The New England Common Assessment Program

Guide to Using the 2005 NECAP Reports

Contact Information

If you have questions after reviewing this guide, please contact the Department of Education for your state.

New Hampshire Department of Education: Tim Kurtz, Director of Assessment, 603-271-3846, 101 Pleasant Street, Concord, NH 03301, www.ed.state.nh.us

Rhode Island Department of Elementary and Secondary Education: Mary Ann Snider, Director of Assessment and Accountability, 401-222-8492, 255 Westminster Street, Providence, RI 02903, www.ridoe.net

Vermont Department of Education: Michael Hock, Director of Assessment, 802-828-3115, 120 State Street, Montpelier, VT 05620, www.state.vt.us/educ/

Table of Contents

Introduction 3
NECAP Background 3
Document Purpose 3

General Guidelines for the Use of NECAP Reports 4
Alignment of Curriculum and the NECAP Tests 4
Use of NECAP Student-Level Results 4
Multiple Data Points Needed for Trend Analysis 4
Regulations Regarding Confidentiality of Student Records 4
NCME Code of Professional Responsibilities in Educational Measurement 4

Understanding the NECAP Student Report 5
Student's Achievement Level and Score 5
Student's Achievement Level Compared to Other Students by School, District, and State 5
Student's Performance in Content Area Subcategories 5
Sample *NECAP Student Report* 7

Understanding the Item Analysis Report 9
Sample *NECAP Item Analysis Report* 11

Understanding the School and District Results Reports 12
Overview 12
Sample *NECAP School Results Report* (cover page) 17
Grade Level Summary Report 18
Sample *NECAP School Results Report* (Grade Level Summary Report) 20
Content Area Results 21
Sample *NECAP School Results Report* (Reading Results) 22
Disaggregated Content Area Results 23
Sample *NECAP School Results Report* (Disaggregated Mathematics Results) 24

Understanding the School and District Summary Reports 25
Sample *NECAP School Summary Report* 26

Appendix A: Overview of Assessment Instruments and Procedures 27
Appendix B: Code of Professional Responsibilities in Educational Measurement 38
Appendix C: NECAP Achievement Level Descriptions 46

Introduction

NECAP Background

The New England Common Assessment Program (NECAP) is the result of collaboration among New Hampshire, Rhode Island and Vermont to build a set of assessments for grades 3 through 8 to meet the requirements of the No Child Left Behind Act (NCLB). The states decided to work together for three important reasons:

- Working together brings together a team of assessment and content specialists with experience and expertise greater than any individual state.
- Working together provides the capacity necessary for the three states to develop quality, customized assessments consistent with the overall goal of improving education.
- Working together allows the sharing of costs in the development of a customized assessment program of a quality that would not be feasible for any individual state.

Document Purpose

The primary purpose of this document is to support local educators' use of test data from the October 2005 administration of the New England Common Assessment Program (NECAP) tests. This document describes and explains the information included in the following NECAP reports:

- NECAP Tests of Fall 2005: *NECAP Student Report*
- NECAP Tests of Fall 2005: *NECAP Item Analysis Report*
- NECAP Tests of Fall 2005: *NECAP School/District Results Report*
- NECAP Tests of Fall 2005: *NECAP School/District Summary Report*

These reports contain information valuable to schools and districts in their efforts to better serve the academic needs of individual students and to evaluate and improve curriculum and instruction. In addition, this document can help school and district personnel communicate with their communities about the NECAP test results. It is important to note that these reports contain results from the student assessment program, and not individual state accountability systems.

Please note that the Appendices contain important information about NECAP assessment instruments and procedures.

General Guidelines for the Use of NECAP Reports

Alignment of Curriculum and the NECAP tests

All test items appearing on the NECAP tests are designed to measure specific NECAP *Grade Level Expectations*. As schools align their curriculum and instructional programs with these standards, test results should reflect student progress towards these standards.

