The perceived effects of state-mandated testing in New Hampshire

Stephen A. Kossakoski

University of New Hampshire, Durham

Follow this and additional works at: https://scholars.unh.edu/dissertation

Recommended Citation

This Dissertation is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact nicole.hentz@unh.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®
THE PERCEIVED EFFECTS OF STATE-MANDATED TESTING IN NEW HAMPSHIRE

BY

STEPHEN A. KOSSAKOSKI
B.S. Ed., Keene State College, 1981
M. Ed., University of New Hampshire, 1993

DISSERTATION

Submitted to the University of New Hampshire in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Education

May, 2000
This dissertation has been examined and approved.

Dissertation Director, Dr. Grant Cioffi
Associate Professor of Education

Dr. John J. Garney
Professor of Education

Dr. Todd A. DeMitchell
Associate Professor of Education
Sonoma State University

Dr. Karen J. Graham
Associate Professor of Mathematics

Dr. Mark V. Joyce, Executive Director
New Hampshire School Administrators Association

Dr. Mary Jane Moran
Assistant Professor
Department of Family Studies

April 19, 2000
Date
DEDICATION

To my wife and best friend, Eileen, whose love, support, and understanding have made this dream possible. Also, to my daughter Kelley and son Kristopher, thank you for always believing in me.

And

To my parents, Richard and Delora, for everything...

In memory,
Of my grandfather, Stanislaw Kossakoski, who instilled in me a love of learning.
ACKNOWLEDGEMENTS

There are many people who provided guidance and encouragement to me along the way to whom I would like to express my appreciation.

First, I would like to express my appreciation and gratitude to Grant Cioffi, my dissertation director, who was always available to share his expertise and to provide encouragement as I navigated through this process.

I would also like to thank my dissertation committee members: John Carney, Todd DeMitchell, Karen Graham, Mary Jane Moran, and Mark Joyce. I truly appreciate the time and expertise that each devoted to this endeavor.

Two committee members deserve special thanks. Todd DeMitchell has served as my advisor through both the M.Ed. and Ph.D. programs during the last decade and has become a valued source of encouragement and support. Mark Joyce has influenced the direction of my career for nearly 20 years. I am very thankful that he was willing to participate in this project.

I wish to extend my appreciation to the Executive Directors of the New Hampshire Association of School Administrators, the N.H. affiliate of the National Education Association, the N.H. affiliate of the American Federation of Teachers, and the N.H. Association of School Principals for expressing their support for this research.

Finally, I would like to extend my sincere appreciation to the third grade teachers who took the time to participate in this study. Your responses to this study confirm my belief that teachers are caring and talented professionals.
# TABLE OF CONTENTS

Dedication ............................................................................................................... iv

Acknowledgements ................................................................................................ v

List of Tables .......................................................................................................... viii

Abstract .................................................................................................................... ix

## CHAPTER

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
</table>
| I. INTRODUCTION ......................................................... 1
| Background ................................................................. 3 |
| Purpose of the Research ............................................... 7 |
| Research Questions ..................................................... 8 |
| Organization of the study .............................................. 9 |
| Definition of Terms ..................................................... 10 |
| II. REVIEW OF THE LITERATURE ........................................ 12 |
| Introduction ............................................................... 12 |
| Historical Context ...................................................... 13 |
| The Influence of Testing .............................................. 25 |
| School Funding and Achievement ................................. 36 |
| NHEIAP Grade Three Test Details ................................. 44 |
| Summary ................................................................. 49 |
| III. METHODOLOGY ........................................................ 50 |
| Instrumentation ......................................................... 51 |
| Population and Sample ............................................... 56 |
| Data Collection Procedures .......................................... 58 |
| Data Analysis ........................................................... 60 |
| IV. ANALYSIS OF DATA .................................................... 67 |
| Demographic Characteristics of Survey Respondents .......... 67 |
| Curriculum and Instruction .......................................... 70 |
| Sources of Pressure .................................................... 79 |
| Attitude ................................................................. 83 |
| Items 54 and 55 ......................................................... 87 |
| Summary of Teacher Comments .................................... 90 |
| Summary ................................................................. 91 |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>V. DISCUSSION</td>
<td>93</td>
</tr>
<tr>
<td>Overview of Study</td>
<td>93</td>
</tr>
<tr>
<td>Discussion of Findings</td>
<td>95</td>
</tr>
<tr>
<td>Limitations of this Study</td>
<td>107</td>
</tr>
<tr>
<td>Recommendations for Further Research</td>
<td>109</td>
</tr>
<tr>
<td>General Recommendations</td>
<td>111</td>
</tr>
<tr>
<td>Concluding Remarks</td>
<td>116</td>
</tr>
</tbody>
</table>

LIST OF REFERENCES ....................................................................................... 117

APPENDIX A: Survey .......................................................................................... 125

APPENDIX B: Cover Letter to Principals ......................................................... 130

APPENDIX C: Principal Response Postcard ...................................................... 132

APPENDIX D: Letter of Support ....................................................................... 134

APPENDIX E: Pre-survey Introductory Postcard .............................................. 136

APPENDIX F: Cover Letter to Teachers ........................................................... 138

APPENDIX G: Follow-up Postcard .................................................................... 140

APPENDIX H: Final Follow-up Letter ............................................................... 142

APPENDIX I: Curriculum and Instruction Factor Analysis ............................. 144

APPENDIX J: Pressure and Attitude Factor Analysis ...................................... 147

APPENDIX K: Teacher Comments ..................................................................... 149

APPENDIX L: Institutional Review Board Approval ......................................... 166
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Student Level and School Level Item Summary</td>
</tr>
<tr>
<td>2</td>
<td>Sample Characteristics</td>
</tr>
<tr>
<td>3</td>
<td>Survey Return Summary</td>
</tr>
<tr>
<td>4</td>
<td>Initial Subscale Outline</td>
</tr>
<tr>
<td>5</td>
<td>Curriculum and Instruction Category Factor Analysis</td>
</tr>
<tr>
<td>6</td>
<td>Pressure and Attitude Category Factor Analysis</td>
</tr>
<tr>
<td>7</td>
<td>Comparisons of Subscale Reliability</td>
</tr>
<tr>
<td>8</td>
<td>Final Subscale Outline</td>
</tr>
<tr>
<td>9</td>
<td>Demographic Characteristics of Respondents</td>
</tr>
<tr>
<td>10</td>
<td>Curriculum and Instruction Subscales: Item Means and Standard Deviations</td>
</tr>
<tr>
<td>11</td>
<td>Curriculum and Instruction: Subscale Means and Standard Deviations</td>
</tr>
<tr>
<td>12</td>
<td>Curriculum and Instruction Subscales: Analysis of Variance</td>
</tr>
<tr>
<td>13</td>
<td>Sources of Pressure Subscales: Item Means and Standard Deviations</td>
</tr>
<tr>
<td>14</td>
<td>Sources of Pressure: Subscale Means and Standard Deviation</td>
</tr>
<tr>
<td>15</td>
<td>Sources of Pressure Subscales: Analysis of Variance</td>
</tr>
<tr>
<td>16</td>
<td>Attitude Subscales: Item Means and Standard Deviations</td>
</tr>
<tr>
<td>17</td>
<td>Attitude: Subscale Means and Standard Deviations</td>
</tr>
<tr>
<td>18</td>
<td>Attitude Subscales: Analysis of Variance</td>
</tr>
<tr>
<td>19</td>
<td>Items 54 and 55: Means and Standard Deviations</td>
</tr>
<tr>
<td>20</td>
<td>Items 54 and 55: Analysis of Variance</td>
</tr>
<tr>
<td>21</td>
<td>Open Response Comments Summary</td>
</tr>
</tbody>
</table>
ABSTRACT

THE PERCEIVED EFFECTS OF STATE-MANDATED TESTING
IN NEW HAMPSHIRE

by

Stephen A. Kossakoski

University of New Hampshire, May, 2000

The purposes of this study were: (1) to examine the perceptions of third grade teachers in the state of New Hampshire concerning the state's mandated testing program, and (2) to determine if there was a relationship between teachers' perceptions and their schools' socioeconomic status. The survey instrument designed for this study contained 61 closed response Likert scale items that measured teachers' perceptions regarding changes in curriculum and instruction, sources of pressure to improve test scores, and attitude toward the third grade test. A stratified random sample of 310 third grade teachers was selected to participate in this study. The percentage of students qualifying for free or reduced price lunches in individual schools was used as the stratifying variable. A total of 257 teachers returned surveys which translated into a response rate of 83%.

The analysis of data revealed that teachers believed that the test forced curricular and instructional alignment with the tested content, but also de-emphasized untested curricular content. Increased emphasis was reported in English language arts and mathematics while a decrease in emphasis occurred in science, social studies, art, physical education, and music. Teachers reported that they increased emphasis on test
preparation, activities using multiple choice formats, and activities that included higher
order thinking skills. The media was identified as the source of greatest pressure to
improve test scores. Test scores were not perceived as accurate indicators of student
achievement or instructional quality and there was little support for their use as a means
of comparing school quality. Teachers did agree that test scores were positively
influenced by the degree to which curriculum and instruction were aligned with tested
content. In general, however, teachers were not supportive of the test or the manner in
which test scores were used.

There were no significant differences among teachers’ perceptions when stratified
by the socioeconomic status of schools. However, there were significant differences
among teachers’ perceptions about the influence of socioeconomic status on test scores.
Teachers who taught in affluent communities were more likely to believe that
socioeconomic status did not affect test scores.
CHAPTER I

INTRODUCTION

The educational testing industry has become a lucrative business in the United States. Spurred on by legislation, states have sought the assistance of commercial contractors in order to aid in the development, implementation, and analysis of statewide testing programs that are now in place in 49 states. While Iowa remains the last state without a mandated testing program, its school districts are required to develop their own comprehensive improvement plans which may include the use of standardized test scores in grades four, eight, and eleven (Iowa Department of Education, 2000). Costs associated with state mandated testing programs were $165 million in 1997 and are expected to reach as high as $330 million in 2000 (Boser, 2000). However, these estimates are based solely on the direct costs associated with the purchase of tests from commercial vendors. Madaus and Raczek (1996) contended that these types of estimates do not include indirect costs and are, therefore, misleading. When indirect costs, such as administrative costs and the opportunity costs of time devoted to the tests by administrators, teachers, and students are factored in, the total cost for state and district testing programs may increase to as much as $22.7 billion. Madaus and Raczek stated that, in 1996, this estimate was equivalent to over twice the federal outlay for elementary and secondary education.
Why are mandated testing programs so popular? Tests have been a part of the fabric of schools since the widespread use of achievement tests began in the 1930s (U.S. Office of Technology Assessment, 1992). All adults who have attended school in the United States have experienced the anxiety of taking some type of test during their school career. For most Americans, tests have become such an expected and accepted method for quantifying the outcomes of learning that few question the reliability or validity of test scores, nor have they questioned the costs associated with implementing and administering these programs. While the use of tests go largely unquestioned by the American public, their role within schools has shifted over time. Tests were once used primarily by schools as methods of sorting students into appropriate academic programs. Since the 1960s, however, tests have become policy instruments that are used by schools, state, and national governments to determine the effectiveness of schools or programs (Madaus & Kellaghan, 1993). The constant barrage of negative reports concerning achievement levels in our nation’s schools has led to the belief that mandated testing programs, as an unbiased and powerful method of ensuring that improvement has taken place, must be implemented. Airasian (1988) described this evolution in the following manner:

Tests have moved from recorders of the effects of educational reform to initiators of reform, from passive instruments to intrusive devices that threaten test takers and seem to dictate the ends of the instructional system. Tests have moved from descriptive indices to certification devices.

The cyclical nature of the process which has led to the acceptance and popularity of tests as certification devices has been summarized by Salganik (1985): (1) an intolerable level of uncertainty about the quality of schools exits in the minds of parents and the
community, (2) parents accept test results over teachers' judgment since they believe that
tests tell them how much has been learned and how good schools are, (3) legislators
know that the public accepts test results as evidence of student achievement and will, in
turn, exert pressure on schools to improve, and (4) educators accept tests in order to gain
public support.

While mandated testing programs have been portrayed and accepted as a benign
technology for measuring student achievement, they may have, as most technologies do,
created unintended consequences that have affected both curriculum and instructional
practices within schools. Given the size and scope of current testing initiatives, relatively
little research has been conducted on the manner in which these programs have
influenced change in schools. In fact, while teachers are required to assume most of the
responsibility for improving student achievement their voices are seldom heard during
public discourse and their opinions are rarely solicited by scholars. This study has sought
to add to the knowledge base in this area by asking teachers in New Hampshire to
describe their perceptions regarding the impact of the state’s grade three testing program.

Background

In 1989, the New Hampshire State Board of Education decided that one of its top
priorities would be to develop an educational improvement and assessment program. As
a result of this action a 27-member steering committee was formed and, in 1991,
submitted a plan to the legislature which detailed the development of curriculum
frameworks and assessment strategies. In June of 1993, the New Hampshire legislature
enacted state law RSA 193-C that established the New Hampshire Educational
Improvement and Assessment Program (NHEIAP). Also in 1993, teams were formed to develop grade three curriculum and assessment frameworks for English language arts and mathematics. The curriculum frameworks were expanded to include all grade levels, with the addition of science and social studies, by the end of 1995. This new legislation required that all public school districts participate in the administration of the NHEIAP tests in grades three, six, and ten (New Hampshire Department of Education, 1998b).

The entire development of the NHEIAP was guided by the following five-step process: (1) define student outcomes at designated levels, (2) communicate standards to all educators, (3) help schools develop improvement and assessment plans, (4) design tests which will evaluate students' progress toward standards, and (5) assist schools in the use of assessment results in a manner which will improve academic achievement (New Hampshire Department of Education, 1995).

The first step in the development of the NHEIAP, the definition of student outcomes, was initiated in 1993 when teams comprised of teachers, administrators, parents, business representatives, community members, and policymakers were formed to develop grade three curriculum and assessment frameworks for English language arts and mathematics. The curriculum frameworks were expanded to include all grade levels with the addition of science and social studies by the end of 1995 (New Hampshire Department of Education, 1995). Since that time a number of addenda to the curriculum frameworks have been made.

The New Hampshire Department of Education (NHDOE) has addressed the steps relating to communication, the development of improvement and assessment plans, and assistance with the use of assessment results on a number of levels. Informational
meetings and workshops have been developed and sponsored by the NHDOE with the assistance of several statewide organizations. Staff at the department have been made available to school districts as consultants to address issues related to the NHEIAP. (New Hampshire Department of Education, 1995). The NHDOE has also developed an improvement planning toolkit and provided districts with test score data and analysis tools (New Hampshire Department of Education, 1999). Additionally, the NHDOE has assisted with Governor Jeanne Shaheen’s Best Schools Initiative, a program designed to support the intent of RSA 193-C and to provide districts and communities with the resources necessary to work on local improvement goals (Learning Innovations, 1999).

The remaining step in the NHEIAP development process was completed when an independent assessment contractor, Advanced Systems in Measurement and Evaluation, of Dover, New Hampshire, was hired by the New Hampshire Department of Education to design and implement the actual assessment tool, scoring program, and regional administration workshops. Development of individual items for all tests was assigned to content area committees comprised of teachers, curriculum experts, and other interested groups. The tests have been administered by teachers in mid-May of each school year and require a total of five hours for completion (New Hampshire Department of Education, 1998b).

Beginning in May of 1994, and again in 1995, all third graders were tested in English language arts and mathematics. Results for these tests were released by November of each year. By 1996, the plan called for grades six and ten to be added to the assessment program and tested in the areas of English language arts, mathematics, science, and social studies (New Hampshire Department of Education, 1995).
The stated purpose of the legislation that has guided the development of the NHEIAP and the grade three testing program is not to establish a statewide curriculum, but to "establish what New Hampshire students should know and be able to do and to develop and implement effective methods for assessing that learning and its application so that local decisions about curriculum development and delivery can be made" (State of New Hampshire, 1993b). While the comparison of schools and school districts is not a stated objective of the NHEIAP, it has become an issue due to the publication of ranked scores by local and statewide media. The Union Leader, New Hampshire's largest newspaper, has been publishing an alphabetical listing of schools and related scores since 1994 (New Hampshire Department of Education, 1994). This list included information on the number of students in the categories of advanced, proficient, basic, and novice. New Hampshire's Commissioner of Education, Elizabeth M. Twomey, wrote an article to accompany the test results in which she stated that the tests should be used only as a baseline measure for improving teaching and performance (Distaso, 1994). The first statewide list comparing school districts was published in November of 1997 by the Boston Globe (Kittredge, 1997, p 1).

In November of 1996 Twomey circulated a letter to all local school board members which contained a statement concerning the proper use of the NHEIAP test scores:

I am sure there will be questions raised about the third-year, end-of-third-grade results. People will analyze and interpret this data in many different ways. I cannot stress enough, however, that it is inappropriate to use the results alone to compare one school district or school to another. The purpose of NHEIAP is to give individual schools and school districts the information they need to develop local improvement plans and the ability to chart their educational improvement over time.
Although Commissioner Twomey cautioned against the use of the third grade tests as a method for drawing comparisons among schools, it continues to be an issue. Popham (1993) acknowledged this type of behavior and described mandated tests as instruments that create a high stakes environment for the educators involved. Furthermore, he identified mandated testing policies as measurement driven reforms that, "are premised on the fundamental notion that if people are to be judged according to certain types of criteria, they will try to excel with respect to those criteria" (p. 31). However, this assumes that the instrument being used to compare the effectiveness of schools has been designed for this purpose in order to provide educators with assessment data that are fair and just. The Detroit Free Press conducted an analysis of Michigan's state test results and found that poverty and other factors outside of the schools' control were strongly linked to results. The newspaper publicly acknowledged that their ranking of school districts, based on raw scores, was neither fair nor accurate and decided to eliminate the practice (Olson, 1998). In New Hampshire, however, public reporting of test scores continues with little caution provided to communities regarding the validity of school to school comparisons based on the NHEIAP test scores.

**Purpose of the Research**

The literature is rich with studies on the effects of testing on instruction, curriculum, and teachers. While these studies provide us with some information that can, no doubt, be generalized to many school environments, none can be said to accurately describe the manner in which a particular state's schools will be affected by a mandated testing program. In the state of New Hampshire, relatively little research has been
conducted on the effects of its state mandated tests. Hall (1999) has documented the
effects of class size, family income, adult education, and gender on the achievement of
third grade students. G. Cioffi and J. Carney (personal communication, November 3,
1997) conducted a study to determine the relationship between the NHEIAP English
language arts test and standardized measures of reading and language arts.
Additionally, data have been collected during the NHEIAP test regarding students'
reactions to a set of questions designed to describe how they prepare for test content.

The purpose of this study was to determine the perceived effects of state-
mandated testing on third grade teachers in New Hampshire. It has also provided
information regarding teachers’ support for this program. Survey research was used for
data collection and statistical analyses have been employed to determine if there is a
relationship between an indicator of socioeconomic status and changes in curriculum,
instruction, and support for the program. This study has begun to address the lack of
information currently available on the manner in which the NHEIAP third grade tests are
affecting New Hampshire’s schools. This study will also add to the knowledge base on
how the technology of testing has influenced the development and progress of our
educational system. The results of this study will be of interest to policymakers,
educators, researchers, and other parties interested in educational assessment issues.

Research Questions

Specifically, this research was designed to answer the following questions:

1. What are the perceptions of third grade teachers regarding the NHEIAP grade
three test’s influence on curriculum and instruction?
2. What is the level of support among grade three teachers for the NHEIAP grade three tests?

3. From which group, or groups, do grade three teachers feel pressure to improve their students' scores on the NHEIAP grade three test?

4. Is there a relationship between changes made to the curriculum and in instruction and the socioeconomic status of the school community as defined by the percentage of students who participate in the federal free and reduced price school lunch program?

5. Is there a relationship between teachers' support for the NHEIAP grade three test and the socioeconomic status of the school community in which they teach as defined by the percentage of students who participate in the federal free and reduced price school lunch program?

6. Is there a relationship between the pressure to improve their students' scores on the NHEIAP grade three test and the socioeconomic status of the school community as defined by the percentage of students who participate in the federal free and reduced price school lunch program?

Organization of the Study

The study is organized into five chapters. The first chapter consists of background information regarding the NHEIAP and the associated grade three test, the purpose of this research, research questions, and the definition of key terms. Chapter two is a review of the literature which consists of the historical background of the testing movement in general and research relating to the purpose of this study. Chapter three describes the methodology used to gather and analyze data. Chapter four presents the
data and the results related to each research question. Chapter five summarizes the findings of this study as they relate to the research questions and hypotheses and presents recommendations for future research. References and appendices follow chapter five.

**Definition of Terms**

**Criterion Referenced Test.** Criterion referenced tests determine a student’s level of achievement along a continuum ranging from no proficiency to perfect performance (Glaser, 1971). Students’ skills or knowledge are measured against a criterion or performance standard which provides information about what the student knows and is able to do. The results of criterion referenced tests illustrate the achievement level of a student, or group of students, and are not comparable among students or groups as are norm referenced tests (Popham & Husek, 1971). The NHEIAP grade three test is a criterion referenced test.

**High Stakes Test.** High stakes tests are designed to elicit certain instructional and curricular changes in schools (Airasian, 1988). The results of these tests are intended to be used as indicators of the quality of instruction within a school. Tests are considered high stakes tests if either one or both of the following statements are true: (1) the results determine whether the examinees are promoted or receive special instructional services, and (2) the results are assumed to reflect the quality of a school and its instruction (Popham, 1993).

**NHEIAP.** The acronym, NHEIAP, stands for the New Hampshire Educational Improvement and Assessment Program.
NHEIAP Grade Three Test. The NHEIAP grade three test refers to the state's mandated test which is administered annually to all third grade students. Students are tested in mathematics and English language arts.

Socioeconomic Status. Socioeconomic status is a term that is traditionally used to allow comparisons between or among groups of students based on indicators such as family income, parents' educational level, and parents' occupation (White, 1982). For the purposes of this study, socioeconomic status will be determined by the percentage of students who qualify for participation in the federal free and reduced price lunch program.

Standardized Test. A standardized test utilizes a consistent set of regulations for designing, administering, and scoring. Standardization is critical in order to guarantee that all students are tested under the same conditions which increases the validity and reliability of the instrument and resulting scores (National Center for Research on Evaluation, 2000).

State Mandated Test. Tests which are required, by legislative edict, to be administered in all schools to all students are considered state mandated tests. In some states, students with disabilities are either excused from the test or the test is modified in some manner. State mandated tests are not necessarily similar in type or in the content that they assess. The majority of states have developed their own tests that are aligned with curricular standards although a minority of states use commercially available standardized tests to measure student performance (Education Week, 1999).
CHAPTER II

REVIEW OF THE LITERATURE

Introduction

The origins of testing can be traced back over 1000 years to when the Chinese initiated a civil service examination for public officials, a practice that continued until 1905 (Black, 1998). The use of tests in American schools, however, dates back only 150 years when in 1845 Horace Mann, as Secretary of the Massachusetts State Board of Education, instituted the first written standardized examinations to classify students in order to increase efficiency and to monitor school systems. (Glaser & Silver, 1994; U.S. Office of Technology Assessment, 1992). Since that time, the popularity of testing in American schools has continued to increase. In fact, Gallup (1998) found that between 66 and 75 percent of those polled supported the use of a national test of public school students. While the popularity and acceptance of tests has increased, the actual technology of testing has changed little when consideration is given to the time and resources devoted to its development. It is commonly accepted that a single test score does not accurately reflect a student’s true understanding, nor can a set of scores be used to accurately compare the quality of instruction among schools, but still such practices persist. Sacks (1999) has characterized our fascination with testing in the following manner: “Like a drug addict who knows he should quit, America is hooked. We are a nation of standardized testing junkies” (p. 6).
What factors have led to testing becoming such an influential and popular part of the American educational culture? How do testing programs influence teachers, instruction, and the curriculum? What role do socioeconomic factors play in the influence of testing on schools? This review of the literature will address these questions by developing a historical context for testing in American schools. Furthermore, a review of research on the effects of testing on teachers, instruction, and curriculum will provide answers to the manner in which school culture has adapted to testing programs.

**Historical Context**

In each of his five State of the Union addresses, President Clinton has consistently mentioned the shift toward a global economy and linked our nation's success to the development of a workforce capable of competing in such an environment. According to the President, there is a direct link between our success as a nation competing in the global economy and the success of our public schools. In order to ensure success he has proposed the development of national academic standards and testing programs (Clinton, 1994; 1995; 1996; 1997; 1998). These recommendations reflect the belief that testing programs will inevitably lead to solutions of societal problems. Black (1998) suggested that the importance of testing in schools evolves as society changes. He asserted that societal changes result in technical challenges which, in turn, lead to the development of new testing programs.

Using Black's explanation of this phenomenon, it can be said that the current call to implement both national and statewide assessment reforms are predictable solutions to...
the challenges created by our rapidly changing "information age" society.

Information age businesses require a different set of academic skills that were not emphasized in traditional industrial age schools. Thus, we are presented with a new technical challenge: to educate an increasingly diverse population with the skills required by information age businesses. When the technical challenge of developing new skills within our workforce exerts pressure on policymakers, they initiate the implementation of testing programs with the hope that these programs will act as levers of reform and accountability in our public schools. Similar patterns of technical challenge have resulted in three major cycles of testing in schools: (1) immigration, war, and the industrial age; (2) Sputnik and civil rights; and, (3) the global economy and the information age.

Immigration, War, and the Industrial Age

Societal change, as a precursor to new waves of testing in schools, can serve as a method of organizing and understanding how testing has become such a powerful tool within our educational culture. The set of societal changes that stimulated the first major cycle of testing in schools occurred around the early part of the nineteenth century. Resnick and Resnick (1985) identified the period between 1880 and 1930 and the interwar period between 1917 and 1939 as stages of time where enthusiasm for testing was developed due to the perceived successes of the Army testing program. Throughout this period the industrialization of American urban and rural communities created a need for a more educated workforce. In the period between 1880 and 1930 the population served by American secondary schools increased from 10 percent of those in the 14 to 16 year old category to 70 percent (Resnick & Resnick, 1985). In the spirit of the "cult of efficiency in industry" (Resnick & Resnick, 1985, p. 11) policymakers and educators
sought methods for dealing effectively with a suddenly large and diverse population of students. During this time, secondary schools experienced a fivefold increase in population due to a dramatic increase in the immigrant population coupled with natural increases in population and the addition of compulsory school attendance laws (Resnick, 1982).

Another societal change that must be considered during this time was the move from an agrarian society to an industrial-based economy and its requirements for a literate workforce (U.S. Office of Technology Assessment, 1992). The United States Army also experienced a technical challenge which led to an increase in testing when faced with sorting and training a large number of new recruits during World War I. While these factors were bringing about rapid change in society, pioneers in the field of mental measurement were disseminating theories and methods for classifying students by heredity and intelligence.

In 1905, Alfred Binet was hired by the French government to create an instrument for sorting students in a manner that would maximize educational resources. This project resulted in the development of a 30 item scale which was designed to determine which students should be identified as "retarded" and placed in special classes. Binet's work was strongly influenced by Sir Francis Galton who advanced the theory that intelligence was determined by genetic inheritance (Black, 1998; Colvin, 1923; Resnick, 1982). Binet had strong reservations about using test data generated by his intelligence scale to classify students and to then reduce cognitive capacity to a single number, a concern that would go unheeded when the Binet scale was introduced in American schools (U.S. Office of Technology Assessment, 1992).
Henry Goddard was credited with bringing Binet's work to American schools. In 1910 he administered his version of the Binet scale to 1,547 elementary students in a New Jersey school district. Goddard used his results to advance the theory that intelligence was a variable that was normally distributed within the school-age population. He argued that his results were also supported by Galton's theory that certain anthropomorphic variables also fit the normal curve. Goddard used the results from this test and his resulting theory of intelligence to assist in the development of the first special education legislation that required separate classes for blind and deaf students and for those whose intelligence was determined by his test to be more than three years below normal levels. However well received, the quality of Goddard's research was questioned since cultural bias, economic background, and language proficiency were not considered as factors that might influence test results (Zenderland, 1998).

While Goddard was recognized as the first to apply Binet's work in a school setting in the United States, it was Lewis Terman, a Stanford psychologist, who was credited with revising Binet's work and promoting the use of intelligence testing throughout the country. Terman, like others of his time, based his work on the belief that intelligence is a genetically inherited constant and that his tests could be used to improve society by identifying and controlling the reproduction of "defectives" and by restricting the influx of immigrants (Chapman, 1988). Terman's version of Binet's work appeared as the Stanford-Binet revision in 1916. This instrument included 90 test items and was designed to identify sub-normal, normal, and superior students and adults. The Stanford-Binet test was standardized to a mean of 100 and a standard deviation of 15 points. The Stanford-Binet became the standard for judging intelligence. Other commercial test
developers validated their tests by proving that there was a strong correlation between their test results and those of the Stanford-Binet. Of course, it was assumed that the Stanford-Binet scale was an accurate method for determining intelligence (Gould, 1996).

Events that occurred during World War I led to the institutionalization of tests as an accepted method for sorting people. Soon after the United States entered the war in 1917, Terman was recruited by Robert Yerkes, president of the American Psychological Association, to work with a group of psychologists to develop intelligence tests for the army. In addition to Terman, Yerkes recruited five other psychologists including Henry Goddard. While the group's original goal was to develop an instrument to identify those who were unfit for service, it was amended to include testing and sorting of all army recruits (Chapman, 1988). Yerkes' group of psychologists developed two tests, the Alpha for the normal population and the Beta for the subnormal population, in approximately one month (U.S. Office of Technology Assessment, 1992). The program fulfilled its mission in 1919, having administered tests to over 1.7 million enlisted men and officers (Chapman, 1988). This was a truly high-stakes test as over 1 million soldiers were sent into combat based on their test scores (U.S. Office of Technology Assessment, 1992).

To the public, the army testing program appeared to be a success and became the launching point for the use of standardized testing programs in schools throughout the country. However, the reality is that the army's testing program was not well received by senior officers, and, in spite of the large number of soldiers tested, there was little support for the psychologists' belief that their tests were valid measures of native intelligence (Minton, 1988). While the army's testing program did not revolutionize the organization, the use of intelligence tests by the army during World War I helped revolutionize the
testing industry and create a great deal of credibility for the emerging testing industry and the science of psychology whether deserved or not.

The perceived success of the army testing program and the subsequent publication of reports by Yerkes and his associates influenced both social policy and school testing programs. In Yerkes’ summary report of the test data, he identified three items that influenced social policy of the time: (1) the average mental age of those tested was slightly above a moron, (2) immigrants fell into the category of moron, and (3) the Negro earned the lowest scores. (Gould, 1996). In 1923, Carl Brigham published *A Study of American Intelligence* in which he used the army test data to support the decline of American intelligence.

According to all evidence available, then, American intelligence is declining, and will proceed with an accelerated rate as the racial admixture becomes more and more extensive. The decline of American intelligence will be more rapid than the decline of the intelligence of European national groups, owing to the presence here of the Negro. These are the plain, if somewhat ugly, facts that our study shows. The deterioration of American intelligence is not inevitable, however, if public action can be aroused to prevent it. There is no reason why legal steps should not be taken which would insure a continuously progressive upward evolution. (p. 210)

The actual influence that these publications had upon public debate is difficult to assess, but the Immigration Restriction Act of 1924 utilized the army test data as a basis for setting quotas for admitting individuals from nations of inferior intelligence. These quotas are estimated to have prevented millions of European immigrants from entering the United States between the end of World War I and the beginning of World War II (Gould, 1996).

A public debate occurred on the pages of the “New Republic” between Terman and journalist Walter Lippmann. Lippman questioned the validity of intelligence tests by
questioning Terman's refusal to acknowledge cultural and environmental influences upon test results. "He is testing the complex result of a long and unknown history, and the assumption that his questions and his puzzles can in fifty minutes isolate abstract intelligence is, therefore, vanity" (Lippmann, 1976, p. 29). Lippman was not alone in his criticism of intelligence testing. In 1930, Carl Brigham reversed his views that he had previously espoused in A Study of American Intelligence.

