number of requirements that limited the extent of ELOB curriculum investigations were put in place system-wide. These had to do with a prescribed, wide-ranging curriculum in subject areas which replaced a more *laissez-faire* attitude on the part of Central office in allowing schools to determine their own curriculum and instructional programs. The following new requirements have impacted the teachers' perceptions of the extent to which expeditions can be used and how they are to be implemented:

- Each school is now required to implement a comprehensive balanced literacy program which is to be adhered to in all subject areas. Most of the models

from which schools may choose, especially at the elementary and middle school levels, recommend homogeneous flexible reading groups so that students are reading materials appropriate to their reading ability. Although advocating practices that are consistent with ELOB, the new directives may have complicated the use of expeditions as an instructional program. For example, teachers have found it difficult to gather reading materials relating to expeditions at all the reading levels of their students. Thus, in some cases the appropriate reading material has been chosen over the content of the expedition. Since the focus of the reform has been on literacy, thus far, teachers are encouraged to implement changes in English Language Arts more in keeping with the newly adopted literacy model than to gear their instructional programs around expeditions. When the two appear to come in conflict, it appears that literacy model changes may have greater support.³⁶

- The District has mandated numerous performance assessments and other standardized tests to be given throughout the year, greatly increasing the amount of time students are formally tested each year. Teachers have found that they have less time for instruction overall. Professional development time during the school day, which was previously used by looping teams to generate, reflect on and modify expeditions, is now increasingly being used to look at student work collectively. While the two agendas are not incompatible--indeed, the Core Practices of ELOB mirror the districts' requirements--there are different messages about how a faculty in a school should go about the reform.
- City wide Curriculum Guidelines now dictate the scope and sequence of learning in subject areas across all grades. This coincides with the State-wide Curriculum Frameworks, which have been newly created and are undergoing refinement. A required curriculum is, thus, emerging and gets modified

³⁶ This is not to say that the balanced literacy approach or sequenced math curriculum advocated by the BPS is at odds with ELOB. Rather, teachers at the Hernandez find they need more time to revise and reconstruct their expeditions in order to serve both purposes well.

periodically, which sets the agenda for what teachers are to teach. In the case of Math, for example, the Hernandez teachers have opted to teach the scope and sequence curriculum of the City, with its District-constructed assessments, rather than integrate more Math in expeditions. Teachers are afraid that students will score poorly on the standardized tests if they do not cover the required Math at the proper time. It will take some time for teachers to revise their expeditions to incorporate new curriculum requirements, and in the meantime, some teachers are opting to give less time to expeditions overall and more time to the materials that the City recommends or requires.

- The increased importance of standardized tests in determining promotion and graduation criteria has an impact on the instructional focus of teachers not only at the Hernandez, but at all BPS schools. There is more class time devoted to test preparation and mock assessments, which are scored according to the Performance Levels established by the various standardized tests. This inevitably leaves less time for more exploratory learning. What effect combining these two contradictory approaches to teaching and learning have on students has not been determined.

In sum, then, the Hernandez' implementation of ELOB in more recent years has not been as comprehensive or as seamless as it was initially, nor is it given the support and consistency that we saw at the King Middle School in Maine, where the results were unambiguously positive. Based on these few years of comparing the same test with similar reporting of results, however, we do not have enough data to draw any strong conclusion about the trend in academic achievement at the Hernandez post ELOB implementation, nor can we conclude with certainty that ELOB has been responsible for either successes or lack of them at the Hernandez given all of the confounding evidence and issues. More consistent test data reports in future years should help clarify some of the picture, but the most informative evaluation of ELOB implementation at the Hernandez would look not only at test data over time, but also provide a detailed and comprehensive picture of instructional practices in all grades.