Use of NECAP Student-Level Results

NECAP results are intended to evaluate how well students and schools are achieving the learning targets contained in the *Grade Level Expectations*. NECAP was designed primarily to provide detailed school-level results and accurate summary information about individual students. NECAP was not designed to provide, in isolation, detailed student-level diagnostic information for formulating individual instructional plans. However, NECAP results can be used, along with other measures, to identify students' strengths and weaknesses. NECAP is only one indicator of student performance and should not be used for referring students to special education or for making promotion and/or graduation decisions.

Multiple Data Points Needed for Trend Analysis

A single year's test results provide limited information about a school or district. As with any evaluation, school and district test results are most meaningful when compared with other indicators and when examined over several years for long-term trends in student performance. This is especially true in small schools where changes in student cohorts from year to year can have a noticeable influence on school results for any given year.

Regulations Regarding Confidentiality of Student Records

The Family Educational Rights and Privacy Act (FERPA) requires that access to individual student results, including those provided in the *NECAP Item Analysis Report*, and the *NECAP Student Report* be restricted to the student, the student's parents/guardians, and authorized school personnel. Superintendents and principals are responsible for maintaining the privacy and security of all student records. In accordance with this federal regulation, authorized school personnel shall have access to the records of students to whom they are providing services when such access is required in the performance of their official duties.

For more information about FERPA please visit the following website: http://www.ed.gov/policy/gen/guid/fpco/ferpa/index.html

National Council on Measurement in Education (NCME) Code of Professional Responsibilities in Educational Measurement

The Departments of Education in NH, RI and VT and Measured Progress adhere to the NCME code. Local educators also have responsibilities under this code. The entire document can be found in Appendix B. More information about NCME can be found at www.ncme.org.

Understanding the NECAP Student Report

The section below discusses the *NECAP Student Report*, which provides schools and parents/guardians with information about individual student performance. Details about the NECAP tests and achievement levels are provided on the cover of the *NECAP Student Report*. Details about the student's performance on the NECAP tests are included on the inside of the report, which is explained in detail below. Parents/guardians are encouraged to contact the student's school for more information on their child's overall achievement after reviewing the *NECAP Student Report.*

The *NECAP Student Report* is divided into three sections.

Student's Achievement Level and Score

This section of the report shows the achievement level attained for each content area. Achievement level descriptions can be found in Appendix C. It also shows the scaled score earned for each content area, as well as a score band that indicates the standard error of measurement surrounding each score. (See page 22 of this report for a definition of standard error of measurement). Achievement level descriptions are provided on the reverse side of the report.

Student's Achievement Level Compared to Other Students by School, District, and State

This section of the report lists the four achievement levels—Proficient with Distinction, Proficient, Partially Proficient, and Substantially Below Proficient—for each content area. This student's performance is noted with a check mark in the appropriate box. The percentage of students at each achievement level is listed for the student's school, district, and state.

Student's Performance in Content Area Subcategories

This section of the report shows the student's performance compared to school, district, and statewide performance in a variety of areas. Each of the three content areas assessed by NECAP is reported by subcategories. For **reading**, with the exception of Word ID/Vocabulary items, items are reported in two ways – Type of Text and Level of Comprehension. The two types of text are Literary and Informational. The two levels of comprehension are Initial Understanding and Analysis and Interpretation. Numbers and Operations, Geometry and Measurement, Functions and Algebra, and Data, Statistics, and Probability are the subcategories reported for **mathematics**. The content area subcategories for **writing** are reported on the Structures of Language and Writing Conventions, displayed in the student's writing and in response to multiple-choice items, and by the type of response – short or extended.

Student performance in all content area subcategories is presented as a table including possible points, points earned by this student, average points earned for the school, district, and state, and the average points earned by students at the Proficient level on the total content area test.

Students in grades 5 and 8 were also administered the NECAP Writing test, which was scored by two independent scorers. To give a more complete picture of this student's performance on the writing assessment, each scorer chose up to three comments from a predetermined list The comments selected by the student's scorers appear in the table at the bottom right-hand corner of the *NECAP Student Report*.

The *NECAP Student Report* is confidential and should be kept secure within the school and district. Remember, the Family Educational Rights and Privacy Act (FERPA) requires that access to individual student results be restricted to the student, the student's parents/guardians, and authorized school personnel.

The following pages contain a sample *NECAP Student Report*.