This review has summarized some of the more recent test findings which show that comparative studies of various national and racial groups may not be made with existing tests, and which show, in particular, that one of the most pretentious of these comparative racial studies—the writer's own—was without foundation. (p. 165)

The intelligence testing movement attained its greatest popularity during the 1920s (Chapman, 1988). However, severe criticism from psychologists and school administrators helped move achievement testing to the forefront as a method of categorizing students. While schools continued to experience rapid growth, educators were eager to adopt testing programs that would assist in the sorting of students without adding the burden of the controversy that plagued the intelligence testing movement.

While the intelligence testing movement was growing, the achievement testing industry was in its infancy and just emerging as an established testing instrument. Joseph Mayer Rice, a medical doctor and publicist, pioneered the use of achievement tests when he tested over 100,000 students between 1894 and 1896 in spelling, penmanship, English composition, and mathematics. Through the publication of his results he hoped to spark interest in the development of objective standards that would be used to measure student achievement and the effectiveness of instruction (Chapman, 1988; Resnick, 1982; U.S. Office of Technology Assessment, 1992). Edward Thorndike advanced Rice's scientific
approach to assessment, in the period between 1908 and 1916, when he developed standardized achievement tests in language, reading, drawing, spelling, handwriting, and mathematics (Chapman, 1988; U.S. Office of Technology Assessment, 1992).

Thorndike's work included an adaptation of Galton's statistical methods to the educational realm. Through his teaching and writing at Columbia, Thorndike had tremendous influence upon the use of scientific measurement and testing in schools. His philosophy helped establish a culture of scientific measurement within schools.

Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quantity as well as its quality. Education is concerned with changes in human beings; a change is a difference between two conditions; each of these conditions is known only to us by the products produced by it--things made, words spoken, acts performed, and the like. To measure any of these products means to define its amount in some way so that competent persons will know how large it is, better than they would without measurement. To measure a product well means to so define its amount that competent persons will know how large it is, with some precision, and that this knowledge may be conveniently recorded and used. (Thorndike, 1918, p. 16)

The educational testing movement reached maturity in the 1930s when it entered a "stage of critical caution." During this time the exaggerated claims of educational measurement experts were tempered (Linden & Linden, 1968, p. 83). E. F. Lindquist, a professor of education at the University of Iowa, developed a battery of tests that reflected the testing industry's maturity. The Iowa Test of Basic Skills (ITBS), for grades 3 through 8, and the Iowa Test of Educational Development (ITED), for grades 9 through 12, were originally developed to evaluate students and schools within the state of Iowa (U.S. Office of Technology Assessment, 1992). The ITED became available to schools outside of Iowa in 1935, while the ITBS was released in 1942 (Kohn, 1977). These achievement tests were designed to measure understanding and application of knowledge rather than the recall of basic facts and were, therefore, not tied to the content of specific
textbooks. The Iowas spawned the development of other tests that were also designed to
diagnose the abilities of students rather than classifying students. The Progressive
Achievement Tests (later renamed the California Achievement Tests), the Metropolitan
Achievement Tests, and the Stanford Achievement Tests were the most popular versions
of this new type of achievement test (Linden & Linden, 1968).

Lindquist also transformed the testing industry on another level by designing a
computerized scoring machine that was capable of scoring 40,000 tests per hour. This
innovation made it profitable for commercial enterprises to enter the business of large-
scale group testing and scoring (Kohn, 1977).

Sputnik, and Civil Rights

World War II had relatively little impact upon the popularity of testing in schools
since testing programs were already firmly entrenched in the educational culture of
schools across the country (U.S. Office of Technology Assessment, 1992). The technical
challenges that contributed to the second major cycle of educational testing included the
Soviet Union's perceived superiority in space exploration and the civil rights movement
(Glaser & Silver, 1994; Madaus & Raczek, 1996; U.S. Office of Technology Assessment,

The implementation of testing programs due to these societal changes marked an
important change in the use of tests. Prior to the 1960s tests were used by local school
districts primarily to sort students. However, the powerful societal changes that occurred
in the early 1960s marked the beginning of a shift toward the use of tests as policy
instruments (Madaus & Kellaghan, 1993). This change in the use of tests went beyond
policy and into the classroom, where tests were used to judge the quality of educational
programs. Salganik (1985) states that the professional judgment of educators as arbiter of school quality gave way to the use of test results.

However, in the new climate of uncertainty about the adequacy of the schools (and thus about the competence of educators), testing introduced a welcome simplicity to the task of restoring both educational quality and public confidence in the schools. Few people were willing to argue with the use of tests as a means of insuring quality control. (p. 608)

Madaus (1985) identified five factors that led to this shift from tests as pedagogical tools to tests as instruments of policy development. First, as previously mentioned, state and federal policymakers were concerned about the quality of education. This led to discussions about educational outcomes and, eventually, to assessments of educational success and failure. Second, tests were implemented to measure the results of federally funded curriculum programs that were established in the 1950s and 1960s. Third, compensatory funds were sought by racial and ethnic groups based on test score discrepancies. Fourth, the National Assessment of Educational Progress (NAEP) was developed and implemented in the 1960s to provide data for public discussion and to highlight problems and progress. Fifth, the Coleman report helped to redefine the effectiveness of schools in terms of how well low-income students performed on standardized tests.

The nature of achievement testing changed in the early 1960s when Glaser (1971) redefined the concept of criterion-referenced tests by differentiating between norm-referenced and criterion-referenced tests. He explained the popularity of norm-referenced tests as being, "brought about by the preoccupation of test theory with aptitude, and with selection and prediction problems" (Glaser, 1971, p. 9). He believed that there was a need to specify, "minimum levels of performance that describe the least amount of end-of-course competence the student is expected to attain" (Glaser, 1971, pp.
9-10). By the 1970s criterion-referenced tests were popular in schools throughout the country with support from professional teacher organizations. Statewide minimum competency tests were examples of criterion-referenced tests that were developed during this period (Peters, Wixon, Valencia, & Pearson, 1993).

The Global Economy and the Information Age

The current cycle of educational testing can be traced to changes in technology that have led to the global economy and the information age. On August 26, 1981 the Secretary of Education, Terrel Bell, created the National Commission on Excellence in Education to study the quality of education in the United States. The Commission's report, A Nation At Risk, reported that the nation's educational system was in decline and in need of urgent attention. This conclusion was based primarily on the comparison of test results over a number of years. Among the report's five recommendations was a call to, "adopt more rigorous and measurable standards" (National Commission on Excellence in Education, 1983).

The publication of A Nation At Risk initiated an "era of criticism and change in U.S. education" (Finn & Rebarber, 1992, p.175). The report was published during a time of great anxiety about the condition of education in the United States and the country's place in the emerging global economy (Finn & Rebarber, 1992; Murphy, 1992b; Peters et al., 1993). The report created action at all levels of government which finally led to states imposing greater control over local school districts. Among the changes implemented at the state level were: (a) increased graduation requirements, (b) regularly scheduled testing programs, (c) sanctions for poor performance, and (d) increased standards for teacher certification (Peters et al., 1993).
Murphy (1992a) identified three waves of improvement reforms that took root in the 1980s and have continued into the 1990s. The first wave (1982-1985) of reform was strongly influenced by the changes and initiatives outlined in *A Nation At Risk*. The second wave (1986-1989) of reform was initiated by those who believed that the first wave reforms were weak and had led only to incremental change. Second wave reforms included: (a) the professionalization of teaching, (b) the development of decentralized and participatory management systems, and (c) the development of at-risk and equity programs to address gaps in first wave reforms. The third wave (1988-present) of reform has centered on the development of the cognitive abilities of students. President Bush is credited with initiating the first national reform program of this era when he and the National Governors' Association met to develop a set of national education goals during an educational summit in 1989. These goals were eventually crafted into legislation and signed into law by President Clinton in 1994 as the Goals 2000: Educate America Act (U.S. Department of Education, ). The public discussion surrounding these events and the resulting funding from the Goals 2000 legislation has led to the development of aligned curricular standards and assessment.

Third wave reforms have led to the development of standards by national and statewide groups. Virtually all states are currently developing, or have developed, standards for key curricular areas. Along with standards, most states have developed statewide testing programs to assess the achievement level of students. Forty-nine states had developed assessments by the end of 1999. More than 66 % of these states have developed their own criterion referenced tests aligned with curricular standards. However, at least 62 % of all states also administer some sort of norm referenced test to
their students. Of the 47 states that report using statewide tests, 70% require students to submit a writing sample as part of their assessment program. Only 32% of the states report the use of performance assessments (e.g., writing prompts, open response items), while three states, Vermont, Kentucky, and New Mexico require the use of portfolios (Wolk, 1997).

The Influence of Testing

Many effects, both positive and negative, have been attributed to the use of testing in schools. Madaus (1988) summarized the effects of high-stakes tests in the following seven principles:

1. The power of tests to influence educators and schools is based on the perceived importance of test scores.
2. Tests will alter and corrupt that which it is supposed to measure in direct proportion to the number of times that they are used as decision-making instruments.
3. Teachers will teach to the test if important decisions are connected to test scores.
4. Over time, high-stakes test content will define the curriculum.
5. Teachers will change instruction to mirror the types of questions used on high-stakes tests.
6. Society will adopt tests as the major goal of schooling if important decisions are tied to test results.
7. The agency that controls the tests will gain control of the curriculum.
These principles highlight two significant adaptations that schools and their communities undergo as a result of the pressure exerted by high-stakes tests. First, school officials and teachers develop a sense of importance based partly on the perceived significance of the tests and the types of decisions that are connected to test results. This sense of importance is further increased when society, and the local community, embraces increasing test scores as the most important objective of schooling. Second, instruction and curriculum are altered in direct proportion to the level of perceived significance associated with the tests. Implicit in this decision is the understanding that curriculum is now controlled by the organization that governs the testing program. These adaptations will be used to organize and analyze the research on testing in this review of the literature.

**Perceived Importance of Testing**

According to Schein (1985) culture can be defined as:

A pattern of basic assumptions—invented, discovered, or developed by a given group as it learns to cope with its problems of external adaptation and internal integration—that has worked well enough to be considered valid and, therefore, to be taught to new members as the correct way to perceive, think, and feel in relation to those problems. (p. 9)

The previous discussion on the three major cycles of educational testing can be examined using Schein's definition of culture to assist in understanding why tests have become such a powerful part of educational culture. As each cycle of reform has appeared and faded, policymakers and educators have increasingly relied upon tests to solve problems. For example, immigration and compulsory education legislation have led to dramatic increases in student population. The influx of new students brought with them a new set of needs that schools were not equipped to handle. The response has been to find a technology to sort students into groups of similar abilities or needs. Using tests
for this purpose has firmly embedded them into educational culture. While many have criticized the use of tests as sorters, they have worked, in Schein's words, "well enough to be valid" (p. 9). Most, if not all, educators have been deeply involved as participants in, and advocates of, the test culture, first as students, and then later as administrators of tests themselves. These experiences have taught educators, again, in Schein's words, "the correct way to perceive, think, and feel in relation to these problems" (p. 9). As such, there is evidence that testing's role as an unquestioned instrument for determining ability or achievement is changing the educational culture.

Madaus (1985) described the implementation of mandated high-stakes testing programs as an effort to use tests as policy instruments rather than as one of many indicators of student achievement. This cultural shift from a useful "classroom tool to infallible arbiter" (Madaus, 1985, p. 617) has caused many teachers to question the validity of testing's new role. Phi Delta Kappa has conducted a series of four nationwide polls (Elam, 1989; Gallup, 1985; Langdon, 1996; Langdon, 1997) to determine teachers' attitudes toward issues affecting public schools. Questions for each of the teachers' polls were based on data gathered in Phi Delta Kappa/Gallup polls of public attitudes toward the schools (Langdon, 1996). Testing was among many issues explored in these polls. In each of the first three polls teachers were asked if promotion from grade to grade should be based on an examination. In the 1985 poll 52% of the respondents were opposed to this practice. Opposition increased slightly, to 59%, in the 1996 poll. However, when teachers were asked if students should be required to pass an exam in order to qualify for graduation from high school the trend was reversed. In the 1984 poll 45% of the respondents did not support graduation exams, but by 1996 opposition had decreased to
29%. Questions regarding testing were not included in the most recent poll (Bracey, 1999).

The Phi Delta Kappa polls suggest that teachers are not in agreement on the proper role of testing. While there is strong support for the use of graduation tests, teachers, in general, do not support the use of tests for other high-stakes uses such as grade to grade promotion or assessing the quality of school. Furthermore, a majority of teachers also believe that there is too much emphasis on achievement testing. These results support the notion that the role of high-stakes testing within educational culture is in a state of change.

While the role of high-stakes tests and educators’ opinions about them may be evolving, public support for the high standards associated with the tests remains strong. In the most recent Phi Delta Kappa/Gallup poll (Rose & Gallup, 1999) of the public’s attitude toward public schools, 72% of the respondents to the poll believed that stricter standards must be in place in order to mitigate social promotion from grade to grade. Why does the public at large support the standards movement and their associated mandated high-stakes testing programs? Airasian (1988) posited that testing programs have been socially validated because they reflect current social values and norms. He has identified three ways in which mandated tests symbolize current norms and values. First, they provide order and control over the curriculum since such control is perceived to be weak. Second, mandated tests are seen by the public as objective and accurate measures of important educational outcomes. Third, the public believes that such programs symbolize the values and outlook they hope will be cultivated in their local schools. The popularity of mandated testing and the public reporting of scores by the media is creating
a culture where “test scores and passing rates on high-stakes test are becoming the educational equivalent to the Dow Jones stock average” (Airasian, 1988, p. 311).

The popularity of the standards movement and mandated testing programs has resulted in new demands being brought to bear on schools and, ultimately, teachers. Corbett and Wilson (1991) compared the impact of mandated testing programs on schools in Pennsylvania and Maryland. Their research included a three phase design: (1) interviews were conducted with 250 central office administrators and a variety of school personnel at 6 sites in each state; (2) interview results were then used to design a questionnaire; and (3) survey results were used to develop follow-up questions. Follow-up interviews were held at 11 of the 12 sites contacted in phase one of this research. One representative from each role group (central office, administrators, principals, and teachers) was asked to complete the phase two questionnaire in Pennsylvania (227 out of 501 school districts elected to participate). The school districts in Maryland were fewer in number but larger in size. Three role group members from each participating school district in Maryland were asked to complete the phase two questionnaire (23 out of 24 school districts elected to participate). Qualitative results were analyzed by constructing data summary charts and then looking for themes. Analysis of the quantitative data was conducted on two levels: (1) within state variations were analyzed using multiple regression techniques, and (2) cross state data were examined for differences using analysis of variance statistics.

Corbett and Wilson (1991) found that the pressure to improve test scores came from parents, the school board, and the local newspaper. Pressure was not exerted to change practice, but only to raise scores. In addition to exerting political pressure to
improve test scores, these same constituencies began to equate test scores with the quality of instruction provided by teachers. The result was that teachers altered practice without examining the educational value of their changes, but only the effect upon test scores. Corbett and Wilson also found that for many educators, increased pressure brought about a "crisis mentality" (p. 133) which resulted in specific and narrow changes in instruction and curriculum. These changes included the use of fewer instructional strategies, a narrowed curriculum within courses, and fewer course offerings.

Smith (1991) also addressed the pressure related to the publication of test scores in a qualitative study of two school districts in Arizona. Teachers were observed and interviewed over the course of more than 15 months. Participants reported pressure from the publication of test scores which resulted in feelings of guilt, shame, embarrassment, and anger. Results indicated that it was not only low performing schools that reacted to pressure but also top scoring districts that used scores as a shield to deter outside influence. Smith found that it wasn't necessarily the tests themselves that created pressure but the political manner in which the test results were used.

Pressure from the media and other stakeholder groups has not resulted in entirely negative effects. Many studies (Corbett & Wilson, 1991; Koretz, Mitchell, Barron, & Keith, 1996a; Koretz, Barron, Mitchell, & Stecher, 1996b; Madaus, 1988; Smith, 1991; Stecher, Barron, Kaganoff, & Goodwin, 1998) support the premise that mandated testing programs will cause schools to align their curriculum with embedded curricular topics. Stecher, Barron, Kaganoff, and Goodwin (1998) surveyed a stratified random sample of 400 elementary and middle school teachers in Kentucky during the 1996-1997 school year to assess the effects of Kentucky Instructional Results Information System (KIRIS)
assessments on classroom practices. Schools were stratified based on low, medium, and high scores received between 1992 and 1996. Principals at each of the selected schools were contacted to provide teachers' names and grades taught. The overall response rate from the original sample of 400 teachers was 70%. In reporting the results of this research, the authors decided to focus on sizeable differences in percentages (approximately 20%) as a guide for discussing statistical significance. Teachers reported that they increased the time associated with KIRIS curriculum standards in writing and mathematics. One caution noted by the researchers is the fact that many teachers reported an increase on the emphasis in all subjects. Possible explanations are that increases have resulted from the integration of all subject areas, or teachers may have gained instructional time by de-emphasizing non-academic activities in their classrooms.

Herman and Golan (1991) surveyed elementary school teachers from 11 school districts in nine states to determine the effects of standardized tests on teachers and learning. Data were received from 341 teachers representing 48 schools in nine states representing each geographical region of the United States. Correlation matrices and analysis of variance statistics were used to determine the relationships between variables. Teacher-reported percentages of students receiving Chapter I services was used as a proxy for socioeconomic status. Teachers reported feeling strong pressure from the media and district administrators to improve test results and moderate pressure from principals, teachers, parents, and the local communities. Job satisfaction remained moderate to strong despite the pressure that was imposed by the stakeholder groups. However, teachers did not believe that testing was helping schools improve and felt that tests did not emphasize curriculum that should be taught to students. Teachers in schools
with improving test scores reported feeling more pressure from their community than teachers in schools with steady or decreasing scores.

**Modification of instruction and curriculum**

As previously discussed, Madaus (1988) identified four direct effects of testing on instruction and curriculum: (1) teachers will teach to the test when important decisions are connected to test scores, (2) instruction will be changed to mirror the types of questions used, (3) curriculum will be defined by the content of high stakes tests, and (4) the agency that controls test content controls the curriculum. A number of researchers have studied the relationship between testing and these principles.

Resnick and Resnick (1992) have addressed the evolution and motivation for teaching to the test:

> Even in school districts where official policy is against 'teaching to the test,' if considerable attention is paid in the press or elsewhere to test scores, teachers will gradually adapt their teaching to the tests. Although many tests are held secure and the items varied from year to year, a teacher who administers the tests yearly has an opportunity to see what is being tested and to adjust class practice to match the test. This process is often exaggerated in school districts with knowledgeable testing directors and curriculum specialists who encourage—even if only implicitly—adaptation to the test in order to demonstrate publicly students' rising academic performance. (p. 57)

The potential problems associated with teaching to the test are potentially serious enough to have been addressed by the Committee on Appropriate Test Use in their report titled: **High Stakes, Testing for Tracking, Promotion, and Graduation** (Heubert & Hauser, 1999). In one of many recommendations relating to high stakes tests, the committee states that there should be a balance between test preparation and "excessively narrow preparation which will invalidate outcomes" (p. 280).
Advocates of measurement driven reform defend teaching to the test as a method of ensuring that mandated curricula and methods of instruction are implemented. Popham (1993) promotes mandated testing as a method of reform based on the premise that "if people are to be judged according to certain types of criteria, they will try to excel with respect to those criteria" (p. 34). The danger, however, is that the focus on test scores narrows the curriculum to the exclusion of activities and topics not covered on the test.

Closely related to teaching to the test are changes in instruction that are implemented as a result of mandated testing programs. These changes can be divided into two broad categories: (1) the use of test preparation activities, and (2) the effects on teaching in general. Teachers have logically determined that the amount of time devoted to test preparation is related to improving test scores. Test preparation activities most commonly include: teaching test-taking skills (Madaus, West, Harmon, Lomax, & Viator, 1992); drill, review, and pull-out remediation (Corbett & Wilson, 1991); and practice tests (Koretz et al., 1996a). Herman and Golan (1991) reported that a significant amount of time was spent in test preparation activities. For example, most teachers spent a week or more preparing for the tests by assigning worksheets with test-like items and teaching test taking strategies. In fact, many studies describe similar examples and confirm that significant instructional time is devoted to test preparation activities in high stakes testing environments (Corbett & Wilson, 1991; Koretz et al., 1996a; Koretz et al., 1996b; Madaus et al., 1992; Smith, 1991).

High stakes testing practices have had an effect upon how teachers view their role. Corbett and Wilson (1991) reported that the pressure exerted by stakeholder groups...
in response to mandated testing programs has led to a crisis mentality which results in the use of quick solutions and narrowed instructional strategies. Smith (1991), in her qualitative study of two Arizona elementary schools, found that published test results created pressure to improve scores which led to test-like teaching strategies. Results indicated that teaching was deskilled, or, as described by Smith (1991),

...the image we project of teachers in the world after testing reform is that of interchangeable technicians receiving the standard curriculum from above, transmitting it as given (the presentation manual never leaves the crook of their arms), and correcting multiple-choice responses of their pupils. (p. 11)

Two effects of mandated testing upon the curriculum emerge from studies conducted on the topic: (1) the curriculum is narrowed, and (2) curriculum is aligned with test content. Curriculum is narrowed through planned changes in content and course offerings (Corbett & Wilson, 1991) or through an unconscious de-emphasis of untested content (Koretz et al., 1996a). Smith (1991) also found that content that was not included in mandated tests was discarded, but also added that administrators influenced which parts of the curriculum were either deleted or emphasized. Jones, Jones, Hardin, Chapman, Yarbrough, and Davis (1999) surveyed a stratified random sample of 470 elementary school teachers in North Carolina to determine the impact of high-stakes testing on teachers and students. Their results demonstrate that a severe narrowing of the curriculum has occurred. Teachers stated that especially in schools that are not performing satisfactorily, “a visitor would have to look hard to see science or social studies lessons taught at all” (p. 201).

While parts of the curriculum may be de-emphasized or discarded, curriculum alignment may result in renewed emphasis being placed on intact content or altogether new content being added to the curriculum. Koretz, Mitchell, Barron, and Keith (1996a)
studied the perceived effects of the Maryland School Performance Assessment Program (MSPAP). A purposeful sample of four school districts (selected for population characteristics) was combined with a random sample of four additional school districts out of the state’s 24 school districts. Within the sample, 112 principals, 113 grade five teachers, and 111 grade eight teachers responded to either mail or telephone surveys. Because of the small sample size and large data set a descriptive approach was used for data analysis. The results showed that 40% of those teachers participating in the study emphasized alignment of their curriculum with tested content, while 85% of the principals in the study believed that curricular alignment was being emphasized. But, when asked about untested material approximately 69% of teachers believed that untested material was de-emphasized while 11% of the principals reported de-emphasizing untested material.

Other researchers have reported similar results regarding the alignment of curriculum with mandated test content. Stecher (1998) found that Kentucky teachers continued to utilize traditional curriculum, but that the state standards emphasized in the mandated testing program were taught with increasing frequency. Herman and Golan (1991), in their study of mandated testing in nine states, concluded that the sequence in which curriculum was taught was adjusted to match the content of mandated tests. Teachers also reported that they look at previously administered tests in order to guarantee that all test content will be covered in their curriculum. As the results of these studies indicate, mandated tests have become efficient mechanisms for state control of curriculum and instructional strategies.
School Funding & Achievement

Research on the relationship between economic resources and student achievement can be studied from two perspectives: (1) the resources allocated to schools, or (2) the resources available to individual students' families. Equality of Educational Opportunity (Coleman et al., 1966) was published by the United States Department of Health, Education, and Welfare and is considered the seminal work regarding the relationship between resources allocated to schools and academic achievement. The authors of the report concluded that schools had little influence on the achievement levels of students. The data also indicated that there was a significant relationship between school-to-school variations in student achievement and family background. Factors included in the measure of family background included urbanism of background, parents' education, structural integrity of the home, family size, items in home (e.g., television, telephone, record player, etc.), reading material in the home, parents' interests (e.g., discussions about school, reading to children), and parents educational desires for their children. This study, also known as the Coleman Report, introduced the educational production function model for comparing school inputs and outputs. As defined by Mosborg (1996), "a production function, educational or otherwise, is a mathematical formula that describes the maximum level of outcome possible from alternative combinations of inputs—bang for the buck" (p. 5).

Hanushek (1986) may well be the most ardent proponent of production function studies in education. However, he has acknowledged that the results of the Coleman Report are "commonly held to be seriously flawed, and its importance is more in terms of intellectual history than insights into schools and the educational process" (p. 1150). But,
at the same time, Hanushek concurred with the Coleman Report by concluding that there is no evidence of a strong relationship between school expenditures and student achievement in a review of 147 production function studies. Studies qualified for inclusion in this study if they met three criteria: (1) published in a book or refereed journal, (2) included an objective measure of student outcome which related to family and school characteristics, and (3) included information regarding the statistical significance of relationships. Vote counting methodology, in which the results of individual studies were tabulated by positive or negative sign and statistical significance was used to analyze data in this study. The characteristics with the most votes were assumed to be significant (Hedges, Laine, & Greenwald, 1994). Based on the results of this analysis, Hanushek argues that teacher/pupil ratio, teacher education, teacher experience, teacher salary, and expenditures per pupil have no significant correlation with student achievement. Hanushek stated that, based on the results of his research, “schools are economically inefficient, because they pay for attributes that are not systematically related to achievement” (Hanushek, 1986, p. 1166). If this is true, it is difficult to understand how schools could become economically efficient without ending the expenditure of all funds, since Hanushek’s variables encompass a very large percentage of all school expenditures.

Production function studies and methodology have been criticized for a number of reasons. Baker (1991) questioned Hanushek’s methodology on the basis of logical and methodological errors. He argues that Hanushek should have looked for relationships between expenditures and achievement gains. Baker defines gains as the improvement seen by the end of a school year. Hanushek addressed achievement levels, which Baker
believed better described differences in the socioeconomic status (SES) of students’ families. Baker performed a simple analysis of Hanushek’s data and found that 65 of the 147 studies addressed expenditures and achievement. Of these 65 studies 13 showed significant positive relationships between expenditures and achievement, while three reported a significant negative relationship. Baker notes that these three studies equal the chance expectation of an error occurring. A chi-square analysis was applied to positive and negative relationships between spending and achievement and determined that there was a significant positive relationship. This analysis of Hanushek’s data confirmed Baker’s contention that there is a positive relationship between increases in revenue to schools and increases in student achievement. Baker summarized Hanushek’s results in the following manner, "...there are so many serious problems with Hanushek’s analysis that no one should take it seriously" (p. 630).

Hedges, Laine, and Greenwald (1994) also reanalyzed Hanushek’s data and corroborated Baker’s assertions by concluding that there is significant support for positive effects from increased resources. Hedges et al. criticized Hanushek’s research based on the weakness of vote counting as a method of analysis. The authors identified the following weaknesses in support of their claim: (1) the magnitude of a relationship cannot be determined; (2) vote counting is a weak method for determining effects and is prone to Type II errors; and, (3) the probability that no effect will be found increases with the number of studies analyzed. Hedges et al. reanalyzed Hanushek’s data by using accepted meta-analysis methods. Two null hypotheses were tested: (1) there is no positive relationship between input resources and output, and (2) there is no negative relationship between input resources and output. It was concluded that there were some
positive effects related to input resources, but little support was found for negative effects. While making a case for a positive relationship between resources and achievement, Hedges et al. caution that the age of the data and the quality of some of the studies included in the original sample make the results of this study suspect.

The second perspective from which to study the relationship between economic resources and student achievement is to shift from school inputs to the home and examine the relationship between SES of students' families and school achievement. Hirth and Mitchell's (1995) work is important because it illustrates how production function studies can fail to uncover inequities caused by SES. This research replicated the methodology suggested by Fortune and O'Neil (1994) in which they proposed the use of t-tests to check for significance versus the use of traditional production function methodology used by Coleman (1966) and Hanushek (1986). This study compared the results of the Indiana Statewide Educational Testing Program (1993), grades 2 and 3, with district per pupil expenditures. The study used free and reduced price lunch percentages as a proxy for SES. No significant differences in achievement levels based on school expenditures were found. However, significant differences were found between high and low SES level elementary schools. Similar differences were also identified between low and high SES levels in middle and high schools. The researchers used their findings to conclude that alternative methods to production function studies should be used to uncover inequities between schools and to determine how expenditure levels affect achievement gains.

White (1982) conducted a meta-analysis of 101 studies that measured the relationship between SES and academic achievement to address the validity of the many measures of SES and to explain why correlations between SES and achievement are
rarely consistent. White listed the following factors to be among the many descriptors associated with the definition of SES: family income, parents' educational level, parents' occupation, occupation of principal breadwinner, source of income, quality of housing, status of dwelling area, size of family, educational aspirations, ethnicity, reading materials, and amount of travel. He concluded that, whenever possible, the term SES should not be used in research studies since there is no common definition of the term. Suggested alternatives for SES include family income, expenditure per pupil, or home atmosphere. The author also stated that an accurate description of how the SES variable was measured would assist in clarifying results.

White's second concern centers on the wide range of correlations (.10 to .80) between SES and academic achievement. He found that the inconsistency of the results could be explained by the fact that using the student as the unit of analysis resulted in positive, but weak correlations with academic achievement. Mean correlations for income (.315), education (.185), occupation (.201), and home atmosphere (.577) support this finding. However, when schools or other groups are used as the unit of analysis then SES measures were usually correlated strongly with academic achievement. Reported mean correlations that supported this conclusion were: income (.767), education (.686), occupation (.586), and school resources (.619). White (1982) offered this summary of the implications of this study upon future research: "When schools or other aggregated groups are the appropriate unit of analysis, traditional measures of SES are usually correlated strongly enough with academic achievement measures to be useful as a covariate, predictor, or stratifying variable" (p. 475).
Payne and Biddle (1999) compared the effects of child poverty in American schools with mathematics achievement using data from Second International Mathematics Study (SIMS) collected in 1982. The data analyzed in this study represented eighth grade students in the United States and 22 other countries. Socioeconomic data on students from the United States were obtained from the School District Data Book. Data for the School District Data Book were collected from the 1990 U.S. Census, other government agencies, and state and local school agencies. Multi-stage correlations were computed to determine classroom level and district level relationships between curriculum and achievement. Finally, a regression analysis was performed using variables and data for all school districts. The results of this study indicate that both child poverty and poor school funding affect student achievement in the United States. Comparisons with other countries show that highly funded school districts with low child poverty would rank second only to Japan in mathematics achievement, while poorly funded, high poverty school districts would rank below all industrialized countries participating in the SIMS study.

While White (1982) and Payne and Biddle's (1999) research was conducted on a nationwide sample, Ferguson (1991) conducted an analysis of previously gathered data on over 838 school districts from one state, Texas, representing 2.4 million students and 150,000 teachers. Data for this study were collected from current Texas state and local sources. Examples include the Texas Examination of Current Administrators and Teachers (TECAT), Texas Educational Assessment of Minimum Skills (TEAMS), 1980 U.S. Census, and other data on SES, school spending, and demographic information. Regression analysis determined that the most significant predictors of increased test
scores in a district were: (1) the number of teachers in a district with high literacy skills, (2) few large classes, and (3) few teachers with less than five years of teaching experience. The results of this study indicated that there was a connection to some traditional SES indicators since teachers selected employers based on the following characteristics: the educational levels of adults in the community, racial make-up of the community, and teacher salaries relative to surrounding school districts. Ferguson’s research is also important for the connection it identifies between the SES of a community and teachers’ employment decisions which, in turn, are directly linked to student achievement.