Student	Grade	School	District	State
	5			

Fall 2005 - Beginning of Grade 5 NECAP Test Results

Content Area	Achievement Level	Scaled Score	This Student's Achievement Level and Score (Below / Partial / Proficient / Distinction; 500 – 530 – 540 – 556 – 580)
Reading	Proficient	548	

Content Area	Achievement Level	Scaled Score	This Student's Achievement Level and Score (Below / Partial / Proficient / Distinction; 500 – 533 – 540 – 554 – 580)
Mathematics	Proficient	540	

Content Area	Achievement Level	Scaled Score	This Student's Achievement Level and Score (Below / Partial / Proficient / Distinction; 500 – 528 – 540 – 555 – 580)
Writing	Partially Proficient	535	

Interpretation of Graphic Display

The line (▮) represents the student's score. The bar (〰〰〰) surrounding the score represents the probable range of scores for the student if he or she were to be tested many times. This statistic is called the standard error of measurement. See the reverse side for the achievement level descriptions.

This Student's Achievement Level Compared to Other Beginning of Grade 5 Students by School, District, and State

	Reading				Mathematics				Writing			
	Student	School	District	State	Student	School	District	State	Student	School	District	State
Proficient with Distinction		5%	5%	15%		0%	0%	17%		5%	5%	10%
Proficient	✓	47%	47%	52%	✓	49%	49%	46%		40%	40%	41%
Partially Proficient		30%	30%	22%		28%	28%	19%	✓	40%	40%	33%
Substantially Below Proficient		19%	19%	11%		23%	23%	18%		16%	16%	15%

This Student's Performance in Content Area Subcategories

Reading		Possible Points	Student	Average Points Earned: School	District	State	Students at Proficient Level
Word ID/Vocabulary		10	9	5.8	5.8	6.5	4.7-7.3
Type of Text*	Literary	21	12	10.5	10.5	11.8	9.6-13.1
	Informational	21	12	9.5	9.5	11.2	7.5-11.7
Level of Comprehension*	Initial Understanding	26	12	11.9	11.9	14.2	10.5-14.9
	Analysis and Interpretation	16	12	8.2	8.2	8.8	6.7-9.9

Mathematics	Possible Points	Student	Average Points Earned: School	District	State	Students at Proficient Level
Numbers and Operations	30	11	11.7	11.7	14.9	9.2-14.5
Geometry and Measurement	13	4	4.5	4.5	5.9	3-6.5
Functions and Algebra	13	8	6.7	6.7	7.4	4.9-8.5
Data, Statistics, and Probability	10	4	3.7	3.7	4.6	2.2-5.1

Writing	Possible Points	Student	Average Points Earned: School	District	State	Students at Proficient Level
Structures of Language & Writing Conventions	10	9	8.5	8.5	8.7	8.1-10
Short Responses	13	3	4.9	4.9	5.4	4.2-6.4
Extended Response	13	9	8.4	8.4	8.4	6.5-10.8

Comments about this student's writing performance:

Writing has some organization.

Writing has some supporting details.

Writing has limited word choice and/or control of sentence structure.

*With the exception of Word ID/Vocabulary items, reading items are reported in two ways - Type of Text and Level of Comprehension

3/9/2006

NECAP Student Report - Fall 2005

This report contains results from the Fall 2005 Beginning of Grade New England Common Assessment Program (NECAP) tests. The NECAP tests are administered to students in New Hampshire, Rhode Island, and Vermont as part of each state's statewide assessment program. The NECAP tests are designed to measure student performance on grade level expectations (GLE) developed and adopted by the three states. Specifically, the tests are designed to measure the content and skills that students are expected to have as they begin the current enrolled grade. In other words, content and skills which students have learned through the end of the previous grade.

NECAP test results are used primarily for school improvement and accountability. Achievement level results are used in the state accountability system required under No Child Left Behind. More detailed school and district results are used by schools to help improve curriculum and instruction. Individual student results are used to support information gathered through classroom instruction and assessments. Contact the school for more information on this student's overall achievement.

Achievement Levels and Corresponding Score Ranges

Student performance on the NECAP tests is classified into one of four achievement levels describing students' level of proficiency on the content and skills required through the end of the previous grade. Performance at Proficient or Proficient with Distinction indicates that the student has a level of proficiency necessary to begin working successfully on current grade content and skills. Performance below Proficient suggests that additional instruction and student work may be needed on the previous grade content and skills as the student is introduced to new content and skills at the current grade. Refer to the Achievement Level Descriptions contained in this report for a more detailed description of the achievement levels.

There is a wide range of student proficiency within each achievement level. NECAP test results are also reported as scaled scores to provide additional information about the location of student performance within each achievement level. NECAP scores are reported as three-digit scores in which the first digit represents the grade level. The remaining digits range from 00 to 80. Scores of 40 and higher indicate a level of proficiency at or above the Proficient level. Scores below 40 indicate proficiency below the Proficient level. For example, scores of 340 at grade 3, 540 at grade 5, and 740 at grade 7 each indicate Proficient performance at each grade level.

Comparisons to Other Beginning of Grade Students

The tables in the middle section of the report provide the percentage of students performing at each achievement level in the student's school, district, and statewide. Note that one or two students can have a large impact on percentages in small schools and districts. Results are not reported for schools or districts with nine (9) or fewer students.

Performance in Content Area Subcategories

This section of the report provides information about student performance on sets of items measuring particular content and skills within each test. These results can provide a general idea of relative strengths and weaknesses in comparison to other students. However, results in this section are based on small numbers of test items and should be interpreted cautiously.

Students at Proficient Level

This column shows the average performance on these items of students who performed near the beginning of the Proficient achievement level on the overall test. Students whose performance in a category falls within the range shown performed similarly to those students. This comparison can provide some information about the level of performance needed to perform at the Proficient level.

Comments about this student's writing performance

Students in grades 5 and 8 took the NECAP writing test which included a writing prompt that required students to produce a written response up to three pages long. Student responses were scored independently by two scorers. Each scorer was able to choose up to three comments from a prepared list to provide feedback about each student's performance on the writing prompt. If both scorers selected the same comment, it is listed only once.

Achievement Level Descriptions

Proficient with Distinction (Level 4) - Students performing at this level demonstrate the prerequisite knowledge and skills needed to participate and excel in instructional activities aligned with the GLE at the current grade level. Errors made by these students are few and minor and do not reflect gaps in prerequisite knowledge and skills.

Proficient (Level 3) - Students performing at this level demonstrate minor gaps in the prerequisite knowledge and skills needed to participate and perform successfully in instructional activities aligned with the GLE at the current grade level. It is likely that any gaps in prerequisite knowledge and skills demonstrated by these students can be addressed during the course of typical classroom instruction.

Partially Proficient (Level 2) - Students performing at this level demonstrate gaps in prerequisite knowledge and skills needed to participate and perform successfully in instructional activities aligned with the GLE at the current grade level. Additional instructional support may be necessary for these students to meet grade level expectations.

Substantially Below Proficient (Level 1) - Students performing at this level demonstrate extensive and significant gaps in prerequisite knowledge and skills needed to participate and perform successfully in instructional activities aligned with the GLE at the current grade level. Additional instructional support is necessary for these students to meet grade level expectations.

Understanding the Item Analysis Report

The *NECAP Item Analysis Report* provides schools and districts with information on the released items. It also includes summary information on the scaled score and achievement level for each student in the school in reading, writing, and mathematics. In addition to showing raw data for students, it provides additional information for each released item. Using this report, together with the actual released items, one can easily identify test items on which groups of students did well or poorly. There is a separate *NECAP Item Analysis Report* for each content area. There is a legend after the last page of data for each content area that defines the terms used.

The data used for the *NECAP Item Analysis Report* are the results of the fall 2005 administration of the NECAP test. The NECAP tests are based on the *Grade Level Expectations* (GLE) from the prior year. For example, the Grade 7 NECAP test, administered in the fall of seventh grade, is based on the grade 6 GLEs. Therefore, many students receive the instruction they need for this fall test at a different school from where they are currently enrolled. The state Departments of Education determined that it would be valuable for both the school where the student tested and the school where the student received instruction to have access to information that can help improve curriculum. To achieve this goal, separate *NECAP Item Analysis Reports* have been created for the "testing" school and the "teaching" school. Every student who participated in the NECAP test will be represented in a "testing" school report, and most students will also be represented in a "teaching" school report. In some instances, such as when the student has recently moved into the state, it is not possible to provide information about a student in the "teaching" school report.