The manner in which SES affects teachers’ attitudes about mandated testing programs was studied by Herman and Golan (1991). In this research of mandated testing in nine states, SES was represented by the percentage of students qualifying for Chapter I services in each class. They found that in schools with high percentages of Chapter I students, there was more likely to be greater emphasis on testing and less attention to non-tested curriculum. Teachers in schools with high percentages of Chapter I students were also more likely to question the fairness of the tests and felt that students’ performance on tests was due to factors not under their control. They also reported that there was a discrepancy between what they believed should be taught and the content that was emphasized by the test.

The relationship between participation in federal free and reduced-price lunch programs, as an indicator of SES, and the academic achievement of high school students was studied by Caldas and Bankston (1997). The researchers obtained, from the Louisiana Department of Education, a data set from the 1990 grade 10 criterion
referenced tests in mathematics and English language arts. Races other than African American or White, as well as 3,068 special education students, were excluded from the study. A four step ordinary least squares regression analysis was performed in order to show the effect of individual variables, a school-level poverty status variable, a school-level family social status variable, and a school-level racial composition variable on school achievement. Individual participation in federal free and reduced-price lunch programs was found to have a small, but significant negative effect on academic achievement. Individual family status was found to have a significant positive effect on academic achievement. Related to this result was the finding that the level of family social status of some students significantly influenced the achievement levels of other students enrolled in the school. This research indicated that family characteristics such as occupational and educational status may be a better indicator of student achievement than qualifying for a free and reduced-price lunch program. However, while family status indicators may, in fact, be better indicators of student achievement than participation in free and reduced-price lunch program, such data are difficult to obtain. It could also be argued that a stronger relationship between participation in free or reduced price lunch programs could have been found if, as suggested by White (1982), the school was the unit of analysis instead of the individual student.

In the state of New Hampshire, Hall (1999) has conducted a study to determine the effect of class size, family income, adult education, and gender on the achievement of third grade students taking the NHEIAP test. Data on class size and student scores (15,799 students in language arts, 15,944 in mathematics) were obtained from the New Hampshire Department of Education on all 1997 third grade assessment tests. Additional
information regarding family income, percentage of the population with income levels below 185% of the poverty level, and adult education level was collected from the 1990 census. The percentage of students receiving free and reduced price school lunches in 1996 was collected from state officials. No statistical analysis was performed as part of this research; comparisons were made through the inspection of percentages and mean scores between variables. Results indicated that there was little distinction between the levels of achievement among students in different class sizes. However, other factors, such as family income, adult education levels, and schools percentage of students participating in federal free and reduced price lunch program were positively, and strongly, related to third grade achievement levels as measured by the NHEIAP test. Hall offered several conclusions based on the results of this study: (1) lowering class size will not necessarily result in higher NHEIAP test scores; (2) NHEIAP test scores are related to factors over which school have little control; therefore, it is not advisable to use these test scores as measures of school accomplishments; (3) schools within the state should be grouped by demographic factors related to achievement levels for the purpose of comparison; and (4) new formulas for the distribution of state aid should include demographic differences, since these factors are linked to levels of achievement.

NHEIAP Grade Three Test Details

The NHEIAP third grade test is a criterion-referenced instrument that is, according to state law, designed to assess the level of proficiency of students in attaining the competencies described in the New Hampshire curriculum frameworks. Tested content is cumulative across grade levels. For example, the grade three test is designed to
measure the skills acquired during the first three years of schooling and not just the content addressed in grade three. The test is divided into two sections identified as common items and matrix sampled items. Common item questions are administered to all students and later released to schools for instructional use. Matrix sampled items are divided among groups of students so that one administration of the test can cover a wide portion of the curriculum frameworks. In 1998, matrix sampled items were organized into eight separate groups, or forms. Each student completed a test consisting of common items in mathematics and English language arts and one of the eight matrix sampled forms in each content area. The details of each student section of the 1998 test are presented in Table 1. Matrix sampling significantly increases the number of questions administered to a cohort of students at a school and increases the number of content areas assessed. Table 1 also shows this breakdown by the total number of questions and their associated types at the school level.

All multiple-choice answers are scanned directly from each third grade student test and machine scored. Open response items are evaluated by trained scorers using zero to four point rubrics that are specifically designed for each item. All writing samples are scored by two raters using a one to six point rubric. If more than one point separates two raters' scores on any writing sample then a table leader is consulted to resolve the difference in scores. Scorers also identify two annotations that reflect positive or negative characteristics of the writing sample to be included with the school report.

Student level reports are sent to schools with information that is to be passed along to the parents of each student. Parents received raw scores on multiple-choice
questions, open response items, and writing samples. Raw scores are translated into either advanced, proficient, basic, or novice proficiency levels in English language arts.

Table 1

Student Level and School Level Item Summary

<table>
<thead>
<tr>
<th>Question type</th>
<th>Mathematics</th>
<th>English language arts</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple-choice</td>
<td>40</td>
<td>43</td>
<td>83</td>
</tr>
<tr>
<td>Open response</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Writing prompt</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question type</th>
<th>Mathematics</th>
<th>English language arts</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple-choice</td>
<td>96</td>
<td>99</td>
<td>195</td>
</tr>
<tr>
<td>Open response</td>
<td>19</td>
<td>19</td>
<td>38</td>
</tr>
<tr>
<td>Writing prompt</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

and in mathematics. Scaled scores within each proficiency level (200 to 300) provide a more detailed picture of each student's performance. Percentages of students within each proficiency level at the school, school district, and state levels are also reported. School and district level reports provide proficiency level results as well as content diagnostic reports. These reports include information about subtopics related to the curriculum frameworks that are to be used to determine strengths and weaknesses in curriculum and instruction. A subgroup performance report provides information about the major
subgroups (i.e., gender, Title I, reading recovery, English language proficiency, educational disability, kindergarten) of the student population. Mean-scaled reports are also included with school district and state level data.

The reliability coefficients for the 1998 test were reported as .98 for English language arts and .97 for mathematics at the school level. Fewer than one percent of the approximately 16,600 student writing samples required arbitration in 1998. This is the basis for a reported high interrater reliability. Content validity was reported as high because the test material is based on the New Hampshire curriculum frameworks. It was reported that construct validity was high since the results are consistent with educators' expectations. Instructional validity was also reported as high since open response items play an important role in the test (New Hampshire Department of Education, 1998b).

Both the writing prompt and the open response questions found in the NHEIAP grade three test can be described as performance based assessment items. Resnick and Resnick (1992) defined performance assessments as devices that use "direct judgments and evaluations of performances rather than indirect indicators of competence" (p. 61). Furthermore, they stated that performance assessments can be designed to positively shape a "thinking curriculum," where thinking and problem-solving are nurtured in all students, rather than constricting instruction and curriculum through the use of mandated standardized tests. In the school setting, performance assessments may be as involved as individual or group performances in front of classroom or community audiences or as simple as responses to an open-ended essay question. Wiggins and McTighe (1998) are not so liberal in their definition of performance assessments. Open-ended questions and problems that ask students to think critically are termed academic prompts. In their
opinion, performance assessments or tasks differ from academic prompts in that they are featured in a setting that replicates the constraints and opportunities that would be found in the "real world." Students who participate in performance tasks or assessments are provided with the task, standards, and criteria in advance in order that they may guide their work. Using this definition, the open response items and writing prompt found in the NHEIAP grade three test would be more accurately defined as academic prompts.

States base their support for performance based assessment items upon a philosophy similar to that summarized by Resnick and Resnick (1992): "Tied to curriculum and designed to be taught to, performance based assessments can become essential tools in educational reform" (p. 72). Madaus (1994), however, stated that performance based assessments are as likely to create unintended negative consequences for schools as any other type of high stakes measurement driven instruction. Klien, Jovanovic, Stecher, McCaffrey, Shavelson, Haertel, Solano-Flores, and Comfort (1997) found that both gender and racial differences were as likely to occur when using performance based assessments as when using traditional multiple-choice tests. Koretz, Stecher, Klein, McCaffrey, and Deibert (1994b) studied the influence of portfolio assessment on instruction in Vermont and found this type of assessment had its flaws. While principals and teachers reported that portfolio assessment was a powerful tool for changing instruction, the unreliability of scoring led to inconsistent interpretation of instructional goals and practices. Other problems included increased demands on teachers' time and school professional development resources.

48

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Summary

The use of tests in schools is a long accepted practice and part of educational culture. Tests have been used to determine intelligence, sort students, diagnose disabilities, indicate levels of achievement, and serve as predictors of future success. Until recently, however, most local school districts remained in control of the types of tests used and the method by which test results would be used and reported. Local school districts in 48 states no longer enjoy this freedom, as the call for accountability and standards has resulted in the design and implementation of state mandated testing programs (Edwards, 1999). The result of this policy change has been to increase public pressure on schools, teachers, and students to increase test scores.

The literature on testing is rich with information on how mandated tests have affected instruction and curriculum. But Madaus (1994) cautions that policymakers should understand that testing is a technology and that we do not yet know the full impact of this technology upon equity, individuals, schools, or the workplace. This is certainly the case in New Hampshire, as little research has been conducted on the effects of the NHEIAP tests on schools, teachers, or students. This study is intended to offer insight into the manner in which these tests are affecting curriculum, instruction, and teachers. This information will be useful to policymakers, educators, and local communities as they continue to work with the results of these assessments.
CHAPTER III

METHODOLOGY

The purpose of this study was to examine the perceptions of third grade teachers regarding the effects of New Hampshire's state-mandated third grade test. Analysis was also conducted to determine if there were differences in teachers' perceptions in schools with differing socioeconomic student populations. More specifically, this study sought to answer the following research questions:

1. What are the perceptions of third grade teachers regarding the NHEIAP grade three test's influence on curriculum and instruction?

2. What is the level of support among grade three teachers for the NHEIAP grade three tests?

3. From which group, or groups, do grade three teachers feel pressure to improve their students' scores on the NHEIAP grade three test?

4. Is there a relationship between changes made to the curriculum and in instruction and the socioeconomic status of the school community as defined by the percentage of students who participate in the federal free and reduced price school lunch program?

5. Is there a relationship between teachers' support for the NHEIAP grade three test and the socioeconomic status of the school community in which they teach as defined by the percentage of students who participate in the federal free and reduced price school lunch program?
6. Is there a relationship between the pressure to improve their students' scores on the NHEIAP grade three test and the socioeconomic status of the school community as defined by the percentage of students who participate in the federal free and reduced price school lunch program?

Instrumentation

A survey instrument consisting of a mailed questionnaire (Appendix A) was developed to address the objectives of this study. Many survey items were designed specifically to address the NHEIAP grade three test while the remaining items were adapted from two related studies; one instrument designed by Koretz et al. (1996b) and the second by Herman and Golan (1991). The Koretz et al. instrument was designed to gather information from teachers about the perceived effects of the Kentucky Instructional Results Information System (KIRIS). Specifically, data were collected on teachers' support for KIRIS, the effectiveness of KIRIS in determining student achievement, classroom assessment methods, test preparation practices, effects on instruction and learning, student incentives and rewards, and school climate. Herman and Golan designed their questionnaire to measure the impact on instruction, curriculum, and teacher planning. They also sought to determine how standardized testing influenced teacher attitudes and to establish which groups were exerting pressure on schools to improve test results.

The original draft of the survey for this study included 44 closed response items, 1 open response item and five demographic items. Based on recommendations from the researcher's dissertation committee, the instructional practices section of the survey was
changed to report separate information on English Language Arts and mathematics. This expanded the total number of general response items to 56. After this initial revision, the survey was piloted with seven third grade teachers, each of whom had previously administered the NHEIAP third grade test. The schools where these teachers were employed were eliminated from consideration for selection to the research sample. Concurrent with the pilot study, information regarding this study was made available to officials at the NHDOE. A content area specialist with the department submitted a list of suggested revisions to the survey as a result of this contact.

Pilot study participants reported that it took an average of 15 minutes to complete the survey. Based upon the information received from the pilot study participants and from the NHDOE content area specialist, the following revisions were made to the survey:

1. Minor changes to the survey directions were made to improve clarity;
2. Language and terms were changed in order to ensure consistency throughout the survey;
3. The categories - myself, New Hampshire Department of Education, and State Government Officials - were added to the attitude section;
4. Three questions (50, 51, 52) relating to tests' ability to describe achievement levels were added;
5. Two questions (54, 55) relating to funding and socioeconomic levels of the community were added;
6. Questions were reorganized to reflect a hierarchical order, where appropriate.
The final questionnaire consisted of 61 closed response items, 1 open response item, and 5 demographic response items. The closed response items were presented in six categories: (1) shifts in instructional time, (2) changes in school-wide instructional emphasis, (3) changes in instructional practices related to the English language arts section of the NHEIAP third grade test, (4) changes in instructional practices related to the Mathematics section of the NHEIAP third grade test, (5) sources of pressure, and (6) personal attitudes. In the first four sections, related to curriculum and instruction, respondents were asked to record their responses on a five-point Likert scale labeled as: (1) decreased greatly, (2) decreased somewhat, (3) stayed about the same, (4) increased somewhat, and (5) increased greatly. Respondents were asked to record their responses to items in the sources of pressure section on a five-point Likert scale labeled as: (1) no pressure, (2) no label, (3) some pressure, (4) no label, and (5) great pressure. Respondents were asked to record their responses to items in the last section, personal attitudes, on a five-point Likert scale labeled as: (1) disagree, (2) somewhat disagree, (3) neither agree nor disagree, (4) somewhat agree, and (5) strongly agree. The format for organizing responses was adapted, with permission, from the Koretz, Barron, Mitchell, and Stetcher (1996b) study that sought to determine teachers’ perceptions regarding the effects of the Kentucky’s mandated assessment program (KIRIS).

In section one, shifts in instructional time, items (1) reading, (2) writing, and (3) mathematics are congruent with the subject areas that are the focus of the NHEIAP grade three test. The remaining items in the section, (4) social studies, (5) science, (6) art, (7) music, and (8) physical education, are subject areas typically found in most elementary schools in New Hampshire. The design of this section attempted to identify how third
grade teachers have changed the amount of time dedicated to each subject area as a result of the NHEIAP grade three test.

In order to determine teachers’ perceptions regarding how the NHEIAP test influenced instructional emphasis in their schools, it was first necessary to identify the areas that the NHEIAP test targeted. Since administrators and teachers used the NHEIAP Educational Assessment Report (New Hampshire Department of Education, 1998a) as the basis for examining results and as a basis for planning improvement efforts, the proficiencies described in this document were used as the basis for items on the questionnaire. Under English language arts there were three items that were used to describe the changes in instructional emphasis: (1) reading and literature, (2) listening and viewing, and (3) writing. There were nine items used to describe changes in instructional emphasis in mathematics: (1) problem solving and reasoning, (2) addition, subtraction, and whole numbers, (3) fractions, (4) adding and subtracting decimals, (5) estimation, (6) understanding and using charts and graphs, (7) identifying, classifying, and comparing geometric objects, (8) measurement, and (9) patterns.

Herman and Golan’s (1991) research, in which they surveyed elementary school teachers from 11 school districts in nine states to determine the effects of standardized tests on teachers and learning, strongly influenced the content of much of this survey due to the similarities between the two studies. Many of the items in each of the remaining sections were adapted, with permission, from the Herman and Golan survey instrument. Changes in instructional practices in English language arts and mathematics were measured using the same set of 10 indicators, under separate headings. The indicators sought to build a picture of the manner in which teachers had changed common
instructional practices as a result of the NHEIAP grade three tests. The instructional practices identified in this section included: (1) use constructed response exercises; (2) emphasis on preparing students for the NHEIAP test; (3) use of multiple-choice, fill-in-the-blank, and matching exercises; (4) use of cooperative/small group learning; (5) implementation of innovative instructional strategies; (6) support for school-wide or grade-level planning; (7) use of project work; (8) use of activities based on drill; (9) use of higher-order thinking and problem solving activities; and, (10) opportunities for students to choose what they want to study.

The items used to identify sources of accountability pressure were also adapted from the Herman and Golan study. This section consisted of eight items in which respondents were asked to identify the amount of pressure that they attributed to each group or individual. Groups, or individuals, identified as pressure groups were: (1) myself, (2) other teachers, (3) principal, (4) district administrators, (5) New Hampshire Department of Education, (6) state government officials, (7) parents, (8) community, and, (9) newspaper/media.

Herman and Golan also sought to acquire data on teacher attitudes toward standardized tests using a set of statements about how the test has affected their school. These statements were adapted to fit the NHEIAP tests and the goals of this study. A set of twelve items were used to measure teachers' attitudes in the following areas: (1) the test as a measure of school quality; (2) the effect of funding and socioeconomic status on test results; and (3) general attitudes and amount of support for the NHEIAP test. The final sections of the questionnaire are divided into an open response section and a section requesting demographic information from respondents.
Population and Sample

The state of New Hampshire does not record information regarding the number of teachers assigned to a particular grade level in any of its schools. Therefore, in order to determine the number of third grade teachers in New Hampshire, an average estimated classroom size of 20 students was divided into the total population of third grade students reported as enrolled (16,641) by the NHDOE on the first day of testing in 1998 (1998a). The results of this calculation yielded an approximate population of 832 third grade teachers. In 1998 a total of 241 elementary schools participated in the NHEIAP assessment program. Since the goals of this study were to determine the perceptions of third grade teachers and to find out if there is a relationship between their perceptions and socioeconomic status, as measured by the percentage of students participating in the federal free and reduced price lunch program, a stratified random sample of schools was selected. This stratifying unit was suggested by Hall (1999) in a study of the relationship between New Hampshire third grade test results and local demographics. The sample was generated by first dividing the population of 241 elementary schools into three groups categorized by the percentage of students that participated in the federal free and reduced price lunch program. Each of these strata were designed to represent approximately a third of the state’s third grade student population. A random sample of schools was then taken from within each stratum until approximately 2000 students were represented in order to obtain a sample that would create a sample of 100 teachers per stratum. This process created three socioeconomic groups of schools with the designations of low free and reduced price lunch percentage schools (Low FRL%), medium free and reduced price lunch percentage schools (Med FRL%), and high free and reduced price lunch
percentage schools (High FRL%). Table 2 shows each of the three socioeconomic strata with the associated number of teachers, students, and schools represented by the sample.

Once the sample was selected, the names of the third grade teachers at each school were required in order to send and track surveys and responses. Principals were contacted by mail and asked to provide the researcher with the names of all third grade teachers at their school. The mailing to principals included a letter describing this study and asking for their assistance in acquiring the names of third grade teachers at their school (Appendix B). Also included in the mailing was a pre-addressed, stamped postcard with space designated for the names of teachers (Appendix C) and a letter of support (Appendix D). The letter of support for this study was endorsed by the directors of each of the following state organizations: the New Hampshire School Administrators Association, the New Hampshire Association of School Principals, the New Hampshire affiliate of the National Education Association, and the New Hampshire affiliate of the American Federation of Teachers.

Table 2

<table>
<thead>
<tr>
<th>Students receiving FRL</th>
<th>Sampled teachers</th>
<th>Sampled schools</th>
<th>NH schools reporting NHEIAP scores</th>
<th>Students represented in sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low FRL% 00.00-12.28%</td>
<td>105</td>
<td>28</td>
<td>66</td>
<td>2028</td>
</tr>
<tr>
<td>Med FRL% 12.61-24.14%</td>
<td>100</td>
<td>34</td>
<td>79</td>
<td>2100</td>
</tr>
<tr>
<td>High FRL% 24.90-55.76%</td>
<td>105</td>
<td>37</td>
<td>96</td>
<td>2142</td>
</tr>
<tr>
<td>Total</td>
<td>310</td>
<td>99</td>
<td>241</td>
<td>6270</td>
</tr>
</tbody>
</table>
Seventy-five of the 102 principals contacted returned the pre-addressed, postage paid response cards. The remaining schools were contacted by telephone and asked to provide teachers’ names. Three principals requested that their schools be excluded from this study. These schools were removed from consideration, resulting in a total sample size of 99 schools. Six schools, who participated in this study, did not disclose the names of third grade teachers, but did identify the number of teachers assigned to third grade classrooms.

**Data Collection Procedures**

Teachers selected for this study were first sent an introductory postcard (Appendix E) briefly describing the study and asking for their cooperation. Berdie, Anderson, and Niebuhr (1986) support the use of introductory letters, or “preletters” but reported that the research is mixed on whether its use increases response rates. They stated that even if there is no positive effect, the use of a “preletter” is a professional courtesy which should be extended to those who will be asked to commit time to a particular research project.

A survey packet was sent to all third grade teachers in the sample three days after the introductory postcard was mailed. The survey packet included: a cover letter (Appendix F), a letter of support (Appendix D), a copy of the survey (Appendix A), and a pre-addressed, stamped return envelope. Each survey was color-coded by percentage of students participating in the federal free and reduced price lunch program to facilitate data entry. A numerical code was placed on the front of each survey to track returns and
to assist with follow-up procedures. A reminder to complete and return the survey was mailed to all non-respondents two weeks after the survey packet was sent (Appendix G). This reminder extended the original deadline by two weeks. A second follow-up letter was sent to all non-respondents two weeks after the first reminder was sent (Appendix H). This mailing included a second, coded copy of the survey as well as a self-addressed, stamped return envelope. At the end of the final deadline period, a final attempt was made to contact the remaining non-respondents by calling each school and requesting that a reminder be placed in each teacher's mailbox. A summary of the number of returns after each mailing is presented in Table 3. Of the 310 teachers surveyed, 257 responded (83% response rate), of which 251 were usable (81% usable rate). A total of 33 respondents (13%) reported they were first year teachers and had not administered the NHEIAP third grade test. Since the goals of this study relied upon teachers' experience with the administration and implementation of the NHEIAP grade three test, all data for first year teachers were removed from analysis.

Table 3

<table>
<thead>
<tr>
<th>Survey Return Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Initial mailing</strong></td>
</tr>
<tr>
<td><strong>First reminder</strong></td>
</tr>
<tr>
<td><strong>Second copy of survey</strong></td>
</tr>
<tr>
<td><strong>Phone call</strong></td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>LOW FRL% (105)</td>
</tr>
<tr>
<td>41</td>
</tr>
<tr>
<td>19</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>83</td>
</tr>
<tr>
<td>MEDIUM FRL% (100)</td>
</tr>
<tr>
<td>39</td>
</tr>
<tr>
<td>22</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>83</td>
</tr>
<tr>
<td>HIGH FRL% (105)</td>
</tr>
<tr>
<td>43</td>
</tr>
<tr>
<td>26</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>91</td>
</tr>
<tr>
<td>TOTAL (310)</td>
</tr>
<tr>
<td>123</td>
</tr>
<tr>
<td>67</td>
</tr>
<tr>
<td>38</td>
</tr>
<tr>
<td>29</td>
</tr>
<tr>
<td>257</td>
</tr>
</tbody>
</table>
Data Analysis

Data obtained from the closed response items were organized into two categories: (1) curriculum and (2) pressure and attitudes. Each of these categories was further divided into subscales. Within the curriculum category six subscales were created while the pressure and attitudes category consisted of three subscales. Table 4 identifies the survey items within each subscale for both the curriculum and instruction, and pressure and attitudes categories. Questions 54 and 55 were not included in any subscale because they measured teachers' perceptions of the effects of socioeconomic status on test results and the effect of funding on test results.

Table 4

Initial Subscale Outline

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Title</th>
<th>Item numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curriculum and instruction category</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS1</td>
<td>Shift in tested curriculum instructional time</td>
<td>1 - 3</td>
</tr>
<tr>
<td>SS2</td>
<td>Shift in non-tested curriculum instructional time</td>
<td>4 - 8</td>
</tr>
<tr>
<td>SS3</td>
<td>Changes in school emphasis in English language arts</td>
<td>9 - 11</td>
</tr>
<tr>
<td>SS4</td>
<td>Changes in school emphasis in mathematics</td>
<td>12 – 20</td>
</tr>
<tr>
<td>SS5</td>
<td>Changes in practice in English language arts</td>
<td>21 - 30</td>
</tr>
<tr>
<td>SS6</td>
<td>Changes in practice in mathematics</td>
<td>31 - 40</td>
</tr>
<tr>
<td>Pressure and attitude category</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS7</td>
<td>Pressure</td>
<td>41 -49</td>
</tr>
<tr>
<td>SS8</td>
<td>NHEIAP test and school quality</td>
<td>50 -53</td>
</tr>
<tr>
<td>SS9</td>
<td>Attitude toward NHEIAP test</td>
<td>56 -61</td>
</tr>
</tbody>
</table>

Confirmatory factor analysis was used to verify the constructs described by each subscale. A principal component analysis with varimax rotation was used as a method of...
verifying the initial subscale outline before calculating the reliability of each subscale.

The results of the factor analysis for the curriculum category are presented in Table 5, while the results for the pressure and attitude category are presented in Table 6 (see appendices I and J for complete results). The loadings for subscales one, three, and four confirmed the conceptual outline as presented in Table 4 with high factor loadings (.42 to .83). The loadings for subscale two suggested that it should be further divided into non-tested academic instructional time (social studies and science) and other instructional time (art, music, and physical education). The alpha levels for each of these new subscales (.879 and .799) were significantly higher than the original subscale (.622). The factor analysis of subscales five and six produced mixed results. While the loadings for items 24 through 27 and 34 through 37 were strong (.51 to .82) the other items in each subscale did not load strongly on factor 2. A comparison between the alpha levels of the original subscales (.722 and .869) and the proposed subscales (.822 and .799) were not significant enough to warrant a change to the original subscales. A summary of the reliability coefficients for each subscale as well as each of the suggested subscales, is shown in Table 7.

<table>
<thead>
<tr>
<th>Table 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Curriculum and Instruction Category Factor Analysis</strong></td>
</tr>
<tr>
<td><strong>Subscale items</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>SS1 Shift in tested curriculum instructional time</td>
</tr>
<tr>
<td>Q1 Reading</td>
</tr>
<tr>
<td>Q2 Writing</td>
</tr>
<tr>
<td>Q3 Mathematics</td>
</tr>
<tr>
<td>SS2 Shift in non-tested curriculum instructional time</td>
</tr>
<tr>
<td>Q4 Social Studies</td>
</tr>
<tr>
<td>Q5 Science</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 5 Continued

<table>
<thead>
<tr>
<th>Subscale items</th>
<th>Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q6 Art</td>
<td>.92</td>
</tr>
<tr>
<td>Q7 Music</td>
<td>.95</td>
</tr>
<tr>
<td>Q8 Physical Education</td>
<td>.89</td>
</tr>
<tr>
<td>SS3 Changes in school emphasis in English language arts</td>
<td></td>
</tr>
<tr>
<td>Q9 Reading and literature</td>
<td>.69</td>
</tr>
<tr>
<td>Q10 Listening/viewing</td>
<td>.43</td>
</tr>
<tr>
<td>Q11 Writing</td>
<td>.76</td>
</tr>
<tr>
<td>SS4 Changes in school emphasis in mathematics</td>
<td></td>
</tr>
<tr>
<td>Q12 Problem solving</td>
<td>.45</td>
</tr>
<tr>
<td>Q13 Addition and subtraction</td>
<td>.49</td>
</tr>
<tr>
<td>Q14 Fractions</td>
<td>.77</td>
</tr>
<tr>
<td>Q15 Add/subtract decimals</td>
<td>.71</td>
</tr>
<tr>
<td>Q16 Estimation</td>
<td>.67</td>
</tr>
<tr>
<td>Q17 Using charts and graphs</td>
<td>.66</td>
</tr>
<tr>
<td>SS5 Changes in practice in English language arts</td>
<td></td>
</tr>
<tr>
<td>Q18 Identify/compare shapes</td>
<td>.72</td>
</tr>
<tr>
<td>Q19 Measurement</td>
<td>.83</td>
</tr>
<tr>
<td>Q20 Patterns</td>
<td>.74</td>
</tr>
<tr>
<td>Q21 Constructed response</td>
<td>.49</td>
</tr>
<tr>
<td>Q22 Prepare students for test</td>
<td>.86</td>
</tr>
<tr>
<td>Q23 Multiple choice</td>
<td>.78</td>
</tr>
<tr>
<td>Q24 Small group learning</td>
<td>.65</td>
</tr>
<tr>
<td>Q25 Innovative strategies</td>
<td>.78</td>
</tr>
<tr>
<td>Q26 School or grade-level planning</td>
<td>.82</td>
</tr>
<tr>
<td>Q27 Project work</td>
<td>.74</td>
</tr>
<tr>
<td>Q28 Drill to build/review skills</td>
<td>.86</td>
</tr>
<tr>
<td>Higher-order thinking</td>
<td>.86</td>
</tr>
<tr>
<td>Q29 activities</td>
<td>.50</td>
</tr>
<tr>
<td>Q30 Student choice of topics</td>
<td>.73</td>
</tr>
<tr>
<td>SS6 Changes in practice in mathematics</td>
<td></td>
</tr>
<tr>
<td>Q31 Constructed response</td>
<td>.68</td>
</tr>
<tr>
<td>Q32 Preparing for test</td>
<td>.69</td>
</tr>
<tr>
<td>Q33 Multiple choice</td>
<td>.86</td>
</tr>
<tr>
<td>Q34 Small group learning</td>
<td>.57</td>
</tr>
<tr>
<td>Q35 Innovative strategies</td>
<td>.70</td>
</tr>
</tbody>
</table>
Table 5 Continued

<table>
<thead>
<tr>
<th>Subscale items</th>
<th>Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q36 School/grade-level planning</td>
<td>.74</td>
</tr>
<tr>
<td>Q37 Project work</td>
<td>.58</td>
</tr>
<tr>
<td>Q38 Drill to build/review skills</td>
<td>.83</td>
</tr>
<tr>
<td>Q39 Higher-order thinking activities</td>
<td>.74</td>
</tr>
<tr>
<td>Q40 Student choice of topics</td>
<td>.84</td>
</tr>
<tr>
<td>Percentage of variance</td>
<td></td>
</tr>
<tr>
<td>12.46 11.00 6.76 6.35 5.95 5.57 5.22 5.03 4.89 4.82 4.40</td>
<td></td>
</tr>
</tbody>
</table>

The loadings for subscales seven, eight, and nine confirmed the conceptual outline as presented in Table 4 with high factor loadings (.47 to .86). The loadings for subscale seven did, however, suggest that it should be further divided into internal and external sources of pressure. The alpha levels for each of these suggested subscales (.771 and .839) were somewhat lower than the original subscale (.869). However, the original subscale was divided as suggested, since the division into internal and external sources of pressure were useful designations that would assist with the analysis of data. The alpha levels for subscales eight (.842) and nine (.800) also supported the original subscale outline. A summary of the reliability coefficients for each subscale, as well as each of the suggested subscales, is shown in Table 7. The final subscale outline is shown in Table 8.

Table 6

Pressure and Attitude Category Factor Analysis

<table>
<thead>
<tr>
<th>Subscale items</th>
<th>Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS7 Pressure</td>
<td></td>
</tr>
<tr>
<td>Q41 Myself</td>
<td>.65</td>
</tr>
<tr>
<td>Q42 Other teachers</td>
<td>.75</td>
</tr>
<tr>
<td>Q43 Principal</td>
<td>.68</td>
</tr>
<tr>
<td>Q44 District Administrators</td>
<td>.59</td>
</tr>
</tbody>
</table>
Table 6 Continued

<table>
<thead>
<tr>
<th>Subscale items</th>
<th>Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Q45 New Hampshire DOE</td>
<td>.82</td>
</tr>
<tr>
<td>Q46 State Gov't. Officials</td>
<td>.86</td>
</tr>
<tr>
<td>Q47 Parents</td>
<td>.61</td>
</tr>
<tr>
<td>Q48 Community</td>
<td>.74</td>
</tr>
<tr>
<td>Q49 Media</td>
<td>.74</td>
</tr>
<tr>
<td>SS8 Test reflects school quality</td>
<td></td>
</tr>
<tr>
<td>Q50 Results reflect individual student achievement</td>
<td>.85</td>
</tr>
<tr>
<td>Q51 Results reflect overall student achievement</td>
<td>.78</td>
</tr>
<tr>
<td>Q52 Results reflect differences among schools</td>
<td>.77</td>
</tr>
<tr>
<td>Q53 Results reflect the quality of instruction</td>
<td>.78</td>
</tr>
<tr>
<td>SS9 Attitude toward NHEIAP test</td>
<td></td>
</tr>
<tr>
<td>Q56 Results influenced by changes in instructional practices</td>
<td>.59</td>
</tr>
<tr>
<td>Q57 Results influenced by alignment with state frameworks</td>
<td>.69</td>
</tr>
<tr>
<td>Q58 Provides feedback on how well I am teaching</td>
<td>.45</td>
</tr>
<tr>
<td>Q59 Produced positive instructional change</td>
<td>.41</td>
</tr>
<tr>
<td>Q60 Produced positive curricular change</td>
<td>.74</td>
</tr>
<tr>
<td>Q61 I support the NHEIAP test</td>
<td>.52</td>
</tr>
<tr>
<td>Percentage of variance</td>
<td>18.78</td>
</tr>
</tbody>
</table>

Table 7

Comparison of Subscale Reliability

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Items</th>
<th>Reliability of original subscale</th>
<th>Items</th>
<th>Reliability of subscales suggested by factor analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curriculum and instruction category</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS1 1 - 3</td>
<td>.738</td>
<td></td>
<td>1 - 3</td>
<td>.738</td>
</tr>
<tr>
<td>SS2 4 - 8</td>
<td>.622</td>
<td></td>
<td>4 - 5</td>
<td>.879</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6 - 8</td>
<td>.922</td>
</tr>
<tr>
<td>SS3 9 - 11</td>
<td>.682</td>
<td></td>
<td>9 - 11</td>
<td>.682</td>
</tr>
<tr>
<td>SS4 12 - 20</td>
<td>.889</td>
<td></td>
<td>12 - 20</td>
<td>.889</td>
</tr>
<tr>
<td>SS5 21 - 30</td>
<td>.727</td>
<td></td>
<td>24 - 27</td>
<td>.822</td>
</tr>
<tr>
<td>SS6 31 - 40</td>
<td>.722</td>
<td></td>
<td>34 - 37</td>
<td>.799</td>
</tr>
</tbody>
</table>
Table 7 Continued

Comparison of Subscale Reliability

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Items</th>
<th>Reliability of original subscale</th>
<th>Items</th>
<th>Reliability of subscales suggested by factor analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pressure and attitude category</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS7</td>
<td>41-49</td>
<td>.869</td>
<td>41-44</td>
<td>.771</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>45-49</td>
<td>.839</td>
</tr>
<tr>
<td>SS8</td>
<td>50-53</td>
<td>.842</td>
<td>50-53</td>
<td>.842</td>
</tr>
<tr>
<td>SS9</td>
<td>56-61</td>
<td>.800</td>
<td>56-61</td>
<td>.800</td>
</tr>
</tbody>
</table>

Note. Cronbach's alpha was used to determine reliability.

Statistical analysis of the data collected from the surveys was performed using the
Statistical Package for the Social Sciences (SPSS, 1999) computer program.

Frequencies, means and standard deviations were used to analyze the data generated by
the Likert scale items and demographic items. Non-response items were coded as
missing values and were not included in any statistical calculations. Subscale means
were calculated for each individual case within each of the three socioeconomic groups to
prepare for parametric statistical analysis.

In order to determine if differences exist among the subscale means of each of the
three free and reduced lunch strata, a one-way ANOVA was performed. Tukey's HSD
multiple comparison procedure was used to confirm significant differences of .05 or
greater. An analysis of individual item means which scored significantly high and low
was conducted. Open response items were coded and summarized.
Final Subscale Outline

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Title</th>
<th>Item numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Curriculum and instruction category</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS1</td>
<td>Shift in tested curriculum instructional time</td>
<td>1 - 3</td>
</tr>
<tr>
<td>SS2</td>
<td>Shift in social studies and science instructional time</td>
<td>4 - 5</td>
</tr>
<tr>
<td>SS3</td>
<td>Shift in music, art, and physical education instructional time</td>
<td>6 - 8</td>
</tr>
<tr>
<td>SS4</td>
<td>Changes in school emphasis in English language arts</td>
<td>9 - 11</td>
</tr>
<tr>
<td>SS5</td>
<td>Changes in school emphasis in mathematics</td>
<td>12 - 20</td>
</tr>
<tr>
<td>SS6</td>
<td>Changes in practice in English language arts</td>
<td>21 - 30</td>
</tr>
<tr>
<td>SS7</td>
<td>Changes in practice in mathematics</td>
<td>31 - 40</td>
</tr>
<tr>
<td><strong>Pressure and attitude category</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS8</td>
<td>Internal pressure</td>
<td>41 - 44</td>
</tr>
<tr>
<td>SS9</td>
<td>External pressure</td>
<td>45 - 49</td>
</tr>
<tr>
<td>SS10</td>
<td>NHEIAP test and school quality</td>
<td>50 - 53</td>
</tr>
<tr>
<td>SS11</td>
<td>Attitude toward NHEIAP test</td>
<td>56 - 61</td>
</tr>
</tbody>
</table>
CHAPTER IV

ANALYSIS OF DATA

This study was designed to gather and interpret data on the perceptions of grade three teachers in three areas: (1) changes in curriculum and instruction due to the implementation of the NHEIAP grade three test; (2) the sources of pressure for improvement of test results; and, (3) the level of support for the NHEIAP grade three test. In addition, this study sought to determine whether there was a relationship among each of the three objectives and the socioeconomic status of the local school community. This data analysis will be organized under the six headings of: (1) demographic characteristics of the survey respondents, (2) curriculum and instruction, (3) sources of pressure, (4) attitudes, (5) socioeconomic and pupil funding items, and (6) summary of open response items.

Demographic Characteristics of Survey Respondents

The sample for this study included 310 third grade teachers selected from three socioeconomic strata as defined by the percentage of students qualifying for the federal free and reduced price lunch program at their school. A total of 257 teachers (83%) responded, of which 251 (81%) were usable for the purposes of this study.

The demographic data collected from respondents is presented in Table 9. Included are data on overall teaching experience, third grade teaching experience, experience administering the NHEIAP grade three test, and the educational level of
Table 9

Demographic Characteristics of Respondents

<table>
<thead>
<tr>
<th></th>
<th>Low FRL% (N=65)</th>
<th>Med. FRL% (N=76)</th>
<th>High FRL% (N=77)</th>
<th>Total (N=218)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Years administering test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2</td>
<td>20%</td>
<td>28%</td>
<td>22%</td>
<td>23%</td>
</tr>
<tr>
<td>3-4</td>
<td>22%</td>
<td>18%</td>
<td>29%</td>
<td>23%</td>
</tr>
<tr>
<td>5-6</td>
<td>58%</td>
<td>54%</td>
<td>49%</td>
<td>54%</td>
</tr>
<tr>
<td><strong>Years teaching grade three</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-6</td>
<td>38%</td>
<td>51%</td>
<td>49%</td>
<td>47%</td>
</tr>
<tr>
<td>7-12</td>
<td>40%</td>
<td>16%</td>
<td>24%</td>
<td>26%</td>
</tr>
<tr>
<td>13-18</td>
<td>8%</td>
<td>21%</td>
<td>14%</td>
<td>14%</td>
</tr>
<tr>
<td>19-24</td>
<td>5%</td>
<td>7%</td>
<td>4%</td>
<td>5%</td>
</tr>
<tr>
<td>25-30</td>
<td>9%</td>
<td>4%</td>
<td>4%</td>
<td>6%</td>
</tr>
<tr>
<td>31+</td>
<td>0%</td>
<td>1%</td>
<td>5%</td>
<td>2%</td>
</tr>
<tr>
<td><strong>Total teaching experience</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-6</td>
<td>14%</td>
<td>16%</td>
<td>17%</td>
<td>16%</td>
</tr>
<tr>
<td>7-12</td>
<td>20%</td>
<td>16%</td>
<td>15%</td>
<td>17%</td>
</tr>
<tr>
<td>13-18</td>
<td>18%</td>
<td>30%</td>
<td>12%</td>
<td>20%</td>
</tr>
<tr>
<td>19-24</td>
<td>25%</td>
<td>14%</td>
<td>22%</td>
<td>20%</td>
</tr>
<tr>
<td>25-30</td>
<td>18%</td>
<td>20%</td>
<td>26%</td>
<td>21%</td>
</tr>
<tr>
<td>31+</td>
<td>5%</td>
<td>4%</td>
<td>8%</td>
<td>6%</td>
</tr>
<tr>
<td><strong>Highest degree earned</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelors</td>
<td>46%</td>
<td>51%</td>
<td>62%</td>
<td>54%</td>
</tr>
<tr>
<td>Masters</td>
<td>54%</td>
<td>49%</td>
<td>38%</td>
<td>46%</td>
</tr>
</tbody>
</table>

Note. All data for first year teachers and teachers who have not administered the NHEIAP grade three test have been removed from analysis.
respondents. The data are organized by the socioeconomic level of schools as indicated by the percentage of students qualifying for free and reduced price lunches. The total column summarizes the results of all respondents.

Teachers’ experience administering the NHEIAP grade three test ranged from one to six years. Low free and reduced price lunch percentage (FRL%) schools had the most teachers with five or more years experience in administering the test (58%) while medium FRL% schools were second (54%), and high FRL% ranked third (49%).

Teachers’ experience teaching grade three ranged from one year to over 31 years. The largest group of teachers responding to this study reported that they had 12 or fewer years experience teaching grade three. Low FRL% schools had the largest percentage of teachers in the 12 or fewer years of experience category (78%). High FRL% schools ranked second (73%) while medium FRL% schools were third (67%) in the 12 or fewer years of experience category.

Years of total teaching experience ranged from one year to over 31 years. Overall teaching experience was more widely distributed among all levels of experience than in the demographic category for third grade teaching experience only. The highest percentage (25%) of teachers in low FRL% schools reported that they had between 19 and 24 years of total teaching experience. The largest grouping of teachers within medium FRL% schools was the 13 to 18 years of experience level (30%). In high FRL% schools 25 to 30 years of teaching experience was where the largest subset of teachers occurred (26%).

Participants in this research were asked to list their highest degree awarded. Overall, the participants in this group had either a bachelors degree or both a bachelors
and a masters degree. Low FRL% school had the greatest percentage of teachers with at masters degree (54%), while medium FRL% schools ranked second (49%), and high FRL% third (38%) in this category.

Curriculum and Instruction

General Description of Data

Data for curriculum and instruction were collected from 40 closed response items organized into five subscales related to curriculum (subscales 1 to 5) and two subscales related to instruction (subscales 6 and 7). For each of the 40 items, teachers were asked to rate the level of change caused by the NHEIAP third grade test on a five-point Likert scale labeled in the following manner: (1) decreased greatly, (2) decreased somewhat, (3) stayed about the same, (4) increased somewhat, and (5) increased greatly.

The means and standard deviations for all items in the curriculum and instruction subscales are shown in Table 10. The data are organized by the socioeconomic level of local school communities as measured by the percentage of students participating in the free and reduced price school lunch program (FRL%). A heading labeled as “Total” includes summary data for all socioeconomic levels. The means for teachers working in low FRL% school ranged from 2.90 to 4.33, indicating a variation in levels of change from decreased somewhat/stayed about the same to increased somewhat/increased greatly. The means for teachers working in medium FRL% school ranged from 2.67 to 4.41 indicating a variation in levels of change from decreased somewhat/stayed about the same to increased somewhat/increased greatly. The means for teachers working in high FRL% school ranged from 2.73 to 4.45 indicating a variation in levels of change from
<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Low FRL%</th>
<th>Medium FRL%</th>
<th>High FRL%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subscale 1: Shift in reading, writing, math instructional time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>Reading</td>
<td>65</td>
<td>3.65 .80</td>
<td>76</td>
<td>3.46 .64</td>
</tr>
<tr>
<td>Q2</td>
<td>Writing</td>
<td>65</td>
<td>4.11 .79</td>
<td>75</td>
<td>4.17 .69</td>
</tr>
<tr>
<td>Q3</td>
<td>Mathematics</td>
<td>64</td>
<td>3.81 .79</td>
<td>76</td>
<td>3.95 .63</td>
</tr>
<tr>
<td>Subscale 2: Shift in non-tested academic instructional time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td>Social Studies</td>
<td>65</td>
<td>2.69 .81</td>
<td>76</td>
<td>2.70 .75</td>
</tr>
<tr>
<td>Q5</td>
<td>Science</td>
<td>65</td>
<td>2.75 .61</td>
<td>76</td>
<td>2.67 .68</td>
</tr>
<tr>
<td>Subscale 3: Shift in other non-tested instructional time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q6</td>
<td>Art</td>
<td>60</td>
<td>2.92 .33</td>
<td>73</td>
<td>2.92 .36</td>
</tr>
<tr>
<td>Q7</td>
<td>Music</td>
<td>60</td>
<td>2.90 .44</td>
<td>71</td>
<td>2.93 .31</td>
</tr>
<tr>
<td>Q8</td>
<td>Physical Education</td>
<td>60</td>
<td>2.95 .39</td>
<td>73</td>
<td>2.96 .20</td>
</tr>
<tr>
<td>Subscale 4: Change in school emphasis in English language arts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q9</td>
<td>Reading and literature</td>
<td>64</td>
<td>3.92 .80</td>
<td>73</td>
<td>3.97 .69</td>
</tr>
<tr>
<td>Q10</td>
<td>Listening/viewing</td>
<td>62</td>
<td>3.79 .68</td>
<td>72</td>
<td>3.79 .63</td>
</tr>
<tr>
<td>Q11</td>
<td>Writing</td>
<td>64</td>
<td>4.33 .74</td>
<td>75</td>
<td>4.37 .63</td>
</tr>
<tr>
<td>Item</td>
<td>Description</td>
<td>Low FRL%</td>
<td>Medium FRL%</td>
<td>High FRL%</td>
<td>Total</td>
</tr>
<tr>
<td>----------</td>
<td>--------------------------------------------------</td>
<td>----------</td>
<td>-------------</td>
<td>-----------</td>
<td>-------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td><strong>Subscale 5: Change in school emphasis in mathematics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q12</td>
<td>Problem solving and reasoning</td>
<td>64</td>
<td>4.25</td>
<td>.67</td>
<td>73</td>
</tr>
<tr>
<td>Q13</td>
<td>Addition, subtraction, and whole numbers</td>
<td>64</td>
<td>3.28</td>
<td>.58</td>
<td>73</td>
</tr>
<tr>
<td>Q14</td>
<td>Fractions</td>
<td>64</td>
<td>3.64</td>
<td>.68</td>
<td>71</td>
</tr>
<tr>
<td>Q15</td>
<td>Adding and subtracting decimals</td>
<td>64</td>
<td>3.59</td>
<td>.64</td>
<td>70</td>
</tr>
<tr>
<td>Q16</td>
<td>Estimation</td>
<td>64</td>
<td>3.62</td>
<td>.65</td>
<td>72</td>
</tr>
<tr>
<td>Q17</td>
<td>Understanding and using charts and graphs</td>
<td>64</td>
<td>3.75</td>
<td>.59</td>
<td>72</td>
</tr>
<tr>
<td>Q18</td>
<td>Identifying, classifying, comparing shapes</td>
<td>64</td>
<td>3.61</td>
<td>.68</td>
<td>70</td>
</tr>
<tr>
<td>Q19</td>
<td>Measurement</td>
<td>64</td>
<td>3.45</td>
<td>.56</td>
<td>72</td>
</tr>
<tr>
<td>Q20</td>
<td>Patterns</td>
<td>64</td>
<td>3.47</td>
<td>.59</td>
<td>72</td>
</tr>
<tr>
<td><strong>Subscale 6: Change in practice in English language arts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q21</td>
<td>Constructed response</td>
<td>65</td>
<td>4.25</td>
<td>.66</td>
<td>76</td>
</tr>
<tr>
<td>Q22</td>
<td>Preparing students for test</td>
<td>65</td>
<td>4.12</td>
<td>.76</td>
<td>76</td>
</tr>
<tr>
<td>Q23</td>
<td>Multiple choice, fill-in-the-blank, matching</td>
<td>65</td>
<td>3.29</td>
<td>.93</td>
<td>76</td>
</tr>
<tr>
<td>Q24</td>
<td>Cooperative/small group learning</td>
<td>65</td>
<td>3.17</td>
<td>.72</td>
<td>76</td>
</tr>
<tr>
<td>Q25</td>
<td>Innovative instructional strategies</td>
<td>65</td>
<td>3.29</td>
<td>1.03</td>
<td>74</td>
</tr>
<tr>
<td>Q26</td>
<td>School-wide or grade-level planning</td>
<td>65</td>
<td>3.22</td>
<td>1.01</td>
<td>76</td>
</tr>
</tbody>
</table>
Table 10 Continued

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Low FRL%</th>
<th>Medium FRL%</th>
<th>High FRL%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Q27 Project work</td>
<td></td>
<td>65</td>
<td>2.97</td>
<td>.81</td>
<td>76</td>
</tr>
<tr>
<td>Q28 Drill to build/review skills</td>
<td></td>
<td>65</td>
<td>3.05</td>
<td>.78</td>
<td>76</td>
</tr>
<tr>
<td>Q29 Activities involving higher-order thinking</td>
<td></td>
<td>65</td>
<td>4.08</td>
<td>.71</td>
<td>76</td>
</tr>
<tr>
<td>Q30 Student choice of topics to study</td>
<td></td>
<td>65</td>
<td>2.71</td>
<td>.90</td>
<td>75</td>
</tr>
<tr>
<td>Subscale 7: Change in practice in mathematics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q31 Constructed response</td>
<td></td>
<td>65</td>
<td>4.18</td>
<td>.63</td>
<td>76</td>
</tr>
<tr>
<td>Q32 Preparing students for test</td>
<td></td>
<td>65</td>
<td>4.11</td>
<td>.77</td>
<td>76</td>
</tr>
<tr>
<td>Q33 Multiple choice, fill-in-the-blank, matching</td>
<td></td>
<td>65</td>
<td>3.31</td>
<td>.86</td>
<td>76</td>
</tr>
<tr>
<td>Q34 Cooperative/small group learning</td>
<td></td>
<td>65</td>
<td>3.22</td>
<td>.78</td>
<td>76</td>
</tr>
<tr>
<td>Q35 Innovative instructional strategies</td>
<td></td>
<td>64</td>
<td>3.39</td>
<td>.90</td>
<td>75</td>
</tr>
<tr>
<td>Q36 School-wide or grade-level planning</td>
<td></td>
<td>65</td>
<td>3.26</td>
<td>.97</td>
<td>76</td>
</tr>
<tr>
<td>Q37 Project work</td>
<td></td>
<td>65</td>
<td>2.82</td>
<td>.75</td>
<td>76</td>
</tr>
<tr>
<td>Q38 Drill to build/review skills</td>
<td></td>
<td>65</td>
<td>3.31</td>
<td>.85</td>
<td>76</td>
</tr>
<tr>
<td>Q39 Activities involving higher-order thinking</td>
<td></td>
<td>65</td>
<td>4.18</td>
<td>.70</td>
<td>76</td>
</tr>
<tr>
<td>Q40 Student choice of topics to study</td>
<td></td>
<td>65</td>
<td>2.71</td>
<td>.86</td>
<td>75</td>
</tr>
</tbody>
</table>
decreased somewhat/stayed about the same to increased somewhat/increased greatly. The means for all teachers ranged from 2.71 to 4.38 indicating a variation in levels of change from decreased somewhat/stayed about the same to increased somewhat/increased greatly. The similarities among means in each socioeconomic category and within the data in the total column would suggest that teachers perceived that, overall, some changes have been made in curriculum and instruction but that there may not be significant differences among teachers in schools of differing socioeconomic levels.

Socioeconomic Analysis

The means and standard deviations for each of the curriculum and instruction subscales, organized by the socioeconomic level of schools, are shown in Table 11. A one-way analysis of variance (ANOVA) was used to compare the subscale means in each of the socioeconomic strata. Cases with missing data within a subscale were excluded from this analysis (Table 12). Subscale 5, change in school emphasis in mathematics, was found to be statistically significant, $F(2, 199) = 5.71, p < .004$. A Tukey HSD test indicated that the mean for medium FRL% schools (3.92) was significantly higher than the mean for low FRL% schools (3.63). The mean for high FRL% schools (3.73) did not significantly differ from the mean of either group. This analysis does not support the hypothesis that there is a relationship between teachers' perceptions of changes in curriculum and instruction, due the NHEIAP grade three test, in schools of differing socioeconomic status since only one of seven subscales designed to measure this effect was found to support this hypothesis. Moreover, the relationship between perceived changes in curriculum and
Table 11

Curriculum and Instruction: Subscale Means and Standard Deviations

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Low FRL%</th>
<th>Medium FRL%</th>
<th>High FRL%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N  M</td>
<td>SD</td>
<td>N  M</td>
<td>SD</td>
</tr>
<tr>
<td>Subscale 1: Shift in reading, writing, math instructional time</td>
<td>65 3.85</td>
<td>.71</td>
<td>75 3.86</td>
<td>.46</td>
</tr>
<tr>
<td>Subscale 2: Shift in non-tested academic instructional time</td>
<td>65 2.72</td>
<td>.68</td>
<td>76 2.68</td>
<td>.68</td>
</tr>
<tr>
<td>Subscale 3: Shift in other non-tested instructional time</td>
<td>60 2.92</td>
<td>.35</td>
<td>72 2.94</td>
<td>.26</td>
</tr>
<tr>
<td>Subscale 4: Change in school emphasis in English language arts</td>
<td>62 4.01</td>
<td>.57</td>
<td>70 4.03</td>
<td>.47</td>
</tr>
<tr>
<td>Subscale 5: Change in school emphasis in mathematics</td>
<td>64 3.63</td>
<td>.46</td>
<td>68 3.92</td>
<td>.52</td>
</tr>
<tr>
<td>Subscale 6: Change in practice in English language arts</td>
<td>65 3.41</td>
<td>.44</td>
<td>73 3.57</td>
<td>.28</td>
</tr>
<tr>
<td>Subscale 7: Change in practice in mathematics</td>
<td>64 3.45</td>
<td>.42</td>
<td>74 3.53</td>
<td>.34</td>
</tr>
</tbody>
</table>
Table 12

Curriculum and Instruction Subscales: Analysis of Variance

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subscale 1: Shift in reading, writing, math instructional time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.01</td>
<td>2</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>Within Groups</td>
<td>76.90</td>
<td>214</td>
<td>.36</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>76.91</td>
<td>216</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subscale 2: Shift in non-tested academic instructional time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.22</td>
<td>2</td>
<td>.11</td>
<td>.24</td>
</tr>
<tr>
<td>Within Groups</td>
<td>96.74</td>
<td>215</td>
<td>.45</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>96.96</td>
<td>217</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subscale 3: Shift in other non-tested instructional time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.18</td>
<td>2</td>
<td>.09</td>
<td>.71</td>
</tr>
<tr>
<td>Within Groups</td>
<td>25.37</td>
<td>206</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>25.55</td>
<td>208</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subscale 4: Change in school emphasis in English language arts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.01</td>
<td>2</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>Within Groups</td>
<td>66.36</td>
<td>204</td>
<td>.33</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>66.37</td>
<td>206</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subscale 5: Change in school emphasis in mathematics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>2.83</td>
<td>2.00</td>
<td>1.42</td>
<td>5.71**</td>
</tr>
<tr>
<td>Within Groups</td>
<td>49.31</td>
<td>199.00</td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>52.14</td>
<td>201.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subscale 6: Change in practice in English language arts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.86</td>
<td>2</td>
<td>.43</td>
<td>2.42</td>
</tr>
<tr>
<td>Within Groups</td>
<td>36.51</td>
<td>207</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>37.36</td>
<td>209</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subscale 7: Change in practice in mathematics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.56</td>
<td>2</td>
<td>.28</td>
<td>1.62</td>
</tr>
<tr>
<td>Within Groups</td>
<td>35.92</td>
<td>208</td>
<td>.17</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>36.48</td>
<td>210</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Indicates significance at the $p < .004$ level.
instruction and the NHEIAP grade three test were only found to be significant between the most affluent communities; there was no significant relationship with schools with the greatest percentage of low income students as measured by the percentage of students participating the free and reduced price lunch program.

Subscale Level Data

Observable differences among the total group means indicated that teachers perceive that instructional time on tested curriculum *stayed about the same increased somewhat* (3.85), while non-tested academic instructional time *decreased somewhat stayed about the same* (2.72). Teachers also indicated that they believed that other non-tested instructional time *decreased somewhat stayed about the same* (2.91). Changes in schools’ emphasis in English language arts was perceived to have *increased somewhat increased greatly* (4.02) while changes in emphasis in mathematics stayed *about the same increased somewhat* (3.76). Teachers reported changes in practice in English language arts at the *stayed about the same increased somewhat* (3.50) level. Changes in practice in mathematics was also reported at the *stayed about the same increased somewhat* (3.52) level.

Item Level Data

Teachers perceived that their emphasis on writing (4.09) *increased somewhat increased greatly* while reading (3.61) and mathematics (3.86) *stayed about the same increased somewhat*. Both non-tested academic subjects, social studies and science, received similar ratings (2.71 and 2.74 respectively) at the *decreased somewhat stayed about the same level*. A similar pattern emerged for other non-tested instructional
time where teachers perceived art (2.90), music (2.90), and physical education (2.92) at the decreased somewhat stayed about the same level.

Teachers also perceived that writing (4.33) received the greatest change in emphasis at the school level with an increased somewhat increased greatly rating. Change in school-wide emphasis in reading and literature (4.03) was also rated in the increased somewhat increased greatly category. Change in emphasis in listening and viewing (3.73) was slightly lower with a stayed about the same increased somewhat rating. In mathematics, problem solving and reasoning (4.38) was the highest rated area of emphasis at the increased somewhat increased greatly level. All other items (3.37 – 3.93) in this subscale were rated at the stayed about the same increased somewhat level.

Three items were rated at the increased somewhat increased greatly level in the subscale six, change in practice in English language arts: (1) constructed response (4.13), (2) preparing for the test (4.13), and (3) activities involving higher-order thinking skills (4.08). The lowest rated item in this subscale was students' choice of topics to study (2.83), which was rated by teachers at the decreased somewhat stayed about the same level. All other items (3.05 – 3.58) received ratings in the stayed about the same increased somewhat range. Subscale seven, change in practice in mathematics, was very similar to subscale six in the manner in which teachers rated each item. The same three items received the highest ratings: (1) constructed response (4.06), (2) preparing for the test (4.10), and (3) activities involving higher-order thinking skills (4.21). The item receiving the lowest rating, student choice of topics to study (2.81), was also the same as in subscale six at the decreased somewhat stayed about the same level. Again, mirroring
subscale six, the remaining items in subscale seven (3.01 – 3.69) received ratings in the range.

Sources of Pressure

General Description of Data

Data for sources of pressure was collected from nine closed response items organized into two subscales related to internal sources of pressure and external sources of pressure. For each of the nine items teachers were asked to rate the level of change caused by the NHEIAP third grade test on a five-point Likert scale labeled in the following manner: (1) no pressure, (2) no indicator, (3) some pressure, (4) no indicator, and (5) great pressure.

The means and standard deviations for all items in the curriculum and instruction subscales are shown in Table 13. The data are organized by the socioeconomic level of local school communities as measured by the percentage of students qualifying for the free and reduced price school lunch program (FRL%). A heading labeled as “Total” includes summary data for all socioeconomic levels. The means for teachers working in low FRL% school ranged from 2.53 to 4.28 indicating a variation in levels of pressure from less than some pressure to less than great pressure. The means for teachers working in medium FRL% school ranged from 2.68 to 4.09 indicating a variation in levels of pressure from less than some pressure to less than great pressure. The means for teachers working in high FRL% school ranged from 2.81 to 4.00 indicating a variation in levels of pressure from less than some pressure to less than great pressure.
<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Low FRL%</th>
<th>Medium FRL%</th>
<th>High FRL%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Subscale 8: Internal sources of pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q41</td>
<td>Myself</td>
<td>65</td>
<td>4.05</td>
<td>.87</td>
<td>76</td>
</tr>
<tr>
<td>Q42</td>
<td>Other teachers</td>
<td>64</td>
<td>2.53</td>
<td>1.22</td>
<td>76</td>
</tr>
<tr>
<td>Q43</td>
<td>Principal</td>
<td>65</td>
<td>3.66</td>
<td>1.09</td>
<td>76</td>
</tr>
<tr>
<td>Q44</td>
<td>District Administrators</td>
<td>65</td>
<td>4.06</td>
<td>1.00</td>
<td>76</td>
</tr>
<tr>
<td>Subscale 9: External sources of pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q45</td>
<td>New Hampshire DOE</td>
<td>61</td>
<td>3.44</td>
<td>1.35</td>
<td>74</td>
</tr>
<tr>
<td>Q46</td>
<td>State Gov't. Officials</td>
<td>62</td>
<td>3.24</td>
<td>1.40</td>
<td>74</td>
</tr>
<tr>
<td>Q47</td>
<td>Parents</td>
<td>65</td>
<td>3.43</td>
<td>.93</td>
<td>76</td>
</tr>
<tr>
<td>Q48</td>
<td>Community</td>
<td>64</td>
<td>3.66</td>
<td>1.01</td>
<td>76</td>
</tr>
<tr>
<td>Q49</td>
<td>Media</td>
<td>64</td>
<td>4.28</td>
<td>.90</td>
<td>76</td>
</tr>
</tbody>
</table>
The means for all teachers ranged from 2.68 to 4.10 indicating a variation in levels of pressure from *less than some pressure* to *less than great pressure*. The similarities among means in each socioeconomic category and with the data in total column would suggest that teachers perceive that, overall, pressure to improve scores exists, but that there may not be significant differences among schools of differing socioeconomic levels in the amount of pressure perceived by teachers to be exerted by these sources.

**Socioeconomic Analysis**

The mean and standard deviations for internal and external sources of pressure subscales, organized by the socioeconomic level of schools, are shown in Table 14. A one-way analysis of variance (ANOVA) was used to compare the subscale means in each of the socioeconomic strata (Table 15). Cases with missing data within a subscale were excluded from this analysis. This analysis did not support the hypothesis that there is a relationship between teachers' perceptions of the amount of pressure exerted by internal and external individuals or groups to improve test scores due to the NHEIAP grade three test in schools of differing socioeconomic status.

**Subscale Level Data**

The total group mean for subscale eight indicated that teachers perceived that internal sources of pressure exerted *more than some pressure* (3.51) to improve student scores. Teachers also indicated that they believed that external sources of pressure exerted a similar level of pressure to improve students' scores (3.41).

**Item Level Data**

Observable differences among the total group means indicated that teachers perceived that the internal sources of greatest pressure to improve test scores came from

---

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 14

Sources of Pressure: Subscale Means and Standard Deviations

<table>
<thead>
<tr>
<th>Subscale 8: Internal sources of pressure</th>
<th>Low FRL%</th>
<th>Medium FRL%</th>
<th>High FRL%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>N M SD</td>
<td>N M SD</td>
<td>N M SD</td>
<td>N M SD</td>
<td>N M SD</td>
</tr>
<tr>
<td>64 3.57 .77</td>
<td>76 3.47 .85</td>
<td>77 3.50 0.96</td>
<td>217 3.51 .87</td>
<td></td>
</tr>
<tr>
<td>Subscale 9: External sources of pressure</td>
<td>Low FRL%</td>
<td>Medium FRL%</td>
<td>High FRL%</td>
<td>Total</td>
</tr>
<tr>
<td>N M SD</td>
<td>N M SD</td>
<td>N M SD</td>
<td>N M SD</td>
<td>N M SD</td>
</tr>
<tr>
<td>61 3.62 .86</td>
<td>73 3.41 .97</td>
<td>74 3.22 1.18</td>
<td>208 3.41 1.03</td>
<td></td>
</tr>
</tbody>
</table>

Table 15

Sources of Pressure Subscales: Analysis of Variance

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subscale 8: Internal sources of pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>0.34</td>
<td>2</td>
<td>0.17</td>
<td>0.22</td>
</tr>
<tr>
<td>Within Groups</td>
<td>162.44</td>
<td>214</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>162.78</td>
<td>216</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subscale 9: External sources of pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>5.23</td>
<td>2</td>
<td>2.62</td>
<td>2.51</td>
</tr>
<tr>
<td>Within Groups</td>
<td>213.88</td>
<td>205</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>219.11</td>
<td>207</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
themselves (3.93), district administrators (3.89), and principals (3.56). Each of these three sources were rated above the *some pressure* level. The internal source rated as providing the least amount of pressure to increase student test scores was other teachers (2.68) which was rated below the *some pressure* level. The external source of pressure to increase student test scores which received the highest rating by teachers was the newspaper/media category (4.10) which was rated as *less than great pressure*. All other external sources of pressure (3.09 – 3.43) were rated above the *some pressure* level.