When reviewing the *NECAP Item Analysis Reports* it is important to note the following:

- enrollment and not tested data are not reported for the "teaching" school;
- not every student is represented in the "teaching" school reports; and
- the subtitle on the report indicates if the report is based on "teaching" or "testing" year. For example, on a grade 4 report, the subtitle "Grade 4 Students in 2005-2006" means that this report shows the item analysis for the school where the student was enrolled at the time of testing. The subtitle "Grade 3 Students in 2004-2005" indicates that this report shows the item analysis for the school where the student learned the grade 3 material he or she is tested on for the grade 4 NECAP.

The top portion of the *NECAP Item Analysis Report* contains seven rows of information.

- The first row lists the **released item number** (not the position of the item in the actual student test booklet).
- The second row lists the **content strand** for the item.
- The third row lists the **GLE code** for the item.
- The fourth row lists the **Depth of Knowledge** code for the item.
- The fifth row lists the **item type**.
- The sixth row lists the **correct response** letter for each multiple choice item.
- The final row lists the **total possible points** for each item.

When reviewing the multiple-choice sections of this report please keep in mind that a (+) indicates a correct response, a letter indicates the incorrect response selected, and a blank indicates that no response was selected. In the columns for the short-answer and constructed-response results, the numbers indicate the points awarded per item and a blank indicates that the item was not answered. All responses to released items are reported in the *NECAP Item Analysis Report*, regardless of the student's participation status.

The first column of this report lists each student alphabetically by last name. The column after the released items shows Total Test Results, broken into several categories. Subcategory Points Earned columns report the points the student earned in each content strand. The Total Points Earned column is a summary of all of the points earned in each of the content areas. The last two columns show the Scaled Score and Achievement Level for each student. For students who are reported as Not Tested, a code appears in the Achievement Level column to indicate the reason the student did not test. The descriptions of these codes can be found on the legend, after the last page of data on the *NECAP Item Analysis Report.* It is important to note that not all items used to compute student scores are included in this report. Only those items that have been released are included. The Percent Correct/Average Score for the school, district, and state are listed at the end of each report after the student data.

This *NECAP Item Analysis Report* is confidential and should be kept secure within the school and district. Remember, the Family Educational Rights and Privacy Act (FERPA) requires that access to individual student results be restricted to the student, the student's parents/guardians, and authorized school personnel.

The following page is a sample *NECAP Item Analysis Report* for mathematics.

Fall 2005 - Beginning of Grade 3 NECAP Tests
Grade 3 Students in 2005-2006
Item Analysis Report
Mathematics

School:
District:
State:
Code:

	Released Items																Total Test Results						
																	Subcategory Points Earned						
Released Item Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16							
Content Strand	NO	NO	NO	NO	NO	GM	FA	FA	DP	DP	NO	FA	DP	NO	NO	GM							
GLE Code	2-1	2-1	2-2	2-3	2-5	2-6	2-1	2-1	2-2	2-4	2-2	2-4	2-1	2-1	2-3	2-7							
Depth of Knowledge Code	2	2	1	1	2	1	2	2	2	2	1	1	3	2	2	2							
Item Type	MC	MC	MC	MC	MC	MC	MC	MC	MC	MC	SA	SA	SA	SA	SA	SA							
Correct MC Response	B	D	C	A	C	A	D	B	D	B													
Name/Student ID — Total Possible Points	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	35	10	10	10		65	
	+	+	+	+	+	+	+	+	+	+	1	1		2	2	1	26	8	9	9		52	351	3
	C	+	+	D	+	+	C	A	+	A	1	0	0	1	2	0	22	6	5	5		38	340	3
	+	+	+	+	+	+	+	+	+	A	1	1	1	2	2	2	29	9	10	8		56	355	4
	+	+	+	+	+	D	+	+	+	+	1	1	0	2	0	2	26	8	8	9		51	350	3
	+	+	+	+	+	+	+	+	+	A	1	0	1	2	2	0	29	5	8	7		49	348	3
	+	+	+	B	B		+	+	+	A	1	1		2	1	2	23	7	9	6		45	345	3
	+	+	B	+	A	C	+		+	D	1	0	0	2	0	0	22	5	5	6		38	340	3
	+	+	+	+	A	D	+	+	+	+	1	1	1	1	2	2	25	6	8	7		46	346	3