**Attitude**

*General Description of Data*

Data for attitude toward the NHEIAP third grade test was collected from ten closed response items organized into two subscales: (1) test reflects school quality, and (2) attitude toward test. For each of the ten items teachers were asked to rate their attitude regarding the NHEIAP third grade test on a five-point Likert scale labeled in the following manner: (1) *disagree*, (2) *somewhat disagree*, (3) *neither agree nor disagree*, (4) *somewhat agree*, and (5) *strongly agree*.

The means and standard deviations for all items in the curriculum and instruction subscales are shown in Table 16. The data are organized by the socioeconomic level of local school communities as measured by the percentage of students participating in the free and reduced price school lunch program (FRL%). A heading labeled as “Total” includes summary data for all socioeconomic levels. The means for teachers working in low FRL% school ranged from 2.00 to 4.09 indicating a variation in levels of attitude from *somewhat disagree* to *somewhat agree strongly agree*. The means for teachers
Table 16

Attitude Subscales: Item Means and Standard Deviations

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Low FRL%</th>
<th>Medium FRL%</th>
<th>High FRL%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N  M</td>
<td>N  M</td>
<td>N  M</td>
<td>N  M</td>
</tr>
<tr>
<td>Subscale 10: Test results reflects school quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q50</td>
<td>NHEIAP results reflect individual student achievement</td>
<td>65 2.55</td>
<td>75 2.51</td>
<td>77 2.38</td>
<td>217 2.47</td>
</tr>
<tr>
<td>Q51</td>
<td>NHEIAP results reflect overall student achievement</td>
<td>65 2.82</td>
<td>75 2.56</td>
<td>77 2.60</td>
<td>217 2.65</td>
</tr>
<tr>
<td>Q52</td>
<td>NHEIAP results reflect differences among schools</td>
<td>65 2.38</td>
<td>76 2.45</td>
<td>76 2.55</td>
<td>217 2.47</td>
</tr>
<tr>
<td>Q53</td>
<td>NHEIAP reflect the quality of instruction</td>
<td>65 2.00</td>
<td>76 2.11</td>
<td>77 2.03</td>
<td>218 2.05</td>
</tr>
<tr>
<td>Subscale 11: Attitude toward test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q56</td>
<td>NHEIAP results influenced by changes in instructional practices</td>
<td>64 4.08</td>
<td>76 3.95</td>
<td>77 3.95</td>
<td>217 3.99</td>
</tr>
<tr>
<td>Q57</td>
<td>NHEIAP results influenced by alignment with state frameworks</td>
<td>64 4.09</td>
<td>76 3.79</td>
<td>77 4.01</td>
<td>217 3.96</td>
</tr>
<tr>
<td>Q58</td>
<td>NHEIAP provides feedback on how well I am teaching</td>
<td>65 2.52</td>
<td>76 2.39</td>
<td>77 2.66</td>
<td>218 2.53</td>
</tr>
<tr>
<td>Q59</td>
<td>NHEIAP has produced positive instructional change</td>
<td>64 3.02</td>
<td>75 3.19</td>
<td>77 3.12</td>
<td>216 3.11</td>
</tr>
<tr>
<td>Q60</td>
<td>NHEIAP has produced positive curricular change</td>
<td>64 3.27</td>
<td>76 3.25</td>
<td>77 3.30</td>
<td>217 3.27</td>
</tr>
<tr>
<td>Q61</td>
<td>I support the NHEIAP test</td>
<td>65 2.60</td>
<td>75 2.79</td>
<td>77 2.73</td>
<td>217 2.71</td>
</tr>
</tbody>
</table>
working in medium FRL% schools ranged from 2.11 to 3.95, indicating a variation in levels of attitude from *somewhat disagree* neither agree nor disagree to *neither agree nor disagree somewhat agree*. The means for teachers working in high FRL% school ranged from 2.03 to 4.01 indicating a variation in levels of attitude from *somewhat agree neither agree nor disagree* to *somewhat agree strongly agree*. The means for all teachers ranged from 2.05 to 3.99 indicating a variation in levels of attitude from *somewhat disagree neither agree nor disagree* to *neither agree nor disagree somewhat agree*. The similarities among means in each socioeconomic category would seem to indicate that there may not be significant differences among teachers' attitudes in schools of differing socioeconomic status regarding the NHEIAP grade three test.

**Socioeconomic Analysis**

The mean and standard deviations for internal and external sources of pressure subscales, organized by the socioeconomic level of schools, are shown in Table 17. A one-way analysis of variance (ANOVA) was used to compare the subscale means in each of the socioeconomic strata (Table 18). Cases with missing data within a subscale were excluded from this analysis. This analysis did not support the hypothesis that there is a relationship among teachers' attitudes in schools of differing socioeconomic status regarding the NHEIAP grade three test.

**Subscale Level Data**

The total group mean for subscale ten indicated that teachers did not agree with the use of the NHEIAP grade three test scores as an indicator of school quality. The mean for this subscale (2.40) indicated a rating of *somewhat disagree neither agree nor disagree*. The total group mean for subscale 11 indicated that teachers' attitude toward
Table 17

**Attitude: Subscale Means and Standard Deviations**

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Low FRL%</th>
<th>Medium FRL%</th>
<th>High FRL%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Subscale 10: Assessment reflects school quality</td>
<td>65</td>
<td>2.44</td>
<td>.96</td>
<td>75</td>
</tr>
<tr>
<td>Subscale 11: Attitudes toward assessment</td>
<td>62</td>
<td>3.28</td>
<td>.82</td>
<td>74</td>
</tr>
</tbody>
</table>

Table 18

**Attitude Subscales: Analysis of Variance**

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subscale 10: Assessment reflects school quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.16</td>
<td>2</td>
<td>.08</td>
<td>.08</td>
</tr>
<tr>
<td>Within Groups</td>
<td>202.68</td>
<td>213</td>
<td>.95</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>202.83</td>
<td>215</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subscale 11: Attitudes toward assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>.32</td>
<td>2</td>
<td>.16</td>
<td>.25</td>
</tr>
<tr>
<td>Within Groups</td>
<td>134.03</td>
<td>210</td>
<td>.64</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>134.35</td>
<td>212</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
the test was positive. The mean for this subscale (3.26) indicated a rating of neither agree nor disagree somewhat agree.

Item level data

The observable differences among item means in subscale ten, test reflects school quality, indicated that teachers do not agree with the use of NHEIAP grade three test scores as measures of overall student achievement (2.65), individual student achievement (2.47), differences among schools (2.47), or of the quality of instruction (2.05). All items were rated at the somewhat disagree/neither agree nor disagree level.

The two highest means in subscale 11, attitude toward assessment, were related to the test’s influence on teachers’ instructional practices (3.99) and curricular alignment with the New Hampshire state curriculum frameworks (3.96). Both items were rated at the neither agree nor disagree/somewhat agree level. Other items that were also rated in this same category, but at a lower level, were related to the test producing positive curricular change (3.27) and positive instructional change (3.11). The lowest rated items in this subscale indicated teachers’ general lack of support for the NHEIAP grade three test (2.71) and related to the test’s ability to provide feedback on the quality of teaching (2.53). Both items were rated at the somewhat disagree neither agree nor disagree level.

Items 54 and 55

General Description of the Data

Items 54 and 55 were not part of any subscale and will, therefore, be treated as separate items for the purpose of this analysis. Question 54 inquired about teachers’ perceptions regarding the effects of a local school community’s socioeconomic status
upon NHEIAP grade three test results. Question 55 sought to measure teachers’ perceptions regarding the influence of per pupil funding levels on test results. For each of the items teachers were asked to rate their attitude regarding the NHEIAP third grade test on a five-point Likert scale labeled in the following manner: (1) disagree, (2) somewhat disagree, (3) neither agree nor disagree, (4) somewhat agree, and (5) strongly agree.

The means and standard deviations for each item are shown in Table 19. The data are organized by the socioeconomic level of local school communities as measured by the percentage of students participating in the free and reduced price school lunch program (FRL%). A heading labeled as “Total” includes summary data for all socioeconomic levels. For item 54, the mean for teachers working in low FRL% school was 3.91 which translated into a rating of neither agree nor disagree/somewhat agree. The mean for teachers working in medium FRL% schools was 4.38 which translated into a rating of somewhat agree strongly agree. The mean for teachers working in high FRL% schools was 4.22 which also translated into a rating of somewhat agree strongly agree. The differences among means may indicate that there are significant differences between two or more socioeconomic strata. For item 55, the mean for teachers working in low FRL% schools was 3.38, which translated into a rating of neither agree nor disagree/somewhat agree. The mean for teachers working in medium FRL% schools was 3.64, which translated into a rating of neither agree nor disagree/somewhat agree. The mean for teachers working in high FRL% schools was 3.52, which translated into a rating of neither agree nor disagree/somewhat agree.
Table 19

Items 54 and 55: Means and Standard Deviations

<table>
<thead>
<tr>
<th>Item Description</th>
<th>Low FRL%</th>
<th>Medium FRL%</th>
<th>High FRL%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Q54 Influence of socioeconomic level on test results</td>
<td>65</td>
<td>3.91</td>
<td>1.03</td>
<td>76</td>
</tr>
<tr>
<td>Q55 Influence of per pupil funding on test results</td>
<td>64</td>
<td>3.38</td>
<td>1.15</td>
<td>76</td>
</tr>
</tbody>
</table>

Table 20

Items 54 and 55: Analysis of Variance

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 54: Test results influenced by socioeconomic status of community</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>8.967</td>
<td>2</td>
<td>4.484</td>
<td>4.871**</td>
</tr>
<tr>
<td>Within Groups</td>
<td>197.9</td>
<td>215</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>206.867</td>
<td>217</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 55: Test results influenced by level of funding per student</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Groups</td>
<td>2.53</td>
<td>2</td>
<td>1.26</td>
<td>1.04</td>
</tr>
<tr>
<td>Within Groups</td>
<td>259.63</td>
<td>214</td>
<td>1.21</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>262.16</td>
<td>216</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Indicates significance at the $p < .009$ level
Socioeconomic Analysis

The means and standard deviations for items 54 and 55, organized by the socioeconomic level of schools, are shown in Table 19. A one-way analysis of variance (ANOVA) was used to compare the subscale means in each of the socioeconomic strata. Cases with missing data within a subscale were excluded from this analysis (Table 20).

Item 54, test results are influenced by the socioeconomic status of the community, was found to be statistically significant, $F(2, 215) = 4.871, p < .009$. A Tukey HSD test indicated that the mean for low FRL% schools (3.91) was significantly lower than medium FRL% schools (4.38) and high FRL% schools (4.31). The means for high and medium FRL% schools did not significantly differ from each other. There was no significant difference among the means in item 55, test results are influenced by the level of funding per student.

Item Level Data

Question 54, test results are influenced by the socioeconomic status of the community (4.22), was rated at the somewhat agree/strongly agree level by teachers. Question 55, test results are influenced by the level of funding per student (3.52), was rated somewhat lower at the neither agree nor disagree/somewhat agree level.

Summary of Teacher Comments

Respondents were provided with an opportunity to comment on any aspect of the NHEIAP grade three test at the end of the questionnaire. Of the 218 experienced teachers who participated in this study, 133 (61%) responded with comments ranging in length from a single sentence to many paragraphs. The comments from each questionnaire
(Appendix K) were coded by topic. The topics mentioned most by teachers are summarized in Table 21. The percentage of responses for each socioeconomic category did not differ significantly. A number of comments related to the objectives of this study provided support for the previous quantitative analysis. Of the teachers who commented, (31%) perceived that the NHEIAP grade three test is inaccurately scored and does not reflect the achievement level of students. Moreover, 26% of those who commented reported that there is pressure exerted on both themselves and students to perform well on the test. Support for the test was confirmed by 20% of the teachers who provided comments while comparisons among schools was perceived to be a problem by 18% of those who commented. The alignment of curriculum and practice was considered to be important by 13% of the teachers who commented. Also mentioned by 21% of those commenting were statements related to a disagreement with instructional methods perceived by teachers as teaching to the test. Other significant findings, not related to the objectives of this study, included the need for revisions to the test by 34% of those commenting. Thirteen percent of the teachers who commented felt that the test was developmentally inappropriate for third grade students.

Summary

This chapter presented the results of an analysis of survey responses from 257 grade three teachers representing 99 schools in New Hampshire regarding their perceptions about the state’s mandated testing program. Analysis of the data revealed the following teacher perceptions: (1) the test has aligned curriculum and practice with the state’s curriculum frameworks; (2) the greatest sources of pressure to improve test scores
Table 21

Open Response Comments Summary

<table>
<thead>
<tr>
<th>Comment</th>
<th>Low FRL% (N = 37)</th>
<th>Med FRL% (N = 48)</th>
<th>High FRL% (N = 48)</th>
<th>Total (N = 133)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revisions to test needed</td>
<td>35%</td>
<td>33%</td>
<td>33%</td>
<td>34%</td>
</tr>
<tr>
<td>Test results inaccurate</td>
<td>27%</td>
<td>33%</td>
<td>31%</td>
<td>31%</td>
</tr>
<tr>
<td>Pressure on teachers/students</td>
<td>24%</td>
<td>27%</td>
<td>27%</td>
<td>26%</td>
</tr>
<tr>
<td>Teach to the test</td>
<td>19%</td>
<td>23%</td>
<td>21%</td>
<td>21%</td>
</tr>
<tr>
<td>Support test</td>
<td>19%</td>
<td>21%</td>
<td>21%</td>
<td>20%</td>
</tr>
<tr>
<td>Comparisons between schools inaccurate/unfair</td>
<td>16%</td>
<td>19%</td>
<td>19%</td>
<td>18%</td>
</tr>
<tr>
<td>Test aligned curriculum and/or practice</td>
<td>14%</td>
<td>13%</td>
<td>13%</td>
<td>13%</td>
</tr>
<tr>
<td>Test developmentally inappropriate</td>
<td>14%</td>
<td>13%</td>
<td>13%</td>
<td>13%</td>
</tr>
</tbody>
</table>

comes from district administrators and the media; (3) test results do not accurately reflect achievement or school quality; (4) teachers’ attitudes toward the test as an instrument for aligning curriculum and practice are positive; and (5) in general, teachers do not support the use of the test. There were no significant differences found among teachers’ perceptions in schools of differing socioeconomic status. However, when asked about the extent to which socioeconomic status influences test scores, teachers in more affluent communities were more likely to rate this effect as less significant than teachers in less affluent communities. A more detailed description and discussion of the findings appears in chapter five.
CHAPTER V

DISCUSSION

Overview of Study

Standards based school improvement and accountability programs now exist in all states except Iowa. While each program differs in some manner, each is organized around four elements: (1) standards that describe what students should know and be able to do, (2) classroom curriculum that is tied to the standards, (3) tests which measure student achievement, and (4) rewards and penalties directed at schools and, in some cases, students (Meier, 1999-2000). While policymakers, state officials, and school administrators all have had roles in the development and implementation, it is the teacher who shoulders the greatest burden. Teachers must assist with the alignment of curriculum, develop new teaching strategies and skills to maximize test scores, teach new curriculum to students, administer tests, interpret scores with the goal of improving instruction, and deal with the pressure created by the public interpretation of test scores. Moreover, the vast majority of teachers have little input into the content of the standards, the method of assessment, or the way in which test results are interpreted or reported. Rarely are teachers asked for their professional opinions regarding the effectiveness or appropriateness of school improvement and accountability programs.
This research was designed to examine teachers' perceptions regarding the NHEIAP grade three test. More specifically, this study sought to answer the following research questions:

1. What are the perceptions of third grade teachers regarding the NHEIAP grade three test's influence on curriculum and instruction?

2. What is the level of support among grade three teachers for the NHEIAP grade three tests?

3. From which group, or groups, do grade three teachers feel pressure to improve their students' scores on the NHEIAP grade three test?

4. Is there a relationship between changes made to the curriculum and in instruction and the socioeconomic status of the school community as defined by the percentage of students who participate in the federal free and reduced price school lunch program?

5. Is there a relationship between teachers' support for the NHEIAP grade three and the socioeconomic status of the school community in which they teach as defined by the percentage of students who participate in the federal free and reduced price school lunch program?

6. Is there a relationship between the pressure to improve their students' scores on the NHEIAP grade three test and the socioeconomic status of the school community as defined by the percentage of students who participate in the federal free and reduced price school lunch program?

The literature review presented in chapter two described the historical roots of current testing practices and discussed how tests have become such an accepted practice.
within school culture. Current studies on the ways in which tests have affected teachers, instruction, and curriculum were used to form the basis for the development of the survey instrument and related analyses. Two studies (Herman & Golan, 1991; Koretz et al., 1996b) and their survey instruments were particularly well matched to the objectives of this research and were adapted, with permission, for use with this study. The survey contained 61 closed response items, a section for general comments, and a request for demographic information. Participants were asked to rate their perceptions on curriculum, instruction, sources of pressure, attitude toward the test, and socioeconomic influences on a 5-point Likert scale.

Participants in this study included third grade teachers who were randomly sampled from schools divided into three socioeconomic categories as determined by the percentage of students who received free and reduced price lunches. The response rate for this study was 83%. Confirmatory factor analysis was used to verify the constructs described by each subscale. A reliability analysis was performed to ensure the internal consistency of subscale items. Frequencies, means, and standard deviations were presented for each socioeconomic category and for all respondents. Parametric statistics were used to determine whether there were significant differences among socioeconomic categories.

Discussion of Findings

Curriculum and Instruction

Madaus (1988) identified seven principles of the power of testing on the curriculum, teachers, and schools. Of these seven principles, the following four are
directly related to the changes in curriculum and instruction addressed by this study: (1) curriculum will be defined by the content of high stakes tests, (2) instruction will be changed to mirror the types of questions used, (3) teachers will teach to the test when important decisions are connected to test scores, and (4) the agency that controls test content controls the curriculum. These four principles will be used to organize the discussion on how the NHEIAP grade three test has influenced curriculum and instruction in New Hampshire's elementary schools.

Madaus' first principle suggests that the NHEIAP third grade test will define the content of each school's curriculum. There is, in fact, evidence from the results of this study that this has occurred as teachers reported that there was a significant shift in instructional time in their classrooms associated with the tested curriculum in reading, writing, and mathematics. At the school level, teachers perceived that instructional emphasis has increased in both English language arts and mathematics. English language arts was perceived as the area receiving the greatest emphasis among the tested curriculum. When asked to indicate which area within English language arts received the most attention, teachers indicated that writing received the most emphasis, followed closely by reading and literature. Listening and viewing were perceived to receive less emphasis than either writing or reading and literature. Within mathematics, problem-solving and reasoning was identified as the area receiving the greatest increase in emphasis. Each of the other tested mathematical topics, fractions, decimals, estimation, charts and graphs, shapes, measurement, and patterns, were perceived to have received increased emphasis but to a lesser degree than problem-solving and reasoning.
Increases in emphasis on tested curriculum causes other curricular areas to share a smaller portion of instructional time resulting in a narrowing of the curriculum. There is some evidence that teachers perceived that this has occurred in New Hampshire elementary schools. Teachers indicated that they had added the most time to instruction related to writing, followed by increases in reading and mathematics. As a result, teachers reported that instructional time in science and social studies has decreased as a result of the NHEIAP grade three test. Moreover, other curricular areas such as art, music, and physical education were also identified by teachers as receiving less instructional time.

The shift in emphasis toward tested curricula has also been reported in studies conducted in Arizona (Smith, 1991), Kentucky (Koretz et al., 1996b), Maryland (Corbett & Wilson, 1991; Koretz et al., 1996a), North Carolina (Jones et al., 1999), Pennsylvania (Corbett & Wilson, 1991), Texas (Gordon & Reese, 1997), and Vermont (Koretz, Stecher, Klein, & McCaffrey, 1994a). The results of these studies, along with the results described in this study, corroborate the fact that state-mandated assessments are effective tools for aligning curriculum and practice with tested content. For example, participants in this study stated that they felt that the NHEIAP grade three test had assisted in aligning curriculum and practice with the state frameworks. However, the fact that teachers perceived that the NHEIAP third grade test also caused a de-emphasis in non-tested curriculum is a cause for concern. Teachers reported that they felt that too much emphasis was being placed on test preparation and the teaching of tested content. Resnick and Resnick (1992) described the power of tests to align and simultaneously

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
narrow curricula in the following manner: what is tested will be taught and what is not tested is not taught.

Madaus' second principle predicted that the NHEIAP grade three test would cause instructional practices to change to mirror the types of questions used in the test. Again, there is evidence that Madaus' principles have held true for this study. In general, the subscales that describe changes in instructional practice indicated that there had been significant change in instructional practice because of the NHEIAP grade three test. Four instructional practices associated with the test - the use of constructed response activities, test preparation activities, multiple choice activities, and activities involving higher order thinking skills - were all identified as areas of significant increase by teachers. But the test did not seem to inhibit the use of other instructional practices not directly related to test questions. The use of cooperative groups, innovative instructional strategies, support for teacher planning, project work, and activities designed for building and drilling basic skills increased slightly. The only practice identified by teachers as decreasing in use was activities where students were allowed to choose their own topics of study. This finding is not surprising given the previous data which indicated that the test had caused a narrowing of the curriculum.

Madaus' final two principles regarding changes in curriculum and instruction brought about by standardized testing programs were not directly measured by the survey used in this study. However, the data did reveal that the NHEIAP third grade test has influenced curriculum and instruction in a manner consistent with Madaus' principles. Madaus' third principle, teachers will teach to the test when important decisions are connected to test scores, is supported by the survey item that asked if there had been
changes in instructional approaches related to preparing students to do well on the test. This item was one of the highest rated items within both the English language arts and mathematics subscale on changes in instructional practices. The score for these items indicated that teachers had significantly increased the number of activities and practices that prepared students to do well on the test. Also, the fact that teachers have aligned the curriculum with the state curriculum frameworks indicates a willingness to teach to the test. Additionally, 21% of the teachers who commented mentioned that they or their colleagues had been teaching to the test. There is also support for Madaus' fourth principle, the agency that controls test content controls the curriculum. Again, no data were collected which dealt directly with this principle, but teachers have indicated that the curriculum has been aligned with the state’s curriculum frameworks and that instructional practices have also been influenced, which supports the notion that the New Hampshire Department of Education is now the agency which controls local curriculum and instructional practices.

Sources of Pressure

Popham (1987) has defined high-stakes tests as assessment instruments whose results are used in at least one of two ways: (1) to make placement or promotion decisions about students, or (2) as a method for determining the quality of instruction. In either case, high-stakes tests can exert considerable pressure on educators and students to produce acceptable results. The results of this study indicate that the NHEIAP grade three test is considered by teachers to be a high-stakes test. Individual schools or school districts may decide whether to use NHEIAP data for decisions regarding student placement or promotions. However, it is clear that the NHEIAP third grade test is
considered by teachers to be a high-stakes test since the media, school administration, and local communities have used the results as measures of the quality of instruction. Teachers reported that, other than themselves, the two most significant sources of pressure from within the school have come from district administrators and their building principal. And, while teachers reported that they have placed the most pressure on themselves, they perceived that the least amount of pressure has come from their colleagues. Outside sources which exerted moderate amounts of pressure on teachers to improve scores included New Hampshire Department of Education officials, parents, the local community, and state officials. The most significant source of pressure, from either internal or external sources, came from the media.

Herman and Golan (1991), in their study of teachers' perceptions regarding the effects of standardized testing in nine states, identified district administrators and the media as the strongest sources of pressure to improve scores. Excluding teachers' self-imposed pressure to improve scores, New Hampshire teachers also identified district administrators and the media as the sources of greatest pressure. The fact that district administrators have sought to improve test scores comes as no surprise, since they are likely to be directly influenced by media scrutiny and the resulting community pressure to improve scores. Teachers understand this but do not agree with the emphasis placed on test results and the resulting pressure to improve scores. One teacher commented, “The negative aspect is that too much emphasis has been put on the test by not only the school board, but the administrators. I also feel that the media has done an injustice to many schools.” Teachers feel betrayed by the public scrutiny of their work when based only on test results since the media's comparison of schools is considered both unfair and
unwarranted. The increased emphasis has led inevitably to increased pressure on
teachers to teach to the test. One teacher voiced her frustration with this reality in the
following manner:

It is unfortunate that schools, parents, SAU offices, and communities in
general are using/abusing/misusing the results of NHEIAP. The entire
focus on 3rd grade instruction is on the results. We have been instructed
that if a skill is not on the test don’t spend any extra time on it. Sad but
true. The pressure on the children is too much. Why not just teach to the
test and let everything else go by the wayside?

Attitude

Three general conclusions about teachers’ attitudes toward the NHEIAP grade
three test can be drawn from the data. The first is that teachers do not believe that the
NHEIAP grade three test results reflect the level of achievement or the quality of
instruction in their school. Teachers indicated that they do not support the idea that test
scores are accurate indicators of either individual or overall student achievement, nor do
they believe that school to school comparisons are valid measures of quality. There is a
level of distrust for what the scores actually measure, as well as evidence that teachers
believe that some tests are scored inaccurately. Teachers have indicated that they believe
that many of the questions on the test are not clear, or as one teacher noted, “I feel many
of the questions are deliberately set up to trick students rather than accurately test a skill.”
Other teachers have reported discrepancies between students’ classroom performance and
the achievement levels indicated by test scores.

I don’t see a great degree of correlation between the students’ classroom
performance and the test scores. Some students in last year’s class, for
example, performed at a proficient level during the year but only scored at
the basic level on the test.

Other teachers have observed that accurate scoring of the test is not assured.
I have noticed inconsistencies in scoring of the writing prompts. Longer pieces were scored higher than shorter pieces, even if the shorter piece was of better quality. It appears that understanding and answering the prompt is not important. A child wrote a piece entirely off topic, and still scored well. Different versions of the test appear to be more difficult than others.

The point of this discussion is not to question the validity or reliability of the test nor to call into question whether scoring procedures are accurate, but to highlight the fact that teachers do not perceive that this test is an accurate reflection of achievement or of the quality of instruction.

The second conclusion that can be drawn from the attitude subscales is that teachers believe that test scores are positively influenced by the degree to which their school's curriculum is aligned to the state's curriculum frameworks. This is an entirely expected result because the content of the test is based on New Hampshire's curriculum frameworks, and, as indicated by the data on curriculum and instruction, most schools have aligned their curriculum and instructional practices. Nevertheless, to assume that all of the content within the curriculum frameworks is being taught has not been supported. Teachers commented that they felt pressured to teach to the content of the test, not that of the curriculum frameworks. In addition, many teachers stated that they were frustrated with the changing content of each year's test, since they based current instruction on the previous year's test only to find a different curricular emphasis in the new version of the test. Teachers also stated that they felt that they could not cover all of the required material by the testing date. One teacher's comments summarized the dilemma of whether to teach to the frameworks or to the content of the test: "Teaching toward meeting the state standards is useful. However, we seem to be more concerned with
teaching toward the test that ‘proves we are meeting the standards.’ What will the emphasis be this year?”

The third conclusion that can be drawn from the data on attitudes toward the NHEIAP grade three test is that teachers are not supportive of the test or the manner in which test results are used. This conclusion is substantiated by the comments volunteered by participants in this study. Over 80% of the comments offered by teachers were critical of the test. While 20% of the teachers who commented supported the use of the test, a large portion of this group also stated that revisions to the test were needed. Among the revisions were suggestions that pertained to changes in the way that students with learning disabilities are tested and scored, reductions in the length of the test, and revision of “tricky” or confusing questions. A number of teachers also mentioned that they felt that the test was developmentally inappropriate for third grade students. Most of these teachers felt that the beginning of grade four would be a more appropriate time to assess the material on the current test.

Two sets of seemingly contradictory data also corroborate the assertion that teachers do not support the NHEIAP grade three test. Teachers indicated that the NHEIAP grade three test had produced positive changes in instruction and curriculum while at the same time stating that, in general, they did not support the test. Logically, it could be expected that favorable ratings in curriculum and instruction would lead to positive attitudes regarding the test. However, given teachers’ beliefs about the test’s deficiencies it is likely that they did not believe that the test was the appropriate method for changing curriculum and instruction. In other words, the fact that positive changes have resulted from administration of the test is not necessarily associated with support for
the test. Other methods for bringing about curricular and instructional change could have been used without the negative consequences identified by the teachers who participated in this study.

**Socioeconomic Status**

This study sought to determine if there was a relationship between teachers’ perceptions in three categories (changes in curriculum and instruction, level of support for the test, and sources of pressure to improve test scores) and the socioeconomic status of communities. The analysis of data did not support the premise that teachers would respond differently depending on the socioeconomic status of the communities in which they taught. When asked if per pupil funding influenced test scores, there were no statistical differences among groups, but all teachers reported that they perceived that there was at least a moderate degree of agreement that funding levels did affect test scores.

There was a significant difference among socioeconomic groups when teachers were asked whether the socioeconomic status of the community, or communities, which a school serves influences test scores. Teachers in more affluent communities agreed that there was some effect on test scores while teachers in less affluent communities believed that there was a stronger relationship between test scores and socioeconomic status. The reason for these differences might be explained by the fact that teachers in less affluent communities see a greater percentage of students from poorer families than do teachers in more affluent areas of the state. As described earlier in this study, Hall (1999) did find that students in poorer communities scored lower on the NHEIAP grade three test than their counterparts in more affluent communities. Teachers from less affluent
communities are more likely to perceive this relationship given the greater percentage of
students from poorer families. A teacher who teaches in a lower socioeconomic
community described her frustration in the following comment:

   Over the past four years my district has spent incredible hours looking
   over the results, targeting the weakest areas, and altering instruction. I
   have served on every committee to help improve grade three scores. We
   continue to score low. The increase is so slight. The bulk score
   novice/low basic.

Teachers in more affluent communities have fewer students from poor families, and they
are more likely to have the resources to address the needs of these students when they do
appear in the classroom. The exact reasons for the differences among teachers working in
schools serving communities of differing socioeconomic strata cannot be answered
conclusively by the data collected for this study but could be a question answered by
future researchers.

General discussion

The description of third grade teachers that emerges from this study, as they learn
to deal with the changes brought about by the NHEIAP grade three test, is that of a group
who have worked to align curriculum with the state curriculum frameworks while at the
same time seeking to change instructional practices in a manner that supports the delivery
of the curricula. As the pressure to improve test scores has increased, teachers have spent
more time adjusting their classroom curriculum, which may differ from their school’s
official curriculum, to the content of the test. As described in the following comment,
teachers do not feel that they are allowed to spend the necessary time working with
students that would help to ensure understanding since the content of the test drives
teachers’ planning and teaching schedules.
After administering the test for six years, I can say that I do know what is covered on the test, and yes, I do try to prepare my students for this experience. However, the wide range of materials and concepts tested is more than I can teach in depth to most third graders. I feel like I’m skimming and hurrying through some of the basics in math to at least introduce more advanced topics so my brighter students might be able to successfully respond to those questions.

As teachers and schools struggle with the need for more time to teach tested content, they inevitably narrow the curriculum. This study has provided evidence that as curriculum and practice have changed to match the test, less time is being spent on science and social studies. The emphasis on art, music, and physical education has, in some cases, also diminished.

Teachers have moved along the natural progression from curriculum alignment to teaching-test taking skills and making instruction more test-like in order to maximize students’ chances of scoring well on the test. One teacher described her school’s method for developing students’ test taking skills:

Our district has set aside a two-week time slot in the early part of the new year for a practice NHEIAP test. Using common items from previous years, students are put through the rigor of taking the test. These items are then scored and results tallied for areas that we as teachers are to stress before the actual testing in May.

Such practices further restrict the already limited time available to teach the broad amount of material outlined in the state’s curriculum frameworks. Another teacher spoke to the issue of devoting time to test taking skills: “I feel more and more we are teaching to the test. The amount of time spent by all the third grade teachers in practice tests is ridiculous.” The NHEIAP testing program and the resulting pressure to improve scores has placed teachers in a situation where they are spending more and more time on an
endeavor that they do not believe is in the best interests of their students. As one teacher stated:

I don’t believe this type of testing is developmentally appropriate for third grade students. I believe we are wasting valuable instructional time teaching students how to take a standardized test. Unfortunately, this pre-teaching is necessary for these young children to have some skill in test taking.

Teachers, in general, are not supportive of the NHEIAP grade three test as it is currently designed, administered, and interpreted. In addition to its perceived negative effects on the curriculum and teaching, teachers doubt whether the results are accurate indicators of student achievement levels or the quality of instruction. Teachers are convinced that public comparisons among schools are not only inaccurate but unfair. Still, many teachers are faced with the fact that the test will be administered each May, that they will be held accountable for the results, and that they do not have a voice in the process.

Limitations of this Study

Research Design

A cross-sectional design was used for this study that provides a single snapshot of teachers’ perceptions at one point in time. Changes to the NHEIAP grade three test, new school administrations, shifting demographics, or the implementation of new state funding structures are a few of many variables that could change teachers’ perceptions over time. It was assumed that the sampling strategy used in this study produced a representative sample of the third grade teacher population. The fact that the sample
represents almost 40% of the third grade teachers in the state, coupled with a high return rate should ensure that this study is representative of the population.

As with any study of this type, it is possible that the perceptions of the respondents do not reflect the reality of the situation. For example, many teachers reported that more emphasis had been placed on mathematics and English language arts. This opinion was not based on empirical data and could, therefore, be called into question. The results of this study will provide researchers with a starting point from which to further investigate these issues.

Survey Instrument

The survey instrument was adapted, with permission, from similar studies (Herman & Golan, 1991; Koretz et al., 1996b) and revised to meet the particular objectives of this research. The survey was piloted prior to use in order to assess content validity. However, the length of the instrument was restricted in order to maximize return rates, which limited the quantity and depth of questioning. Also, differing interpretations of the terms used in the survey could affect results.

Generalizability

The sample for this study was selected from the State of New Hampshire which may limit generalizability because the state’s population is 98% white (U. S. Census Bureau, 1999b) with a statewide average income that is higher than the United States average (U. S. Census Bureau, 1999a). Readers of this study should also take into account New Hampshire’s lack of state funding (at the time of this survey) and local control mechanisms before generalizing the results to their state’s environment. While teachers in New Hampshire who have responsibility for administering tests in grade six
and grade ten may agree with the perceptions expressed by participants in this study.

generalizability to these populations is cautioned. Given the high return rate and the
sampling procedures used for this study, the results should be generalizable to all grade
three teachers in New Hampshire.

Recommendations for Further Research

This study represents the first attempt to uncover the way in which the NHEIAP
grade three test has affected teachers, curriculum, and instruction by asking teachers to
report their perceptions and opinions. Given this study’s exploratory nature, follow-up
research is warranted to probe the deeper meanings of the results reported here and to
uncover other effects of mandated testing. This study relied heavily on quantitative data
to determine the perceptions of third grade teachers. A follow-up study where teachers
are asked to respond to, and expand upon, the results of this study would assist in
developing a deeper understanding of the results. For example, a study could be
conducted to determine the types of curricular and extracurricular opportunities that have
been sacrificed as a result of the increased attention to testing skills and tested content.

NHEIAP tests are also administered at the sixth and tenth grade levels. Studies similar to
those described above would enable comparisons of the effects of mandated tests at each
level. The results would be useful to researchers, policymakers, and educators as they
search for the consequences of standardized state mandated tests in New Hampshire and
other states.

The goals of this study did not address the way in which test scores affect teachers
attitudes and expectations. However, there is some evidence that test scores can affect
these attitudes. Herman and Golan (1991) found that teachers in schools with increasing test scores felt there was more pressure to improve test scores than did teachers in schools with decreasing scores. A follow-up study could be designed to determine if such a relationship exists in New Hampshire.

The relationship between socioeconomic status and state mandated tests requires more study. While no relationship was found between teachers’ perceptions and socioeconomic status, it is possible that the relationship exists in certain locations (e.g., urban v. rural areas) but that it is too small to be statistically significant across the state. This theory is supported by the data from one survey item that directly asked teachers whether socioeconomic status affected students’ ability to perform on the NHEIAP grade three test. It was determined that, in fact, there were differences between the answers that were provided by teachers in communities of differing socioeconomic status. This finding should be studied further to determine whether students in less affluent communities in New Hampshire are at a disadvantage when taking the NHEIAP grade three test.

Finally, state mandated testing programs in New Hampshire and other states require tremendous amounts of resources to produce, administer, and interpret. Future research should be designed with the intention of determining whether or not these programs are prudent expenditures of time and money. If they are true measures of student achievement and stimulate meaningful improvement in our schools, then they should become permanent fixtures in educational culture. If, however, they are found to be one of many valid and reliable indicators of student achievement, then the resources,
time, and emphasis placed on mandated tests should be tempered and allocated to other
effective improvement efforts.

**General Recommendations**

Every March, 64 college basketball teams participate in a tournament designed to
determine the national champion of the sport. Inevitably, one team emerges as a
“Cinderella” team and captures the hearts and minds of the fans who actively watch this
typically small and unknown school compete and triumph over much larger, better
funded, and more talented teams. Nevertheless, the school’s quest for a championship
usually ends when a superior team ends the dream. The question is, have they failed? If
we value only the final set of numbers which are used to determine winners and losers in
the athletic area, then yes, the “Cinderella” team must be judged to be less successful
than the teams which move on in the tournament. For the fans of college basketball who
watched the team and participated in the excitement, however, a much different picture
develops. These fans would use terms like overachieving, and hardworking to describe
the team. An outside observer, who, like many who live in our culture which is
increasingly defined by numbers and statistics, may look at the final score and use terms
such as loser, underachiever, or not up to standards to describe the team.

This scenario applies equally well to publicly reported scores in New Hampshire
and in other states with similar assessment programs. Schools, teachers, and students are
increasingly judged by how their state test scores compare to other schools. If scores do
not compare favorably, regardless of the situation or increases in student achievement
observed at the classroom level, then these schools are judged as failing, inadequate, or
not up to standards. The New Hampshire Commissioner of Education has repeatedly cautioned against this practice to no avail. The findings of this research also indicate that teachers do not believe that these comparisons are justified and that this test alone is not an effective instrument for judging student achievement or the quality of a school. This research is also consistent with the findings of others researchers (Corbett & Wilson, 1991; Koretz et al., 1996a; Koretz et al., 1996b; Smith, 1991; Stecher et al., 1998).

Salganik has cautioned against ranking and comparing schools based on test scores and has depicted a set of likely outcomes if this practice is continued. As test scores are absorbed into the norms of school culture, we can expect that there will be an increasing reliance on technical evidence, increased use of tests as measures of school quality, increased regulation of schools, and a decreased reliance on professional judgment (1985).

Recommendations

The legislation that governs the development and implementation of the NHEIAP (State of New Hampshire, 1993a) includes a statement of purpose that is divided into six sections. The first purpose statement describes how the state’s political and economic health depend on the establishment of a statewide school accountability and assessment program. The remaining six statements will be used to organize the following recommendations. These recommendations are based on the results of this study and are not meant to address all facets of the state’s assessment program.

Statement of Purpose II. A statewide education improvement and assessment program built upon the establishment of educational standards specifying what students should know and be able to do is an important element in educational improvement. Such a program also serves as an effective measure of accountability when the assessment exercises or tasks
are valid and appropriate representations of the curriculum standards that students are expected to achieve.

**Recommendation.** The results of this study indicated that teachers believe that test scores are not an accurate indicator of student achievement or the quality of instruction. If the instrument that measures the standards is flawed in reality, or only perceived to be flawed, the effect will be the same: it will not serve as an effective measure of accountability. Teachers also report that valuable instructional time is used to teach test taking skills in an effort to increase scores. The possibility that some of the criticisms made by teachers are due to misunderstandings regarding the test’s design, or the curriculum standard’s intent, serves to highlight the need for a more inclusive and collaborative model of development and revision. A review of the state’s grade three test is recommended with input, of some form, from all elementary teachers. In addition, an ongoing review of the positive and negative effects of the assessment program should be initiated.

**Statement of Purpose III.** Widespread participation in the establishment of a statewide education improvement and assessment program is essential. Consultation with educators at all levels, business people, government officials, community representatives, and parents must occur in the development of educational standards. In turn, widespread dissemination of those standards, once established, must occur. Teachers, administrators, and school board members must be fully apprised of these state-developed standards. They must, in turn, communicate these expectations to students and parents, and find and implement methods to enable students to acquire and apply the requisite knowledge and skills.

**Recommendation.** Teachers have stated that they have struggled with the breadth of the state curriculum frameworks. In an effort to alleviate this problem teachers have tried to teach only the content and skills that they thought would be most useful on the test. Additionally, teachers have questioned the developmental appropriateness of the
curriculum standards and test content. A review and revision of the state's curriculum frameworks and the state assessment program should be initiated with input from all elementary teachers.

**Statement of Purpose IV.** In addition, the assessment results must be reported to students, parents, teachers, administrators, school board members, and to all other citizens of New Hampshire in order that informed decisions can be made concerning curriculum, in-service education, instructional improvement, teacher training, resource allocation, and staffing.

**Recommendation.** This study has identified the media as the greatest source of pressure to increase test scores. Also, teachers commented that they believed that comparisons among schools based on test scores were unfair and detrimental to the school improvement process. While it is easy to blame the media for reporting and ranking schools by test scores, it is also misplaced blame. Schools must share some of the responsibility for failing to develop alternative sources of assessment that can be used to augment the state's test scores. Many respondents to this study described schools that are wrestling with ways to improve NHEIAP test scores with few resources left to work on new assessment plans. The state should assume the responsibility of assisting districts with developing new methods of assessment that will demonstrate what students know and are capable of doing.

**Statement of Purpose V.** A critical part of this program is the local education improvement and assessment plan. In order for an assessment program to give an accurate picture of student performance, it must include more than a one-time measure. Local school districts should devise and implement measures that focus on the continuing growth of individual students, and report the results to parents along with those obtained from the state-developed tool.

This recommendation is closely linked to the previous recommendation. Again, many schools are working to improve NHEIAP test scores and have not developed new
methods for measuring student achievement. Very few teachers who responded to this study mentioned other indicators of student growth; when they did, they described the use of commercially available, standardized, norm referenced tests. The state should continue to support the development of local education improvement and assessment plans and intensify its efforts to assist schools in the development of new indicators of student growth.

**Statement of Purpose VI.** The purpose of the statewide education improvement and assessment program is not to establish a statewide curriculum. It is, rather, to establish what New Hampshire students should know and be able to do and to develop and implement effective methods for assessing that learning and its application so that local decisions about curriculum development and delivery can be made.

**Recommendation.** The results of this study show that teachers believed that the tested content of the NHEIAP grade three test has established a de facto statewide curriculum. The results of this study also indicated that curriculum that is not included on the test has received less emphasis. It is obvious that schools have relinquished some degree of local control of the curriculum in order to maximize test scores. Some previous recommendations, if implemented, may serve to return some measure of control of the curriculum to local schools. If schools develop new indicators of student growth and show that their students are achieving at acceptable levels, they may find that the pressure to raise test scores and the focus on teaching test related skills and content might subside. However, a collaborative process where teachers are an integral part of the design process is required in order to develop shared meanings and ownership. Again, this is an area where the state could provide schools with resources and assistance.
Concluding Remarks

New Hampshire teachers are concerned about the way in which the NHEIAP grade three test is affecting curriculum, instruction, and the students whom they teach. To ignore these voices would be to ignore the professional opinion of a large, experienced, and well-educated teaching force. Concerns were consistently voiced by teachers who teach in affluent and poor communities, and by those who work in high scoring and low scoring schools. In a state as small as New Hampshire it is inconceivable that only a small percentage of teachers were involved in the planning process. For this assessment program to be truly effective it must include teachers as partners in the improvement process and not reduce their role to that of a technician who delivers the prescribed curriculum and administers the state’s tests. As currently constituted, the NHEIAP is dependent on the statewide publication of test results to generate pressure on schools and teachers to improve test scores. The underlying assumption of such high-stakes testing programs is that increases in test scores are associated with improvements in student learning. Wagner has stated (1998), “Pressure in the form of increased emphasis on test scores may get the attention of some teachers, but it will not create the organizational learning required to obtain significantly better results” (p. 513). The development of a collaborative approach for improving the NHEIAP, with teachers as partners in the process will ensure that organizational learning will shift the focus from the technology of the test to the improvement of student learning.
LIST OF REFERENCES


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Learning Innovations. (1999). Best Schools Leadership Institute team member toolkit. Stoneham, MA: Learning Innovations a Division of WestED.


Perceptions of New Hampshire Third Grade Teachers

Directions: In this questionnaire, the term "NHEIAP test" refers to the New Hampshire Educational Improvement and Assessment Program third grade assessment. Please respond to the following questions by either circling or filling in your response. Once completed, return the survey in the pre-addressed, stamped envelope that has been provided. Please be assured that your responses will be held in the strictest of confidence. No teacher or school will be identified in the reporting of survey results.

Thank you!

In what ways, if any, have you shifted the available instructional time among the following subject areas since you first administered the NHEIAP test? Consider balancing increases in time with decreases in time.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Decreased Greatly</th>
<th>Decreased Somewhat</th>
<th>Stayed the Same</th>
<th>Increased Somewhat</th>
<th>Increased Greatly</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Reading</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2. Writing</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3. Mathematics</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4. Social Studies</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5. Science</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6. Art</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>7. Music</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>8. Physical Education</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

How has your school changed its instructional emphasis on the following areas since first administering the NHEIAP test? If you do not know how your school has changed its instructional emphasis in a particular area, label the row "don’t know."

<table>
<thead>
<tr>
<th>Subject</th>
<th>Decreased Greatly</th>
<th>Decreased Somewhat</th>
<th>Stayed the Same</th>
<th>Increased Somewhat</th>
<th>Increased Greatly</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENGLISH LANGUAGE ARTS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Reading and literature</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>10. Listening/viewing</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>11. Writing</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>MATHEMATICS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Problem solving and reasoning</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>13. Addition, subtraction, and whole numbers</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>14. Fractions</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>15. Adding and subtracting decimals</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>16. Estimation</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>17. Understanding and using charts and graphs</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>18. Identifying, classifying, and comparing geometric objects</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>19. Measurement</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>20. Patterns</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
To what extent have you changed instructional practices as a result of the NHEIAP test in English Language Arts?

<table>
<thead>
<tr>
<th></th>
<th>Decreased Greatly</th>
<th>Decreased Somewhat</th>
<th>Stayed About the Same</th>
<th>Increased Somewhat</th>
<th>Increased Greatly</th>
</tr>
</thead>
<tbody>
<tr>
<td>21. Use of instructional exercises which allow for constructed response (e.g., short essays)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>22. Emphasis on preparing students to do well on tests.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>23. Use of multiple choice, fill-in-the-blank, and matching exercises</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>24. Use of cooperative/small group learning</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>25. Implementation of innovative instructional strategies</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>26. Support for school-wide or grade-level planning</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>27. Use of extended project work (e.g., research or other projects requiring a week or more)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>28. Use of activities relying on drill to build/review basic skills</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>29. Use of instructional activities which involve higher-order thinking and problem solving</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>30. Opportunities for students to choose what they want to study</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

To what extent have you changed the following instructional practices as a result of the NHEIAP test in Mathematics?

<table>
<thead>
<tr>
<th></th>
<th>Decreased Greatly</th>
<th>Decreased Somewhat</th>
<th>Stayed About the Same</th>
<th>Increased Somewhat</th>
<th>Increased Greatly</th>
</tr>
</thead>
<tbody>
<tr>
<td>31. Use of instructional exercises which allow for constructed response (e.g., short essays)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>32. Emphasis on preparing students to do well on tests.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>33. Use of multiple choice, fill-in-the-blank, and matching exercises</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>34. Use of cooperative/small group learning</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>35. Implementation of innovative instructional strategies</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>36. Support for school-wide or grade-level planning.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>37. Use of extended project work (e.g., research or other projects requiring a week or more)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>38. Use of activities relying on drill to build/review basic skills</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>39. Use of instructional activities which involve higher-order thinking and problem solving</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>40. Opportunities for students to choose what they want to study</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
To what extent do you feel pressure from the following sources to improve your students' scores on the NHEIAP test?

<table>
<thead>
<tr>
<th>Source</th>
<th>No Pressure</th>
<th>Some Pressure</th>
<th>Great Pressure</th>
</tr>
</thead>
<tbody>
<tr>
<td>41. Myself</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>42. Other teachers</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>43. Principal</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>44. District administrators</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>45. New Hampshire Department of Education</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>46. State Government Officials (Governor, State Legislators, etc.)</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>47. Parents</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>48. Community</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>49. Newspaper/media</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

The following items inquire about your personal attitudes. Read each item and indicate your degree of agreement.

<table>
<thead>
<tr>
<th>Item</th>
<th>Disagree</th>
<th>Somewhat Disagree</th>
<th>Neither Agree nor Disagree</th>
<th>Somewhat Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>50. NHEAP test results accurately reflect individual student achievement at my school</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>51. NHEAP test results accurately reflect overall student achievement at my school</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>52. NHEAP test results accurately reflect differences in student achievement between my school and other schools in the state</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>53. NHEAP test results accurately reflect the quality of instruction at my school</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>54. NHEAP test results are significantly influenced by the socioeconomic level of the community (% of students receiving free and reduced price lunches, educational attainment level of parents, % of single parent households, etc.)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>55. NHEAP test results are significantly influenced by the level of funding per student</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>56. NHEAP test results are significantly influenced by changes in instructional approaches</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>57. NHEAP test results are significantly influenced by curriculum alignment with state curriculum frameworks</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>58. The NHEAP test gives me important feedback on how well I am teaching</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>59. In general, I believe that the NHEAP test has produced positive instructional change</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
In general, I believe that the NHEAP test has produced positive curricular change

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
</table>

In general, I support the use of the NHEAP test

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
</table>

You are welcome to provide additional comments about the NHEAP test in the space below.

Demographic Information: Please record the appropriate information in the space provided.

1. I administered the NHEAP test in May of 1998

2. Number of years teaching

3. Number of years administering NHEAP test

4. Highest degree earned

Thank you for taking the time to complete this questionnaire. By expressing your opinions regarding the NHEAP test you are helping to expand our understanding of how the NHEAP third grade test is influencing education in our state. A copy of the results of this study will be available at my website at the following address: http://www.nh.ultranet.com/~koss. Please feel free to contact me with any questions or concerns that you may have about this survey.

Sincerely,

Steve Kossakoski
P.O. Box 481
Kingston, NH 03848
Phone: 555-555-5555
E-mail: smail@e-mail

Questions adapted from the following sources, with permission:


October 8, 1999

«Administrator»
«School»
«Address»
«Town» NH «Zip»

Dear Principal «PrinLstName»:

As you are well aware, the New Hampshire Educational Improvement and Assessment Program (NHEIAP) has had a tremendous impact on all schools in the state. However, while the NHEIAP third grade assessment was first administered over five years ago, little research has been conducted on its impact at the classroom level. My doctoral research, at the University of New Hampshire, is focused on analyzing the perceptions of how the third grade test has influenced teachers, the curriculum, and instruction.

Your school has been randomly selected as one of approximately 100 schools that will be asked to participate in this study. I will be sending a survey to all third grade teachers at each of the selected schools. The survey will ask teachers to respond to questions on curriculum, instruction, and their attitudes relating to the test. A pilot test of the instrument revealed that most participants completed the survey in less than 15 minutes. The results of this research will be available to all interested parties on my web site this spring.

The purpose of this letter is:

1. To inform you of this research and ask that you encourage your teachers to participate in this project. In order to ensure the reliability of this research it is important that a high percentage of invitees participate. Teachers should expect to be contacted within the next three weeks.

2. To ask that you complete the enclosed postcard with the names of all third grade teachers at your school. The New Hampshire Department of Education does not record this information so your response is critical to the success of this project. You can either mail the card to me or send the names via e-mail.

I know that this is a busy time for you and I thank you for the time that you will devote to this activity. I sincerely believe that this research will provide interested parties with useful information regarding the improvement of education in our state. If you have any questions about this research project, feel free to contact me by phone or e-mail.

Please return the enclosed postcard or send an e-mail to me as soon as possible. Your assistance in this matter is greatly appreciated. Thank you!

Sincerely,

Steve Kossakoski

Research Web Site: http://www.nhultranet.com/~koss
School: ____________________________

Please print the names of all third grade classroom teachers below.

1. ____________________________________________________________
2. ____________________________________________________________
3. ____________________________________________________________
4. ____________________________________________________________
5. ____________________________________________________________
6. ____________________________________________________________
7. ____________________________________________________________
8. ____________________________________________________________
9. ____________________________________________________________
10. ____________________________________________________________

Please return this postcard ASAP. Thank you!

TO: Steve Kossakoski
P.O. Box 481
Kingston, NH 03848
Dear New Hampshire Educator:

We, the executive directors of the New Hampshire School Administrators Association, the New Hampshire Association of School Principals, the New Hampshire Affiliate of the National Education Association, and the president of the New Hampshire Federation of Teachers, urge you to participate in the teacher survey on the New Hampshire third grade assessments conducted by Stephen Kossakoski.

As you know, since May of 1994, third grade students have been tested in the areas of English, Language Arts, and Mathematics as part of the New Hampshire Educational Improvement and Assessment Program (NHEIAP). While many changes have been implemented by individual school districts as a result of this program, little research has been conducted on the manner in which teachers are coping with the challenges created by the implementation of the NHEIAP. This study will address this issue and will inform policymakers, educators, community members, and other parties interested in educational assessment issues.

Again, we hope that you will participate in this endeavor, and thank you for your time and consideration.

Sincerely,

Mark V. Joyce/Exec. Dir.
NH School Administrators Association

Ms. Peggy McAllister, Exec. Dir.
NH Association of School Principals

Mr. Mel Myler, Exec. Dir.
NEA-NH

Mr. Jim Rust, President
NH Federation of Teachers (AFT)
October 27, 1999

Dear NH Third Grade Teacher:

I am a doctoral student in the Ph.D. program in Education at the University of New Hampshire and I am seeking your assistance in my dissertation study of the New Hampshire Educational Improvement and Assessment Program (NHEIAP). As you are well aware, the NHEIAP has had a tremendous impact on all schools in the state. This study will provide teachers with a vehicle for voicing their professional opinions about the NHEIAP third grade assessment.

The third grade teachers in your school have been randomly selected to participate in this study. I will be sending a survey to you within one week. The survey will ask that you respond to questions on curriculum, instruction, and your attitudes relating to the test. A pilot test of the instrument revealed that most participants completed the survey in less than 15 minutes.

In order for this study to accurately reflect the thoughts and views of New Hampshire's third grade teachers, I am urgently asking for your participation in the study by promptly returning the forthcoming survey.

Thank you in advance for your time and cooperation!

Sincerely,

Steve Kossakoski

Steve Kossakoski
POB 481
Kingston, NH 03848

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
October 30, 1999

Dear «Prefix» «LName»:

I am a doctoral student in the Ph.D. program in Education at the University of New Hampshire and I am seeking your assistance in my dissertation study of the New Hampshire Educational Improvement and Assessment Program (NHEIAP) grade three assessments. As you are well aware, the NHEIAP grade three assessments have had a tremendous impact on all schools in the state. However, while these assessments were first administered over five years ago, little research has been conducted on their impact at the classroom level. This study is focused on analyzing how the third grade assessments have influenced teachers, the curriculum, and instruction.

You have been selected as one of over 300 teachers representing 103 randomly selected schools in New Hampshire that have been asked to participate in this study. The enclosed survey should take approximately 15 minutes of your time to complete. In order to ensure that this study accurately reflects the views of third grade teachers in New Hampshire it is important that you complete and return the enclosed survey. The results of this research will be available to all interested parties on my web site this spring.

Please be assured that your responses and comments will be held in strict confidence. The information gathered from your survey will be compiled with the data received from other teachers. No teacher, school, or school district will be identified as part of this research. The code number on your survey will only be used to verify the return of the completed survey and to prevent you from receiving reminder letters.

I know that this is a busy time of the school year for you and I thank you for participating in this project. I sincerely believe that this research will provide educators, policymakers, parents, and communities with important information regarding the improvement of education in New Hampshire. If you have any questions about this research project, feel free to contact me by phone or e-mail. Please complete and return the survey in the enclosed pre-addressed stamped envelope by Monday, November 15, 1999.

Thank you for your time and cooperation!

Sincerely,

Steve Kossakoski

Research Web Site: http://www.nh.ultranet.com/~koss
APPENDIX G
November 16, 1999

Dear NH Third Grade Teacher:

Two weeks ago you should have received a survey and a letter asking for your participation in my dissertation study on the NHEIAP third grade test. If you have already returned your survey, please accept my sincere appreciation. If you have not returned the survey, please know that your participation is important in order for this research to accurately represent the opinions of all third grade teachers. Please complete and return your survey on, or before, Tuesday, November 30, 1999.

If you did not receive a survey or if it was misplaced, feel free to contact me by e-mail (mail@mail) or telephone (555-555-5555) and I will send you another promptly.

Thank you in advance for your time and cooperation!

Sincerely,

Steve Kossakoski
POB 481
Kingston, NH 03848
December 2, 1999

Dear «Pfx» «LName»:

I am writing to you once again to ask that you participate in my dissertation research on the NHEIAP third grade assessment. A survey was mailed to you at the end of October, however, to date I have not received your completed survey. I am enclosing another copy of the survey along with a return envelope. In order for this research to be representative of the opinions of grade three teachers in New Hampshire it is very important that I receive a high percentage of survey returns. Your opinions regarding the NHEIAP third grade test are very important to the success of this study.

If you have already returned your survey, I sincerely appreciate your assistance in this matter — please excuse this reminder. If you have not responded, I hope that you will please take 15 minutes to complete the survey and return it to me within the next few days. Please be assured that your responses and comments will be held in strict confidence. The information gathered from your survey will be compiled with the data received from other teachers. No teacher, school, or school district will be identified as part of this research. The code number on your survey will only be used to verify the return of the completed survey.

NEW TEACHERS: If you are teaching 3rd grade for the first time and do not feel that you can complete the survey, please note this fact at the top of the survey and return it to me as soon as possible.

I know that this is a busy time of the school year for you and I thank you for participating in this project. I sincerely believe that this research will provide educators, policymakers, parents, and communities with important information regarding the improvement of education in New Hampshire. If you have any questions about this research project, feel free to contact me by phone or e-mail. Please complete and return the survey in the enclosed pre-addressed stamped envelope by Monday, December 13, 1999.

Thank you for your time and cooperation!

Sincerely,

Steve Kossakoski

Research Web Site: http://www.nh.ultranet.com/~koss
Curriculum and Instruction Category Factor Analysis

<table>
<thead>
<tr>
<th>Subscale items</th>
<th>Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shift in tested curriculum instructional time</td>
<td></td>
</tr>
<tr>
<td>Q1 Reading</td>
<td>.05</td>
</tr>
<tr>
<td>Q2 Writing</td>
<td>.05</td>
</tr>
<tr>
<td>Q3 Mathematics</td>
<td>.17</td>
</tr>
<tr>
<td>Shift in non-tested curriculum instructional time</td>
<td></td>
</tr>
<tr>
<td>Q4 Social Studies</td>
<td>.05</td>
</tr>
<tr>
<td>Q5 Science</td>
<td>.07</td>
</tr>
<tr>
<td>Q6 Art</td>
<td>-.05</td>
</tr>
<tr>
<td>Q7 Music</td>
<td>-.07</td>
</tr>
<tr>
<td>Q8 Physical Education</td>
<td>-.05</td>
</tr>
<tr>
<td>Changes in school emphasis in English language arts</td>
<td></td>
</tr>
<tr>
<td>Q9 Reading and literature</td>
<td>.23</td>
</tr>
<tr>
<td>Q10 Listening/viewing</td>
<td>.37</td>
</tr>
<tr>
<td>Q11 Writing</td>
<td>.22</td>
</tr>
<tr>
<td>Changes in school emphasis in mathematics</td>
<td></td>
</tr>
<tr>
<td>Q12 Problem solving</td>
<td>.45</td>
</tr>
<tr>
<td>Q13 Addition and subtraction</td>
<td>.49</td>
</tr>
<tr>
<td>Q14 Fractions</td>
<td>.77</td>
</tr>
<tr>
<td>Q15 Add/subtract decimals</td>
<td>.71</td>
</tr>
<tr>
<td>Q16 Estimation</td>
<td>.67</td>
</tr>
<tr>
<td>Q17 Using charts and graphs</td>
<td>.66</td>
</tr>
<tr>
<td>Changes in practice in English language arts</td>
<td></td>
</tr>
<tr>
<td>Q18 Identify/compare shapes</td>
<td>.72</td>
</tr>
<tr>
<td>Q19 Measurement</td>
<td>.83</td>
</tr>
<tr>
<td>Q20 Patterns</td>
<td>.74</td>
</tr>
<tr>
<td>Q21 Constructed response</td>
<td>.02</td>
</tr>
<tr>
<td>Q22 Prepare students for test</td>
<td>.03</td>
</tr>
<tr>
<td>Q23 Multiple choice</td>
<td>.08</td>
</tr>
<tr>
<td>Q24 Small group learning</td>
<td>.19</td>
</tr>
</tbody>
</table>

Note. Loadings over .4 are shown in bold face type.
<table>
<thead>
<tr>
<th>Factors</th>
<th>Subscale items</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q25 Innovative strategies</td>
<td>.21</td>
<td>.78</td>
<td>.07</td>
<td>-.02</td>
<td>.15</td>
<td>.00</td>
<td>.10</td>
<td>.06</td>
<td>.14</td>
<td>-.16</td>
<td>.13</td>
</tr>
<tr>
<td></td>
<td>Q26 School or grade-level planning</td>
<td>.08</td>
<td>.82</td>
<td>-.02</td>
<td>.31</td>
<td>.14</td>
<td>-.18</td>
<td>.00</td>
<td>-.03</td>
<td>.00</td>
<td>.04</td>
<td>-.24</td>
</tr>
<tr>
<td></td>
<td>Q27 Project work</td>
<td>.07</td>
<td>.74</td>
<td>.01</td>
<td>.03</td>
<td>.06</td>
<td>-.04</td>
<td>.28</td>
<td>.04</td>
<td>.09</td>
<td>-.03</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td>Q28 Drill to build/review skills</td>
<td>-.04</td>
<td>.15</td>
<td>.03</td>
<td>.00</td>
<td>.10</td>
<td>-.07</td>
<td>-.05</td>
<td>.86</td>
<td>.05</td>
<td>.13</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>Higher-order thinking</td>
<td>.11</td>
<td>.16</td>
<td>.07</td>
<td>.22</td>
<td>.50</td>
<td>.39</td>
<td>.04</td>
<td>.23</td>
<td>.04</td>
<td>-.38</td>
<td>.16</td>
</tr>
<tr>
<td></td>
<td>Q29 activities</td>
<td>.07</td>
<td>.38</td>
<td>.18</td>
<td>-.08</td>
<td>.11</td>
<td>-.14</td>
<td>.73</td>
<td>-.03</td>
<td>.21</td>
<td>-.07</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>Q30 Student choice of topics</td>
<td>.02</td>
<td>.02</td>
<td>-.04</td>
<td>.12</td>
<td>.21</td>
<td>.68</td>
<td>-.07</td>
<td>-.13</td>
<td>.04</td>
<td>.18</td>
<td>.13</td>
</tr>
<tr>
<td>SS6</td>
<td>Changes in practice in mathematics</td>
<td>.15</td>
<td>-.03</td>
<td>.13</td>
<td>.11</td>
<td>.69</td>
<td>.28</td>
<td>.11</td>
<td>.03</td>
<td>-.18</td>
<td>.27</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>Q31 Constructed response</td>
<td>.13</td>
<td>-.06</td>
<td>-.03</td>
<td>.03</td>
<td>.07</td>
<td>.15</td>
<td>.11</td>
<td>.19</td>
<td>.01</td>
<td>.86</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>Q32 Preparing for test</td>
<td>.23</td>
<td>.57</td>
<td>.04</td>
<td>-.08</td>
<td>-.24</td>
<td>.27</td>
<td>.23</td>
<td>.05</td>
<td>.09</td>
<td>.08</td>
<td>.42</td>
</tr>
<tr>
<td></td>
<td>Q33 Multiple choice</td>
<td>.34</td>
<td>.70</td>
<td>.00</td>
<td>-.08</td>
<td>-.05</td>
<td>.26</td>
<td>.10</td>
<td>.01</td>
<td>-.04</td>
<td>-.07</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td>Q34 Small group learning</td>
<td>.25</td>
<td>.74</td>
<td>-.06</td>
<td>.25</td>
<td>.06</td>
<td>.09</td>
<td>.08</td>
<td>-.04</td>
<td>-.02</td>
<td>.14</td>
<td>-.29</td>
</tr>
<tr>
<td></td>
<td>Q35 Innovative strategies</td>
<td>.17</td>
<td>.51</td>
<td>-.01</td>
<td>.01</td>
<td>-.13</td>
<td>.27</td>
<td>.58</td>
<td>.08</td>
<td>.06</td>
<td>.11</td>
<td>-.07</td>
</tr>
<tr>
<td></td>
<td>Q36 School/grade-level planning</td>
<td>.13</td>
<td>-.13</td>
<td>-.04</td>
<td>.13</td>
<td>-.01</td>
<td>.06</td>
<td>.02</td>
<td>.83</td>
<td>-.07</td>
<td>.26</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>Q37 Project work</td>
<td>.21</td>
<td>.09</td>
<td>.00</td>
<td>.11</td>
<td>.25</td>
<td>.74</td>
<td>.08</td>
<td>.13</td>
<td>-.11</td>
<td>-.16</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>Q38 Drill to build/review skills</td>
<td>.19</td>
<td>.24</td>
<td>.11</td>
<td>-.08</td>
<td>-.05</td>
<td>.03</td>
<td>.84</td>
<td>-.03</td>
<td>.15</td>
<td>.09</td>
<td>-.03</td>
</tr>
</tbody>
</table>

Percentage of variance: 12.46 11.00 6.76 6.35 5.95 5.57 5.22 5.03 4.89 4.82 4.40

Note. Loadings over .4 are shown in bold face type.
Pressure and Attitude Category Factor Analysis

<table>
<thead>
<tr>
<th>Subscale items</th>
<th>Factors</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS7 Pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q41 Myself</td>
<td></td>
<td>.16</td>
<td>-.03</td>
<td>.28</td>
<td>.65</td>
</tr>
<tr>
<td>Q42 Other teachers</td>
<td></td>
<td>.13</td>
<td>.06</td>
<td>-.10</td>
<td>.75</td>
</tr>
<tr>
<td>Q43 Principal</td>
<td></td>
<td>.43</td>
<td>-.03</td>
<td>-.03</td>
<td>.68</td>
</tr>
<tr>
<td>Q44 District Administrators</td>
<td></td>
<td>.58</td>
<td>-.05</td>
<td>.11</td>
<td>.59</td>
</tr>
<tr>
<td>Q45 New Hampshire DOE</td>
<td></td>
<td>.82</td>
<td>-.11</td>
<td>-.19</td>
<td>.13</td>
</tr>
<tr>
<td>Q46 State Gov't. Officials</td>
<td></td>
<td>.86</td>
<td>-.11</td>
<td>-.14</td>
<td>.04</td>
</tr>
<tr>
<td>Q47 Parents</td>
<td></td>
<td>.61</td>
<td>.03</td>
<td>.11</td>
<td>.36</td>
</tr>
<tr>
<td>Q48 Community</td>
<td></td>
<td>.74</td>
<td>-.05</td>
<td>.08</td>
<td>.33</td>
</tr>
<tr>
<td>Q49 Media</td>
<td></td>
<td>.74</td>
<td>-.16</td>
<td>.15</td>
<td>.12</td>
</tr>
<tr>
<td>SS8 Test reflects school quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q50 Results reflect individual student achievement</td>
<td></td>
<td>-.04</td>
<td>.85</td>
<td>.06</td>
<td>.03</td>
</tr>
<tr>
<td>Q51 Results reflect overall student achievement</td>
<td></td>
<td>.02</td>
<td>.78</td>
<td>.22</td>
<td>-.04</td>
</tr>
<tr>
<td>Q52 Results reflect differences among schools</td>
<td></td>
<td>-.11</td>
<td>.77</td>
<td>.05</td>
<td>-.02</td>
</tr>
<tr>
<td>Q53 Results reflect the quality of instruction</td>
<td></td>
<td>-.13</td>
<td>.78</td>
<td>.09</td>
<td>.06</td>
</tr>
<tr>
<td>SS9 Attitude toward NHEIAP test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q56 Results influenced by changes in instructional practices</td>
<td></td>
<td>.18</td>
<td>-.01</td>
<td>.59</td>
<td>.23</td>
</tr>
<tr>
<td>Q57 Results influenced by alignment with state frameworks</td>
<td></td>
<td>.01</td>
<td>.00</td>
<td>.69</td>
<td>.19</td>
</tr>
<tr>
<td>Q58 Provides feedback on how well I am teaching</td>
<td></td>
<td>-.21</td>
<td>.45</td>
<td>.47</td>
<td>.10</td>
</tr>
<tr>
<td>Q59 Produced positive instructional change</td>
<td></td>
<td>-.09</td>
<td>.41</td>
<td>.72</td>
<td>-.11</td>
</tr>
<tr>
<td>Q60 Produced positive curricular change</td>
<td></td>
<td>.00</td>
<td>.29</td>
<td>.74</td>
<td>-.20</td>
</tr>
<tr>
<td>Q61 I support the NHEIAP test</td>
<td></td>
<td>-.11</td>
<td>.61</td>
<td>.52</td>
<td>-.20</td>
</tr>
</tbody>
</table>

Percentage of variance

|         | 18.78 | 17.96 | 13.98 | 11.84 |

Note. Loadings over .4 are shown in bold face type.
Teacher Comments

I do believe that testing can provide important data that shows progress over time for some. However, as I stood aside and watched some of the children shed tears as they were trying to take the test, I began to question if some parts of the test were DEVELOPMENTALLY appropriate for a 9 year old child. I feel there are certainly some parts that set the children up for failure. How valid then are the results? I feel the written response to a prompt is scored higher for length than for content. It is difficult and frustrating as a teacher because I can see where the children have come from. If I have a student who comes to me scoring ones and is scoring threes by testing time, isn't that progress? If so, is sure doesn't show when testing scores are released. Regardless, I am proud of them all!

I participate in the L.A. NHEIAP committee and value the encouragement given to teachers on the committee.

I believe there are questions in the test that should be “thrown out” but are still included (or styles of questions) that set the students up for failure. I believe that more areas of “modifications” should be allowed.

It's still a reading test. It separates the average readers from the critical readers.

Every May I tell myself I've taught the skills and enriched the curriculum, but if the students have a bad day at home or on the bus coming to school, all that “baggage" travels with them for the day. Also, no one takes into consideration the borderline MR children who are barely functioning at grade level. The NHEIAP test is only important to them if it is important to their parents! The NHEIAP test does not accurately profile many students in my class.

I think that if the test were used for the purpose of aligning curriculum, it's done its job. It feels that results are not necessarily used in a constructive way—perceived competition between districts, political ramifications, etc.

Thoughts, questions, concerns are as follows:
1. I am concerned that 3rd grade is the level chosen for testing. Kids don't seem to be developmentally ready for the higher level thinking being asked of them. Many 3rd graders are just beginning to put reading skills together.
2. Our hands are tied by our administration and school board to do what we think would work. There is no readiness, no retention policy, no ability grouping. Kids who are not at grade level or who aren’t meeting the proficiencies are given “modified" programs so that they are successful (not successful according to grade level). There is no way that all of these modifications are preparing children for the test or to able to score at advanced or proficient levels.
3. Fluctuation of what is emphasized depending on previous year's test results.
4. Math test is language based. Unfair to children unable to read.
5. Test has done the job of aligning and improving curriculum.
6. Teacher and student creativity has been lost.
7. Too much breadth, not enough depth. Should be doing few things well!

The implementation of the NHEIAP has caused curriculum to become more uniform in our school and in NH schools. I feel I am doing a better job of teaching writing and problem solving than I did in the past. However, I truly question much of the impact on the children I teach in third grade. Our school district puts a lot of pressure on teachers to get their students to perform. I personally feel a lot of pressure. I’m sure that the 8 year olds I teach feel it also. Our day is packed because of the curriculum which must be taught. Much of the problem solving and higher level thinking is much above the understanding of your average 8 year old. I feel we are rushing children to grow up and they will be no better off academically in a few years. Children need time to absorb what they are learning. I feel that this test is too much pressure for young children.

I am concerned and frustrated with the Special Education students who are required to take this test. These students are tested every three years and have an IEP developed to meet their needs. The third grade test is written on a much higher level than most of those students are able to read. There may be some accommodations for certain sections and certain parts; however, does the group of people who developed this test actually physically watch these IEP students take the test? Many of these students are very frustrated since the test is not at a level of their own abilities; therefore, a lot of guessing takes place. Why have IEPs developed or accommodations made? I had an IEP student last year take the test
and on the Language Arts section he got 7 out of 38 questions and 11 out of 32 on the math correct. Did that tell me anything that I did not already know about him? NO—he's go and IEP—that's what I knew and I knew that he was unable to read the test. So what did that prove to anyone having him take the test???? What about the districts that still have a Readiness program? Some students are not developmentally ready. I have some students in my classroom this year that just turned 8 in September. Some students don't test well (I know I don't). Over the past 6 years that our district has given the test we have tried to adjust our curriculum. All good and well however with an aligned curriculum to meet the standards we need to stop adjusting and start teaching. Our district is very adamant about doing well (although we "aren't" suppose to be comparing ourselves to other districts or schools within our district) and has our third graders take a practice test as well as work an 11 week math packet program based around a year's previous test. We as teachers have expressed our displeasure with doing it this way, making our teaching very choppy, and want to TEACH the students as the professionals we were hired into this district.

The NHEIAP test has done its job to improve and align curriculum. Now it's time to allow teachers the time to teach. The following are just many thoughts I have regarding this testing:

1. Teachers' hands are tied regarding curriculum instruction. We are school board directed. We teach what we are directed to teach.
2. The test results have become a competitive tool between towns.
3. There is fluctuation of what is emphasized for us to teach based upon previous years test items and scores.
4. These test results have caused more meetings, workshops on test taking skills; mini tests for students, etc. for ways of improving test scores.
5. The primary school experiences have been left out and the focus on children's needs has shifted. We no longer have time to see that children enjoy learning. We are stuffing information down them and asking them to complete tasks that they may not be developmentally ready to complete. One example is our reading program now has numerous open response writing activities that administrators have developed four point rubric guides for. No longer do we read for enjoyment.
6. Our district has set aside a two-week time slot in the early part of the New Year for a practice NHEIAP test. Using common items from previous years, students are put through the rigor of taking the test. These items are then scored and results tallied for areas that we as teachers are to stress before the actual testing in May.

Is this test valid? Is it necessary to put third graders through such a stressful experience? Have parents been given the opportunities to see this test? Are the test results a true evaluation of schools and teachers? Is the child's background that may include a list of family issues considered? When we first administered the NHEIAP we were all overwhelmed with the format. We immediately revised our curriculum. We had a Reading Specialist come in and teach the children how to approach a Writing Prompt. We revised our math curriculum and two years ago purchased a new Math Textbook, Harcourt and Brace which aligned more with the NH Frameworks. School Board unfortunately put a lot of emphasis on our scores and were very negative. This trickled down to our administration who held numerous meetings with third grade teachers trying to figure out how to improve our scores. Much to the dismay of the teachers one method they chose was having children take numerous practice tests. We as teachers felt this took time away from teaching curriculum. One year our school came in 3rd in the state. Needless to say we were patted on the back. I'm surprised there wasn't a parade. This particular class happened to be outstanding, and was the last class to have had Readiness. I joined the State Math Assessment Committee to find out more about the Test. I figured if I was going to complain about the test I should find out more about it. It was very enlightening. One good thing about being on the committee was becoming more aware of the Math Standards and proficiencies. Another good thing was understanding questions are chosen at random. We used to think there were more open-ended questions on Graphs and that we needed to concentrate on graphs etc. When creating the test we try to make sure that most of the proficiencies are addressed, but there is not a pattern or special emphasis. This was good to know. When we teach we need to cover all the material and not focus on certain areas as we had done in the past. There are positive things about the test. It has forced teachers to become more aware of the State Frameworks. The negative aspect is that too much emphasis has been put on the test by not only School Board, but administrators. I also feel the media has done an injustice to many schools. It is important that people understand that demographics and economics affect the scores. Whether or not a school has Readiness or retention is also important. I have ten summer birthdays this year in my class. This is certainly going to affect my scores if they are not developmentally ready for some of the concepts I am addressing in third grade. So much time being spent preparing for a test has prevented many teachers from spending time on creative projects they used to enjoy doing. I have tried to keep art and music and my curriculum because I feel creativity is so important. Writing is more than a writing prompt. I feel our school has lost sight of that. Everything is about open-ended questions or prompts.
I also feel that the test is only a piece of the assessment puzzle. Some of my best students are not good test-takers. Some of my best math students have a great deal of difficulty reading. Therefore the math test is not a good assessment of what they can achieve.

We are always trying to learn new and innovative ways of teaching. Not that I feel we have aligned our curriculum with the frameworks I would like to relax a little about the NHEIAP. I don't want to live and breathe the Assessment. I don't want our district to be judged on where we place in the state. I don't want to be compared to other schools in our district, who by the way have Readiness and Retention. It's like comparing apples and oranges. I wish someone would write an article and explain why some schools have a better change at scoring higher. Lots of factors have to be taken into consideration: class size, learning disabled population taking the test, the number of children getting support, the number of 3rd grade classes in the school......need I go on. Sorry if I rambled. Hopefully I having given you some information that will be beneficial to you. We really appreciate the time you are spending on this survey.

I think the NHEIAP has been the single most important thing that I have been involved in during the last 15 years of my professional career. It is significant because it has changed my practices and those of my fellow teachers. We have been forced to look at children in a new light. We did a lot of building kids up with false confidences. We had some real low performers because we didn't push hard. Instead we kept telling them their work was great (praise to motivate). We don't do that anymore. It's so cut and dry here. In one hand we hold the standards in the other, the indicators of progress. We report honestly about each child, give feedback to each student, and encourage all students to work hard and get smart.

The money spent on the NHEIAP test could be put to better use by providing more money to the districts so they in turn could provide more enrichment opportunities, additional up to date classroom materials and teacher's aids. Teachers and schools are only a small part of the educational picture. Parents need to be more supportive of schools and be more willing to assist their children when they struggle. Above all else they need to offer their children encouragement and a "Yes, you can do it--just try." The NHEIAP test is just one test. There are other ways to assess student achievement. District scores go up and down, this is often because of group abilities --- some groups are stronger academically then others. This is not taken into account.

Our school did best on the first test in comparison with others because there was not time to prepare and we were already doing things that aligned with state frameworks. Each year the bar is raised. Comparing one year to the other cannot be justified. It is a new group of third graders learning the same material; but the questions insinuate that since they could do it last year, it should not be difficult this year. I have many slower, "non-readers," "can't add" students this year. If the bar is raised another notch, these students will just stare a the paper. I can bring them up to doing well on one of the first state test levels.

I believe teachers feel a great deal of pressure to teach to the test. We are always trying to cram additional lessons in --especially in math. Some units are not taught until May or June.

Our district had problem solving, higher order thinking skills, and cooperative group learning in place prior to these tests being first administered.

It bothers me that this test is administered to students that have a coded reading or math disability. Though this is accounted for on the test, it is still averaged into the count. It concerns me to see students frustrated and crying because they know how they will score--novice! On the other hand, I think it really separates those "very bright" students from your "average" student. At the lower primary grades many parents believe their child is "gifted" and it reinforces the explanation "gifted." It really demands higher levels of thinking.

The test has improved our Math instruction. We are closer to the frameworks and doing more problem solving. I feel intense pressure as a grade 3 teacher and it does not trickle down enough to Grade 1 and 2. We are the ones who get the results and worry about the testing. I find the writing sample contrived and limiting. The prompt is not always developmentally appropriate. The students need time to prewrite and reflect before writing. My students write beautifully in class before the test, but the do not always perform as well on the writing test. The reading test is fine -- I think the passages are a bit boring sometimes, but overall it works. The test is just one snapshot -- I do not agree with educators who use the results too heavily. Third graders are very young to do a test such as this. I think it should be at 4th grade.

I feel more and more we are teaching to the test. The amount of time spent by all the 3rd grade teachers in practice tests is ridiculous. I feel especially dismayed by the writing prompts -- personally I couldn't write an interesting coherent piece on those topics. Over and over I've seen excellent writers produce mediocre work on the test. I also feel that some of the math (i.e., decimals) is not appropriate for 3rd
graders. The amount of time it would take for understanding is not worth it. And being both a 3rd and 4th grade teacher I feel the end of 4th grade would be a much more appropriate time to give a test of this magnitude. The pressure of preparing for this test makes me seriously consider not looping and just teaching 4th grade.

It is a poor indication of true math ability for students who are below average readers and not coded. Part of the test should be strictly computation to give the previously mentioned group a chance to be successful. Last year, when attending a meeting about the NHEIAP test in Nashua, I mentioned my previously stated comments to the presenter (on a one + one) in charge of the math section of the test. I found him to be not very pleasant or understanding of my point. He kept saying that they wanted the test to be unlike any other. Well, it is so much that it doesn't give some children a chance to be somewhat successful. On the other hand, much of the math work is geared beyond the third grade level — what a disappointment. Have the people who made the math section actually been in a classroom?

I've taught a 2/3 multi-age, 2nd grade and, this year, a 3/4 multi-age. I did not administer the test my first year teaching—my 3rd graders went to another classroom for the test. I feel SOME pressure to help my students perform well on the third grade test this year—however, no more than I felt as a 2nd grade teacher. I always keep in mind that pencil/paper test measure only the performance of a student for one moment I their educational experience. I spent my summer scoring the MCAS for Advanced Systems in Dover. The experience left me feeling very negative towards standardized assessments and how they are scored. I think if more people realized how these tests are scored we could focus on the big business this creates for some companies and put our energies into educating our children effectively.

—The writing sample question and procedure don't reflect the writing process which is emphasized in our district—prewriting discussion, brainstorming, etc.
—The math section actually measures reading in many questions
—The math section has a disproportional emphasis on geometry which is often covered at the end of grade 3—after the test
—I question scoring procedures and qualifications of those who score. Why was a totally blank test book which had been marked void scored? The dismal results were included in our school's scores, which of course, appeared in the newspaper.

While the tests have changed instruction in hopes of improving test results, they don't reflect the attitudes and maturity levels individual students have that are taking the test in any given year. This could impact scores.

1-I think this test does not accurately reflect student achievement due to high test anxiety and the fear of doing poorly. Some students do not test well and tend to do worse than they would on a classroom exam.
2-The writing prompts have been very abstract (especially 1998-1999's test) and students who normally enjoy writing and produce well freeze up.
3-Most classrooms "teach to the test" so parents, school board, etc. will see good results.

Although I first welcomed the NHEIAP test as giving us new direction in instruction, I quickly learned that we were required to teach to the test even if our students were not developmentally prepared to succeed. I spent two years w/a below average population trying to get them to complete 5 paragraph essays and trying to develop ways of getting fractions and decimals across to them when what they needed was basic skills. It was a very frustrating ordeal. Then—the test itself is so anxiety and stress producing that many students who could do better succumb to a case of nerves. Some years the physical conditions in the classroom worked against the students. My room is always too hot and in the spring it can be quite uncomfortable. I have seen many a good student test poorly due to these extraneous conditions. The loss of class teaching time required to administer the test that time of the year is not worth what we get in results.

The test requires a great deal of maturity and focus which many third graders are not ready for. It expects a level of development in reasoning, inferences, and problem solving which are not developmentally appropriate for the average third grader. They give up by the end of the week. There are children who have fun with the test, but for many more it is a grueling experience. Parents just laugh when they see the sample passages or problems. Our school has a strong curriculum and our community has high expectations. This test has just put more pressure into a packed schedule.

The test should be administered at the fourth grade level.

I don't believe this type of testing is developmentally appropriate for third grade students. I believe we are
wasting valuable instructional time teaching students how to take a standardized test. Unfortunately, this pre-teaching is necessary for these young children to have some skill in test taking. I believe we are beginning to move away from an emphasis on the process of learning to an emphasis on product learning. Through these tests we are saying that all 8 or 9 year olds should know the same "stuff" at the same time. I especially resent the news media and politicians using the test scores to berate school districts and their teachers. These test scores should only be used as one measure of how instruction can be improved.

The NHEIAP test doesn't allow for children with different learning styles to perform well. It's designed for children with strong language and overall maturity. If districts want to keep results to themselves, interpret the test properly and intelligently decide what the results mean; then the test could have some value. This is not what the results are used for and therefore any possible benefit of the test is negated by improper interpretation, incompetent understanding of public education on the part of the state and public display of test results of 8 year old children.

I was under the impression curricular change (at the state level) occurred prior to testing. It has moved local districts toward developing common curricula. I am most disturbed by the political involvement in education. At the same time I believe education needs to improve programs and increase accountability. The "test" does not do this. I strongly believe this test could be better. It has some excellent components and is connected to curriculum. However, its structure as well as its length are two serious drawbacks.

I feel that this assessment should focus on the skills learned by 3rd grade students from K through 3rd grade, and not singly on what they have learned in grade 3. Although I realize that this is the focus, I believe that more emphasis needs to be put on teaching all educators, K-3, about what their role is in the third grade assessment. We need to work together and teach strategies the same way and place emphasis on the same skills that need work in the district! I also think that educators deserve NHEIAP results mid-summer so as to better prepare for the following school year. Doesn't this make sense if improvement is our goal?

I feel there should be an additional component to the test regarding the children's individual potentials. For example last year my students tested lower than the year before however it does not reflect on the test score totals that I had several (1/3 of my class) with emotional and academic needs including 3 Sp. Ed. Students, 3 students in title one reading and an emotionally needy student. Another problem is in the Math section. When we test for Math understanding and problem solving what gets in the way of this assessment is the child's poor reading skills. I had 2 students who were great in Math but did poorly on the math test because the reading was too difficult.

The test asks thought provoking questions which is good. The test asks some questions which involve multiple steps. These can be very tough.

I realize that this type of assessment tool is necessary to reveal whether students are knowledgeable, flexible in thinking, understanding and capable of understanding complex ideas. However the type of students who enter school today are less prepared to accept the responsibilities to listen, learn and think independently. Therefore preparing students for this type of assessment requiring independent thinking and responding is a daunting challenge.

The test puts a tremendous amount of pressure, anxiety and stress on the 8 and 9 year olds who take it. It's an awfully young age to be doing that to. I personally like the math part of the test and think the L.A. is too difficult for this grade level. It doesn't follow the developmental curve and destroys those students just barely learning how to read.

The tests have resulted in more consistency in curriculum throughout the state. This is probably a good thing. The actual questions, however, are capricious and at times totally inappropriate for students of this age. We have had problems with the quality of the videos used. I personally feel there is not enough effort made to avoid gender biased questions (Video about big machinery that is inherently more interesting to boys as one example.) Some schools unfairly teach to the test (with prior knowledge of writing prompts or video topics) making comparisons between schools meaningless. Thanks for asking.

I have had concerns that the math part could create "math fear." Some students are not ready for this rigorous math—higher level thinking. There are so many variables involved in how a student scores that this test doesn't take into account. There is still not a "level playing field." Some inclusion/sp. needs students are excluded and some not. The test itself is "too clever" and "tricky" for children. Some parts are vague or confusing. The length is difficult for many students at this age, especially when they care to do their best and are stressed and exhausted. Some parts are good and some are developmentally
inappropriate—we feel the pressure to "teach" and "reteach" material that we know is not developmentally correct for students. In another year the students would be ready to learn some material easily as compared to being forced to learn it earlier. Do you think a 2-3 split grade can do as well as a straight 3rd? We've now lost readiness and don't have kindergarten in the public school. We're supposed to teach "developmentally," yet the test is just the opposite. Teachers are forced to do what administration dictates, but when not successful it is always the teacher's fault. He/she did a poor job. Personally, my students have done very well. I'm proud of the good job that I do, but I am also feeling worked to death! We teachers can never get enough done, and yet we are always given more to do, either by the local administration, state, or community pressure. The one only needs to turn the TV on or read a newspaper, to learn again about the poor job we are doing. Enough said—poor me! I hope some insightful info. is shared with the educational field and public.

I don't not think that many of the requirements on this test are developmentally appropriate for third graders. We teach the writing process and then ask the same children to give an organized, edited, final draft.

I think it is a form of child abuse. This test is very inappropriate.

My background is in Early Childhood and I strive to provide a learning environment that is developmentally and individually appropriate. Essential elements include social interaction, active interaction with manipulatives and oral communication. Children in first and second grade especially require this kind of environment. The 3rd grade testing situation is artificial! Training students to become familiar with the format of test taking needs to be introduced earlier than 3rd grade, in small increments. Increased emphasis must be placed on this training in 3rd grade in order to ensure familiarity and a more relaxed comfort level for students. The results do not accurately represent a student's ability or achievement level. The overall curriculum planning and implementation with emphasis on content. Teaching methods, learning environment, social/emotional development of students are components that affect learning. The frameworks provide a focus for curriculum renewal. The 3rd grade test does not provide an accurate measure of student achievement or curriculum implementation! The overall test results do not reflect the various factors that come into play during this one-time testing situation. I believe that the districts who score the highest have spent a great deal of time training their students to take the test. If feel that the frameworks do help districts focus on curriculum implementation and teaching practices --- not training students for this test.

Well, thank you, but I hardly know where to begin. I appreciate the opportunity to offer feedback, yet this is, for me, an enormous issue. The whole idea that a "test" can improve instruction or curriculum is ludicrous; the assessment and much of the state frameworks is in opposition to the entire last decade of research on learning and brain function; the "accommodations" allowed for the learning disabled and/or children with unorthodox processing abilities are laughable at best and discriminatory at least. N.H. is taking steps backward, not forward into the next century of education.

I believe the third grade test is better for fourth grade students. The test and groups of students change quite a bit each year so I feel comparing year to year scores is like comparing apples to oranges. One example, was last year's writing sample, it differed greatly from past years' prompts. The test is also supposed to be used to help each individual school improve their curriculum but it ends up pitting towns against each other or townspeople refusing teacher raises because of the scores. They need more classroom teachers designing the test.

The third grade test is not developmentally appropriate for an average third grader and is therefore a poor test. It requires an adult degree of awareness and reflective thought that goes beyond what's reasonable to expect for most third graders, except those who've been well schooled in what the test is like, and I don't feel we should be wasting classroom time teaching children to run like chats through this smug adult maze called the NHEIAP! The goals of creating reflective readers, writers, and students who can use the tools of mathematics towards practical ends are worthy, but this test has the effect of demoralizing teaching staffs, parents, and communities. Why not have each school offer documentation of what they've done for students at all levels to push them along toward these goals which could be reviewed and responded to by a state board with a visitation to verify that what they said is true?

I do not like having students do a writing sample in one sitting. If it must be done that way, I feel the writing prompt should not be about people. I also think there is too much fraction to decimal, decimal to fraction.

I feel that I've changed the way I teach so that my instructional material and techniques are more in line with the states framework. I use the results as a tool for my continued growth as a teacher.

155
Last year, I had a large percentage of special needs students. They had a difficult time with the test, and their scores reflected this. It's hard not to take their scores personally when many students score novice or basic. I'm always looking for growth and disappointed when it doesn't occur.

This test tends to build a status level within the school system. Great teachers with classes containing "difficult" students (ED-ADHD-PDD-no parental support, etc./low IQ levels) cannot get results that lesser teachers with high IQ students with parental support get.

Positive: teachers have taken a close look at their instruction techniques, curriculum has been worked on to coordinate skills K-6 (elem.). Negative: scores being publicly compared school/school district/district, less time being spent on the arts—it's difficult to justify art projects, plays, etc., tests being administered differently (no adherence to directions)

I do not like the fact that we are unable to help ESL and coded students during the reading/language arts section of the test. We give them support all year long, and we burst their bubble during test time. All it does is make them feel bad about themselves. Most are done in 5 minutes—just plugging in guesses. Students who do not take the test should not be counted as a 0 in figuring your classroom scores. It does not give an accurate picture of your class. I teach the children to use all their tools when writing, including Quickword, and then they are not allowed to use it for the test. This is not fair to them as well.

I strongly feel that my students on I.E.P.'s are at a disadvantage during the test, and they often feel inadequate. It tends to reinforce their feelings that they are not successful. I feel the test is not truly geared for the maturational stages of most 3rd graders. It does isolate the top percent, but it also frustrates many learners. The emphasis on fractions/decimals is unrealistic. Now we find many schools' curriculums "driven" by the test. The pressure is intense. Every year we write prescription plans to raise student success.

I feel it is not beneficial to children of special needs to be "forced" to take this regular ed test. Also, teaching to the test is to prove what?—Done in all districts. Children who make remarkable progress in a year's time are not congratulated, but shown to be still below their peers when given results. I feel all this testing is just another way for administrators to compare and compete. Thanks for letting me "air" my feelings. Good luck to you in your endeavors. What is your feeling?

I would like to see some research about the connection of parent income and education and how well the child does. It seems within our own district, schools with higher parent income/education/involvement have more students at basic and above. Also, some of the material (1999 Writing Prompt for third grade, narration of the listening/viewing video) were not appropriate for third grade.

There are times I feel the test is more a measurement of intelligence, but since intelligence correlates with academic success I guess that is OK. It is just frustrating how we focus on cooperative learning and discussing our problems all through 3rd grade and suddenly students must achieve independently—They are frustrated. Socio-economic level and maturity seems to have everything to do with the individuals' success with the test. Still—raising the bar helps everyone I know my institution has improved but the test is only one reason.

I believe students feel pressure from their parents, the school staff and themselves to do well on this test. (My own child thought she would "fail" the grade if she didn't do well.) I feel this effects the way they perform. Many time I have seen students who are "my best" writers "freeze" because this is a test and write an essay well below their usual quality on the essay portion of the English Language test. Further, I don't believe this test reflects the quality of instruction as much as it reflects the great differences in opportunity, encouragement and genetics our students bring to school. I can be the world's best jockey but I can't win the Kentucky Derby on a donkey. I can be the world's best teacher but I can't get proficient grades from a student whose mother did drugs while pregnant.

I strongly feel the children should be given a choice between two different writing prompts. Many times the children dislike the prompt and "shut down." Perhaps a choice between topics would help. Another suggestion would be to place the writing prompt after the Lang. Arts Reading sections. Often the first testing day is difficult due to the fact that this is for many their first formal testing situation that they have ever faced. Thank you!

Teaching to the test puts too much pressure on teachers and students. Children no longer have a well-rounded education, only facts on the test.

This test is not an accurate assessment of students who take it. (At least at the 3rd grade level). The readability is an average 5th grade level and places undue stress on 8 year olds. Third grade is the first year these students have ever seen a "content type test," never mind one of this proportion! Many
questions are designed to "trick" students, and many more are above their conceptual level. Good points? Hell, they learn how to take a test (after practicing those skills), and perhaps write a bit better (although the prompts are not appropriate!) An adult would have to put real effort into an essay about why! Let's replace this test with one that is an average 3rd grade readability, one with interesting prompts to write to, and one that isn't designed to trick students who are just beginning to take tests! Thanks!

I am very concerned over the usage of school scores—we were guaranteed by the Dept. of Ed. That scores would never be released/ranked by school. As you know, however, that's exactly what they've done.

While I do not have a problem administering the test, I feel that the state/city/governing bodies are caught up in the numbers. In the beginning, Dept. of Educ. Promised us that the results would be used for feedback and curriculum development ONLY. This past year all the schools were ranked publicly in the newspaper (something they assured us would never happen!!!)

I. The math assessment is very reading oriented. I have had very good math students perform at a "basic" level because they could not read all that was expected of them. Conversely, I have had "B" math students perform at an advanced level (which I never saw in day-to-day learning) because they were excellent readers and could figure out what they were being asked to perform.

II. Will the "raising of the bar" ever stay/slow down? I feel that I no sooner figure out the expectations and work with my kids to head in that direction and the test makers raise the ante. There's a part of me that doesn't feel we can keep the numbers low in NOVICE range when the expectations keep rising. Last year's test had 6 (?) or more open-ended math questions all in 1 section—one after another—many of my students just quit on it because they got overwhelmed—could these have been spread out and not clumped together? Lastly, I take a great deal of pride in teaching and doing a good job—I want my kids to do well—I just wish the tests were a little less demanding/more realistic... These kids are only 8 and 9 years old!!! The expectations are just INCREDIBLE/HIGH. Are there any 3rd grade teachers setting up these assessments or just grown-ups trying to see if they can think like 3rd graders??? I thank you for giving some 3rd grade teachers the opportunity to tell you how we feel...

Although I generally support the concept of raising our educational expectation for all students, I feel deeply cognizant of the pressure for my students to perform very well on the assessment. The frameworks have positively improved curriculum expectations, however, it seems the expectation for all students is to perform at the proficient and advanced levels. I don't feel this is a realistic expectation for many kids. I am annually frustrated by the testing administration. The questions are often extremely difficult to the students and not really assess what they know. Test designers do not understand how an 8 year old thinks and the manner in which many test questions are designed is not developmentally appropriate. I feel I spend far too much time teaching test taking skills in order to prepare my students for the assessment. We are also required annually to come up with lengthy action plans to address areas in the testing that need improvement. I am appreciative of the fact that the test does allow for certain accommodations for individual student needs. I find the writing sample expectations particularly invalid. It seems that quantity is valued over quality. I would rather see an 8 year old write a well-planned, concise paragraph than a rambling multi-paragraph offering. I find that I have had to abandon many worthwhile projects and activities that I've done for years because I simply cannot afford to take the time to do them. There is so much to teach and prepare the children for. I think it is a shame that many of the experiences have had to be eliminated from my teaching. I know the children enjoyed and benefited from them. Thank you for asking for the opinions of myself and other classroom teachers.

Teaching toward meeting the State Standards is useful. However, we seem to be more concerned with teaching toward the test that "proves we are meeting the standards." What will the emphasis be this year? The writing prompts differ widely in their inspiration to a child's writing. Fractions are either emphasized or not. What must be crammed in before the second week in May? The publication of the ranking of schools in the state is a violation of what teachers felt was a promise this would not happen. How is this helpful?

This is my first year teaching third grade and administering the test. The pressure I put on myself was immense—I really wanted to have the children "practice," talk, test all year so they wouldn't feel overwhelmed. When it came time they responded very well. After they reflected it "wasn't as hard" as they thought it might be. Our class results seemed to be an accurate reflection of their skills. It was a very diverse class=extreme highs/lows. We'll see how this year goes.

This test should NOT be given in 3rd grade. Third graders are NOT developmentally ready to do such a test. Fourth grade would be far more appropriate. This is also a lot of undue stress put on a 3rd grader.
I have been teaching 3rd grade for 10 years. I feel this test is developmentally not appropriate for third graders. It is hard and it seems to get more difficult each year.

I feel this test is made to be complicated and tricky. For concrete 8 and 9 year olds, the design is confusing and frustrating. A more straightforward design of questions and data can still discern knowledge. All too often there are questions in multiple choice format that do not offer a correct answer—or offered two correct answers. Remaking the test each year adds to the problem. The videos have, for the most part, been less than adequate. And, the language should really be on a 3rd grade level.

I feel that it is a little too difficult. Take math for example. Most of the questions deal with a student's ability to reason and apply. That is excellent. However, there should also be a section on basic computation (as there is on the CAT's exam). All students cannot deal on a higher plane and it is good feedback for the teacher to know if at least the student is able to basically compute—a requirement of many jobs in life.

This same vein, whereas all math questions involve some amount of reading, one doesn't know if the student did not get it right because he couldn't do it or he couldn't read it. Though modifications are allowed on the math test for very low readers, there are poor readers who do not qualify for that and have to read the questions themselves, so one doesn't know if the problem is a lack of understanding in math or problem with the reading. I would to see a small section on pure computation.

I believe the Mathematics section has become increasingly difficult in the open-ended response section...some problems excessively difficult for 3rd graders (inappropriate)

Last year's writing prompt was not a good prompt to assess a third grader's writing. The prompt was difficult for many to respond to; therefore, how could this be a just sample to critique on individual's writing?

I question the reading level of the test!!! I think it is written well beyond a third grade reading level. I could understand if some of the test was on level and some above. However I think the entire test is written at too difficult a reading level to test other skills well.

I'm quite concerned over several parts of this test. The math, at times, seems "brutal" in what I truly think are trick questions for a third grader. I also believe that there should be some third grade computation facts to give some sense of success on the test if this is where a student's math strength is. Another concern I have is the developmentally young students we seem to be getting more of in the last 5 or 6 years. My school has a pretty strong philosophy of no retention. They seem crushed. I'm also concerned that the very first day is the writing piece. I'd like to see that after the reading portion when they're used to the testing situation and may be more relaxed. There is much instruction being given to "teach to the test" rather than just quality, well-rounded instruction for "where the child is."

The test should be one of many tools by which the student's growth is measured. One group of third graders does not reflect on the upcoming class. The profile of each class is too individual to be a "stand alone" indicator of students/school/teacher achievement.

I have felt all along that the NHEIAP's are nothing but the State Department of Education's attempt to "test" how well the teachers are teaching to their curriculum guidelines. They have caused much undue stress, which thanks to the State Dept. of Ed.'s position of power is only supported and magnified by the media. Furthermore, it is a very difficult test for third graders with test questions that are often poorly worded, misleading or just plain confusing. If a test of this nature is necessary I do not feel it should be administered until grade 4. In the meantime, I will continue to teach my class in the most meaningful ways I know how. Each class is very different as are their learning styles, therefore each year I change my instructional style to meet the needs of my students and I will not worry about the NHEIAPs.

I have been serving on the state committee to make up the third grade math test. After seeing what goes into it, I have a lot more respect for the test and the process. It is made up by my teachers and tries to accurately reflect the state frameworks. However, I feel that some parts of the state frameworks in Math, specifically addition of fractions and decimals, are beyond the developmental level of third graders and should be changed. I also have concerns about third graders taking the test — it would be more effective in fourth grade.

50 & 51 — This has to be answered as two parts. I think the MATH questions are for the most part carefully thought out and would get a 4. I think the LANGUAGE ARTS & LISTENING sections are poorly done and my response would be a 1. The WRITING section would get a 3. It can be argued both ways.

60. The change is positive if you agree with the standards. If you think you should be teaching the WHOLE child and not at TEST-TAKER, the answer would be a 1.
Our school is very small and usually there is only one class of third graders. When percentages are reported in the newspaper our school can look like it did very well or very poor and it is really only one or two students. I really don't think have the scores plastered in the Sunday paper is helpful. What is helpful is disaggregating the data and looking at what was weak, seeing if there is a hole in the instruction and trying to improve. Improvement can only happen if all teachers K-3 are involved in the plan. The test is not just a test of third grade. It is important to remember the other grades are part of this too.

This is my third year at this school. I know that the school has increased the emphasis of language arts instruction since I have been there. I'm not sure if instructional emphasis in math has changed since the first administration of the NHEIAP. We do focus more on writing across the curriculum.

Reported results (media) don't reflect special needs/chapter 1 students which can impact school

Because of NHEIAP, we (staff) have spent hours and hours trying to create curriculum that will meet the NH standards—realistically, I would need ~25 more school days to meet them. Many of the standards involve higher level thinking abilities and are moved to the 3rd grade curriculum. Math and social studies have added a large amount of time needed to "prepare" the students. Lengthening the school day and the school year is the only way I feel I will be able to teach everything.

I feel NHEIAP is taking over our lives. Just let me teach, I promise they're learning!!

I know that every school is different, but I personally feel that my instruction hasn't really changed as a result of the testing. In contrast, it seems that the tests are more reflective of our teaching content, as is. Something that is not addressed in this survey much is the preparation needed for the students. Multiple choice test strategies need to be taught and practice tests from years past need to be used to increase student comfort level. Third graders need to be taught these test taking skills—this needed practice take away valuable teaching time at the end of April. So many skills are on the math test and with 5-6 weeks left of school after the test, there is still more to teach—so there's the question of doing a cursory job of all skills for the test or continuing with in depth study of the math strands and then not getting into everything. Does this then put our kids in a disadvantage because teaching to the test does not seem right in this instance? A lot of weight is put on these scores. Some of the test questions are terrible! There seems to be a lack of consistency with the type/amount of preparation amongst classrooms and schools. So much depends on the type of stories, writing prompt, etc.... I just teach the curriculum the ways my students learn best; present the whole test taking days as "a different type of activity" similar to their practice ones and encourage them to do their best. If pressure is taken off the kids, maybe they'll do better. Most enjoy it, occasionally there are tears. I've kind of rambled on—no one has really asked my opinion about this and as you can see, I have mixed feelings. Some accountability is needed. Good luck with your pursuits!

Interestingly, I am currently participating in a seminar at UNH on Curriculum Design. We have spent considerable time dissecting national/state/local frameworks and I feel I have a solid knowledge base in terms of the reasons behind the implementation of the NHEIAP, which is more politically than philosophically anchored. Although the "test" does not reflect my own basis for how children are most effectively and developmentally assessed, I take the test with a grain of salt, use the results objectively, understand that its but a single piece of how children are assessed, and one that does not drive instruction for me. Rather, I continue to try to develop a community of learners that centers on the individual possibilities for each of my students. Teaching to any test is not a paradigm shift I would make. Media hype however, does affect me because the information is arbitrary in regards to socio-economic basis of communities. "Gene banks" count in this assessment!!! I'm saddened that so many districts have succumbed to pressures to improve scores and so many teachers feel compromised. Alfie Kohn's The Schools We Deserve deconstructs this standards, benchmark, testing mania quite nicely.

Thank you for this survey. So far you are the only person to ask my opinion of NHEIAP test. I think the week of testing is pure hell. I am a strong believer in Howard Gardner's Multiple Intelligence. These children (in my opinion) do not get tested fairly. Some students are the best story tellers, however, writing skills may be weak. What does the test measure? Also from teaching MI, some students do not sit for long periods of time and during testing that is the only thing the students can do. Lastly, there is so much talk about this test all year! What pressure! Teaching should be for children to learn not a competition among towns. Again, thank you for asking my opinion.

159
Over the past 4 years my district (lower socio-economic) has spent incredible hours looking over results, targeting weakest areas, and altering instruction. I have served on every committee to help improve grade 3 scores. We continue to score low. The increase is so slight. The bulk score novice/low basic. We, as teachers see growth and feel good about the instruction our students receive. Writing appears to be our weakest area. We have serious concerns about the scoring of this portion of the test (length only matters—not quality). Also, upon hearing teachers from other districts, the question arises to the level of support some students receive during the test. We follow the protocol—not everyone does!

I have noticed inconsistencies in scoring of the writing prompts. (2 pieces received scores of 2, and were miles apart in quality and quantity). Longer pieces were scored higher than shorter pieces, even if shorter piece was of better quality. It appears that understanding and answering the prompt is not important. A child wrote a piece totally off topic, and still scored well. Different versions of the test (A, B, C, etc.) appear to be more difficult than others.

Our school invested in Sahon Math last year. It was felt the students needed consistency in the math program. I feel Sahon lacks in many areas that Math Their Way does not. We are now asked to “read” the scripted lessons and there is no time in the day to supplement with hands-on activities, cooperative group activities, or problem solving. We have completed a 4 year process for a writing curriculum which I believe will show better test scores in the future. We also teach students good test-taking strategies in all areas throughout the year. We question the scoring on the writing piece. It seemed very “subjective” rather than using a rubric to obtain a more standard score. “see attached (copies of scored tests attached)

Score=2
I would like to be is Leesu my lif wood be fun
Score=2
If I could be a famous person, I would like to be Will Smith. I would like to be Will Smith because I would spend my life as a rap singer. I would want to be Will Smith also because I would have awesome moves. Score=2
If I could be a famous person I would be the President. This is why I would be the President because I would get to ride in a limo. There would be people to clean up after me. I could do what ever I wanted to. If I wanted to go some were someone would have to take be there. I could just it easy. Life would be relaxing. That's what I would be.

The test is good in that it provides structure for my instruction. It shows a measurable growth between 3rd and 6th grade. Parts of it in the past have been developmentally challenging for 3rd grade students. Some of the things the test results are being used for are questionable. They are many variables that need to be taken into account—motivation, the physical state of the student at the time and the seriousness of the student(a 3rd grader!) and the length of sitting that are not consistent with everyday effort and output of that student.

I have a hard time believing that some 3rd grades do so much better that others. I can't help but think there is some dishonesty going on, by the classroom teachers. This caused by the pressure to "rate high" on the test. Possible solution: Have the tests monitored by surrogate "teachers."

Due to the small amount of test items for each skill, scores are easily skewed by misunderstanding or misreading items. Also, students processing skills are not looked at in the written language essay. Rough drafts, revising and editing are only viewed through the final product....Strange!

I feel the test is difficult for the average child. Some students get very nervous during the test and the results are not accurate.

It is not that I disagree with testing children but these tests, at least some of the questions, seem to be tricky. The wording of many questions is not how children think. It will take a number of years for teachers and students alike to change our way of thinking.

It is unfortunate that schools, parents, SAU offices and communities in general are using/abusing/misusing the results of NHEIAP. The entire focus on 3rd grade instruction is on the results. We have even been instructed that if a skill is not on the test don't spend any extra time on it. Sad but true. The pressure on the children is too much. Why not just teach the test and let everything else go by the wayside?

I think the level of sophistication of the third grade test is unreasonable. Are they testing reading (l. arts) and math competency or how clever students are on "trick" questions. We don't always teach 8 year olds
the way they test them.

A waste of my student's time and my instructional time. More time is spent teaching testing strategies than curriculum materials. Testing lacks norms, so the "Stanford" is also given. A huge block of prime instructional time is devoted to administering these tests, for the purpose of "pitting" one school against another and one community against another. A complete waste of "tax" payer moneys.

52-Socio-economic factors as well as % of sped or learning disabled students are a factor here.
56-Teach to the test--any test--and % goes up.
58-For this test, yes. I've taught for 30 years. This test is only one small indicator of "how well I am teaching."
60-My personal change, but the district hasn't really changed--unfortunately.

It is difficult to compare scores form one year to the next because there is no base line -- also can't compare nationally. New superintendent is back to basics drill review to learn facts (understanding not a strong point) we'll see...

My recent decision to think about another job is somewhat influenced by instructional mandates I have received partly due to the NHEIAP. My district/admin. has declared its support of this test and openly teaches to the test. At the annual district meeting, test scores are thrown about in support of various viewpoints. I am sure the time will come when schools are compared openly in our regional (5 towns) district. After that, teachers, too, will be tracked. This is already done, although on the Q.T. At my school, with a teacher request by parents a sure thing, a disturbing trend has developed. Children of notable community parents are being placed in classes (by request) whose "atmosphere" is traditional and whose classes, do quite well on the test. This leaves children who are not championed by their parents to be placed in the remaining spots. These children could generally be described as less successful at school (trad.) activities. (the painters, wigglers, loud mouths etc.) In summary, children whose parents value behavior and activities that commonly result in better school success are packing in classes whose atmospheres reflects the same. Less popular (not necessarily poor in atmosphere--just not traditional) classrooms are being filled with the other children. (Each are becoming condensed versions of themselves.) I wonder if survival of the fittest is going to result in classes of privileged (by nature OR environment) children taught by teachers who are NHEIAP driven (by nature or environment). Yes, I fear for my job, not to mention the state of education.

The only fault I find with the test is the time factor. I do not think it's realistic to expect a third grader to read for an hour plus at one sitting. I would like to see the reading broken down into small segments.

As a teacher who uses collaborative learning opportunities and integrated curricular approaches I feel undue pressure to cover more content than to make student connections and meaning making. Deep understanding takes time and multiple opportunities to experience the same concept in more than one place. The state test has many good points, however it should be considered only a piece of the picture not the whole pie as it is unfortunately seen as. It does raise the bar to aim for, but unfortunately not all students are interested in the need to perform well.

Interestingly enough, the first year the assessment test was given to third graders was the year I returned to the teaching profession after a ten year absence. I was amazed to see this test! My first impression was that it was a test to seek out "gifted and talented" students as it certainly didn't fit what my third graders were able to do. After administering it for six years, I can say that I do know what is covered on the test, and yes, I do try to prepare my students for this experience. However, the wide range of materials and concepts tested is more than I can teach in depth to most third graders. I feel like I am skimming and hurrying through some of the basics in math to at least introduce more advanced topics so my brighter students might be able to successfully respond to those questions. I am glad that the media seems to have "backed off" on its approach in reporting the test results. It's now so much a feeding frenzy; town vs. town; school vs. school; teacher vs. teacher. We do use the results to beef up our teaching strategies and approaches, BUT we also realize that even with the best teaching and materials, some third graders will not score any higher than novice. Interestingly, I don't believe any student in my district has ever been fully excused from taking this test, and we have had some very challenged students. I teach in a town where the property taxes are high; parental involvement minimal, and Title I and SPED, overwhelmed. I live in an affluent town. Which community's scores do you think are better? In concluding, I am pleased to say that no student has ever "freaked out" during testing. However, I am a basket case during that week as well as the last week in October when scores are released. If I didn't enjoy teaching kids at this age, I probably would request another grade just to avoid this test!!

Generally, I feel the test reflects where my students are except for the writing prompt. I have had some very strong writers "bomb" on this section simply because the prompt didn't inspire them, or because
they usually need some "brainstorming" to get started. Enough said.

I feel too much emphasis is placed on these tests. Children can't enjoy a fun curriculum anymore, teachers can't teach things they would like to teach because of always being asked how does it fit in the frameworks? Will it be on the test that the students take? Some students are terrible test takers and always will be so they will never perform well on the tests in any grade. People also forget that K, 1st and 2nd grade teachers also play a big part on how well students perform on these tests. On half of a year in 3rd grade does not teach students everything they'll need to know on these tests!!

I enjoy sharing the common questions with the students. They are enjoyable reading selections and math problems. The NHEIAP is a fun, interesting test for children AND one that supports a most important element for our 21st century young people: THEIR THOUGHTS/OPINIONS/HYPOTHESIS. It also values strong communication skills, another important emphasis of the future. I enjoy the challenge it gently clues me into—it teaches me what's important in a manner that teachers enjoy: an experiential way rather than the lecture/presentation method of learning about the latest teaching practices.

Having a small rural school, I feel my students' results are unfairly judged by comparisons with other schools. When only ten students take the test and three are special education it reflects poorly on the whole class. I think the test needs to be more balanced.

I don't see a great degree or correlation between the students' classroom performance and the test scores. Some students in last year's class, for example, performed at a proficient level during the year but only scored at the basic level on the test. Some of this was due to lack of effort on the test. The greatest correlation is seen at the highest and lowest ends of the spectrum. I see a difficulty in giving the test in early May on material to be taught in Grade 3. It forces the teacher to try to introduce as much as possible prior to the test or to be testing the children on material they have not had. I feel that this test would be better given in Oct. of Grade 4. The students would have completed Grade 3 material and would have had time to adjust to 4th grade. The teachers could have enough time to review key concepts before the test is given.

With so many different forms, I don't see how the NHEIAP can fairly or accurately measure a child's ability. I have had children do what I thought was poorly on a certain (form) question or part of the test when I thought they would have done much better had they had a different (form) question or part. Also, the tests do not take into account children with learning disabilities. Each group or class of children is different too. I would also like to see this test given at the beginning of 4th grade rather than May of 3rd grade. My reasoning here is that especially in math we have not had a chance to cover all the concepts needed to do well on the test.

Children below grade level in reading will NEVER do well on this test and it saddens me. It does not show how much growth they have made it only shows what they don't know. We need more "positives" and fewer "negatives" and this is a BIG NEGATIVE. It is a waste of valuable instructional time.

Many of the questions and problems presented in the third grade test are developmentally inappropriate to the age of child. I am tired of politicians and business determining what children should know and when.

The NHEIAP test is often perceived as the "ultimate" indicator of student progress. This of course, is not the case. The newspaper and other media further complicates matters or promotes that perception by making comparisons between schools within a district and comparing school districts. Parents often see these published scores as an indicator of their school's worth.

I think there are other factors that affect the NHEIAP scores. Time of year, heat, the way they are administered being different from the way we teach. Over the years the feedback has been useful to our district and we have realigned all curricula areas. I've sat on all committees so I am very well informed, my issue is with people who choose to do nothing to improve their instructional practices. This isn't fair to the children nor to the other children whose test scores are compiled as a group.

It isn't so much that we have shifted available instruction time—as we have put much more emphasis on writing answers—ie—explaining math in written form. This is extremely difficult for those borderline kids—can do the computation but are low in language areas—but not SPED students (ie. Chapter 1 kids). The number of multi-part open response questions in the math is also exhausting. If that many are truly needed—let us vary the test schedule—mixing language and math. Variety carries energy. We don't teach language all day Monday and Tuesday and then math all day Wed. and Thursday. Thank you.
Whether or not accelerating the curriculum robs students of true mastery. I also wonder how class size, curriculum. I am always looking for new materials and better techniques for teaching. I continue to debate which the school board and administration attend to test results — has made me constantly reassess my

I feel many of the questions are deliberately set up to trick students rather than accurately test a skill. Some are poorly worded. I especially feel last year’s prompt contained poor wording. Many children did not understand what they were being asked to do. A much simpler, more easily understood prompt could have said: Pretend to be a famous person. Write about what your life would be like as that person.

While I do believe that standardized test have some worth in the realm of assessing our students, I find it sad that so much attention is brought to these tests. The fact of the matter is—some people don’t test well. Also, my students do not respond well to writing prompts. The things that I like about the NHEIAP is that it requires the need to bring staff together to talk about how we can be better at what we do. We should be doing this whether the NHEIAPs agree or not. I do not like when we become expected to teach to the test—something that I find boring. I know some teachers that spend much of the year with their students responding to writing prompts. I see that as learning that is not meaningful. It that what we want children to be engaged in?

This test indicates the child’s thought process very clearly. It holds the school and home accountable for bettering education and demanding educational responsibility to and for the child. The schools have gotten the message but parents haven’t.

My school never goes over the tests with the teachers in grades 1 and 2. There is tremendous pressure on Gr. 3 teachers to do well but they come in to us w/o even knowing how to write a sentence. We never analyze responses or look to ways of improving….we just have the third grade teachers slave away at improving them.

I think that the NHEIAP has been somewhat positive in guiding instruction. However, I think it is not right to use testing scores information to compare schools against each other because a lot of people do not understand all of the other variables among schools, such as socioeconomic level, number of special ed kids, teacher to student ratio, etc.

I believe that the NHEIAP test is influenced by many factors including: socioeconomic issues at home, transient students, services and test anxieties.

My last year’s class consisted of 20 students in September. Seven moved away before May! Of the remaining 13 original students I had: —3 who were new to the school that year (1 of the 3 students had gone to 5 schools by the end of 3rd g!) —7 had IEPs —7 who were in Chapter 1. From approx. March to May (before testing) I had 3 new students which all had IEPs. One of the new students was at kindergarten level where he copied his name from his name tag on his desk, and confused letters with numbers! This information does not get printed in the newspapers to explain the low scores.

I have very mixed feelings about the testing. I feel it both helps and hinders the teaching/learning process. It is difficult to cover the entire 3rd grade math and language arts curriculum six weeks prior to the end of the school year. In order to have children feel somewhat familiar with the material on the test I have to accelerate their lessons. This is particularly difficult for students with learning disabilities or who are developmentally "young." I can’t say that I feel good about how overwhelmed the students are by the testing. There were 4 children who cried last year because they didn’t understand parts of the test. Two were students who typically struggle with independent work, but two were students who usually do very well. These tears came despite the fact that I’d spent weeks teaching test taking skills, had explained that the testing was not a measure of how smart they were, but evaluated how our teaching and methods were working. I start each test morning with “breakfast.” Students and their families donated healthy cereals and bananas (I supplied the bowls, milk, and spoons). We started each morning with a bowl of cereal. This relaxes the class and addresses the problem of low energy—and attention for those children coming to school without having had breakfast. I have to admit that having to administer the test knowing the results for our school are published in the local paper (mine is the ONLY 3rd grade class); and seeing the level to which the school board and administration attend to test results — has made me constantly reassess my curriculum. I am always looking for new materials and better techniques for teaching. I continue to debate whether or not accelerating the curriculum robs students of true mastery. I also wonder how class size,
school funding, and teacher experience effect test results. Of further concern is the fact that each year the assessment is designed to evaluate slightly different skills. This constantly leaves me feeling anxious about my preparation. I have often thought that the test also put challenging problems at the beginning (particularly in math)—which tends to overwhelm students and sets an anxious tone for the rest of the test. I apologize if this seems scattered or if I've packed too much into my comments...it just seems good to have the chance to express my concerns...finally! Thank you!

I feel that too much time is spent practicing for the test. This takes away from other important things the students could be learning. We also take time to have whole school meetings after the results have been released. We work on a plan of action that doesn't seem to make a difference.

I understand the test has created some positive, much needed change but... it has also created a lot of teacher and student stress, teaching to the test and unfairly comparing schools. The greatest pressure I feel comes from our principal and assistant principal. Because were are located in an urban area, when newspapers publish the results, I feel they are "embarrassed" when other schools in the area do better. However, these schools are either located in higher socioeconomic neighborhoods or they are located in the inner city—a low socioeconomic area which qualifies that school for federal grants, programs, etc., so that students receive much more individual attention. They do not feel these differences between schools should matter in the test results (they've even said this to the third grade teachers). They will not admit that our "gene pool" is lower than some of the other neighborhood schools. They tell us "you're the hardest working third grade teachers in the state...BUT" since we're in the classroom we must do things differently—Great Pressure! Also, first and second grade teachers don't seem to take any ownership for the test: it's often referred to as "the 3rd grade tests" and NOT "End of the third year test." EXTREME PRESSURE for conditions I have NO control over.

Pressure comes from administration and our individual school has problems since we do not qualify for Title I we don't receive the tutors we need. There is great difficulty for finding EH tutors and even children who are coded do not receive the support that their IEPs call for.

The pressure comes from administrators—the difference between schools—socioeconomic and extra help for schools should be considered. Pressure from the newspaper publication does not promote a healthy atmosphere for learning. It is defeating to be constantly compared to other schools. I'm afraid we are really starting to teach to the test and this is crazy in third grade. I am all for a test, even a tough test. I know that it is good for students to strive for a high mark. Most of the questions are a nice challenge. Some of the questions however, are purposefully tricky. Why? It doesn't take a genius to "trick" an 8 year old. A difficult test is one thing, a tricky one is another. Amazing the way the media blows off the test when results are status quo or slightly higher! Watch out for any dip. Thanks for this opportunity.

I worry about my "young" 3rd graders—those students who are developmentally young and/or chronologically young, and whether the NHEIAP is too aggressive for them to deal with at "this time." I reluctantly embrace "upping the ante" in class—but the test itself is very long and very tough. I'm not one to throw young kids into the deep end of the pool.

With the implementation of the NHEIAP I seen more and more schools try to teach to the test. Much of the process is confusing to students, parents, and the community. With many teachers not up to date on current practices, the process approach to learning is not being used correctly. The degree of anxiety provoked by these tests is ridiculous. Socioeconomic conditions weigh heavily on results and all students are figured into final results. Deaf children, autistic, Down's syndrome, MR— they are all figured together whether they can read, write, or communicate. This is wrong.

Having to take this test forced the district to look at curricula and update text book series. The first year's results (and the test itself) was a big shock to administration in our district. I feel that the test is given at the wrong time. It may be better suited to 4th grade in Oct. or Nov. after all concepts have been introduced and reviewed. Because of snow days, etc., we have had as much as six weeks of school after the test. Many concepts at the end of the third grade are not even introduced by the second week of May. Many children are not developmentally ready for decimals, equivalent fractions in the third grade.

I do feel that some of the examples on the test are a bit "tricky" for the children (I wish I could remember one to tell you). Also most of the math that is on the test is 3rd grade math. When the test is given in early May we have not yet completed teaching all the concepts.
I don't believe parents should only take test scores to evaluate their child. Some students don't test well, others lack effort and get discouraged because of the amount of testing and the time of year given. I think parents know their child best and should consider their every day work habits to know their child's academic level.

Too many tests are currently being given to our third graders! In our reading program alone, we give 6 theme tests, and 3 benchmark tests! In math, we give district-wide beginning and ending tests for each year. Then, we have the NHEIAP. We will also be participating in the IOVA test this spring. Why???

I have felt the test is geared to a high level student with above grade level reading ability. Those students, average or lower, often get frustrated taking the test. Those students on IEPs should not be subjected to this test -- another format should be used. I feel I have changed my teaching style to best teach with the test in mind—but I am not sure this has made me a better teacher.

I think the NHEIAP test has some merit. I would however support each school district developing their own assessment that would match up with the N.H. frameworks. Children who struggle with reading or who are below grade level or coded cannot be successful in either the Lang. Arts or Math section of the NHEIAP. We need to come up with a better way to assess what children actually know.

The design of the test questions is an area I disagree with. Many questions have very confusing "wording" that often confuses the students. In order to show improvement in test results I have used many different phrases i.e., Blooms' Taxonomy—in tests I administer, or oral discussion questions in class—as a result of the NHEIAP test.

I would like to see a breakdown of reading levels for the test the children are expected to read, comprehend and make inferences about.
The Institutional Review Board for the Protection of Human Subjects in Research has reviewed the protocol for your project as Exempt as described in Federal Regulations 45 CFR 46. Subsection 46.101 (b) (2), category 2.

Approval is granted to conduct the project as described in your protocol. Changes in your protocol must be submitted to the IRB for review and approval prior to their implementation.

The protection of human subjects in your study is an ongoing process for which you hold primary responsibility. In receiving IRB approval for your protocol, you agree to conduct the project in accordance with the ethical principles and guidelines for the protection of human subjects in research, as described in the Belmont Report. The full text of the Belmont Report is available on the OSR information server at http://www.unh.edu/osr/compliance/belmont.html and by request from the Office of Sponsored Research.

There is no obligation for you to provide a report to the IRB upon project completion unless you experience any unusual or unanticipated results with regard to the participation of human subjects. Please report such events to this office promptly as they occur.

If you have questions or concerns about your project or this approval, please feel free to contact me directly at 862-2003. Please refer to the IRB # above in all correspondence related to this project. The IRB wishes you success with your research.

For the IRB,

Kara L. Eddy, MBA
Regulatory Compliance

c: File
Grant Cioffi