Released Item Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16					
Percent Correct/Average Score: School	88	100	88	75	63	50	88	75	100	38	1	0.6	0.4	1.8	1.4	1.1	25.3	6.8	7.8	7.1	
Percent Correct/Average Score: District	88	100	88	75	63	50	88	75	100	38	1	0.6	0.4	1.8	1.4	1.1	25.3	6.8	7.8	7.1	
Percent Correct/Average Score: State	74	76	87	76	59	53	74	81	92	46	0.9	0.6	0.4	1.5	1.1	0.7	23.2	6.1	6.9	6.3	

401-705-28400

Understanding the School and District Results Reports

Overview

The *NECAP School Results Report* and the *NECAP District Results Report* provide NECAP results for schools and districts based on the testing of local students in grades 3 through 8. A separate school report and district report has been produced for each grade level tested.

Although text in this section refers only to the *NECAP School Results Report*, educators and others who are reviewing the *NECAP District Results Report* should also refer to this section for applicable information because the data reported, report format, and guidelines for using the reported data are identical for both the school and district reports. The only real difference between the reports is that the *NECAP District Results Report* includes no individual school data.

IDENTIFICATION

The box in the upper right corner of each page shows the school name, district name, state, and school and district code.

BASIS FOR RESULTS

Results in the *NECAP School Results Report* are based on common items (with one exception described on page 22 of this document), and represent the aggregate of individual student scores (achievement level results and scaled scores).

TEACHING YEAR VS. TESTING YEAR

The data used for the *NECAP School Results Report* are the results of the fall 2005 administration of the NECAP test. The NECAP tests are based on the *Grade Level Expectations* (GLE) from the prior year. For example, the Grade 7 NECAP test, administered in the fall of seventh grade, is based on the grade 6 GLEs. Therefore, many students receive the instruction they need for this fall test at a different school from where they are currently enrolled. The state Departments of Education determined that it would be valuable for both the school where the student tested and the school where the student received instruction to have access to information that can help improve curriculum. To achieve this goal, separate *NECAP School Results Reports* have been created for the "testing" school and the "teaching" school. Every student who participated in the NECAP test will be represented in a "testing" school report, and most students will also be represented in a "teaching" school report. In some instances, such as when the student has recently moved into the state, it is not possible to provide information about a student in the "teaching" school report.

MINIMUM NUMBER OF STUDENTS NEEDED TO GENERATE REPORTS

To ensure confidentiality of individual student results and discourage generalizations about school performance based on very small populations, the Departments of Education in NH, RI and VT have established that groups of students must be larger than nine in order to report results in any particular reporting category. Consequently, schools with a very small number of students enrolled in a grade tested may not show results in some sections of their school report. A school report was not generated for any school that tested fewer than ten students at a particular grade; results for students in these schools are included in district- and/or state-level results.

MAKING COMPARISONS AMONG STUDENTS, SCHOOLS, AND DISTRICTS

The Departments of Education in NH, RI and VT do not encourage or promote comparisons among schools and districts. NECAP was designed so that each individual school or district can evaluate its performance against a set of *Grade Level Expectations* and achievement standards.

Scaled scores are the most suitable statistic to use when comparing NECAP results among students, schools, and districts. When interpreting the meaning of these comparisons, however, it is important that decision-makers—teachers, administrators, and policy-makers—fully recognize that any single test is a limited measure of student performance. Since some apparent differences in scaled scores may not be statistically or educationally significant, some guidelines for comparing results are explained on the following page.

COMPARISONS OF SCHOOL- AND DISTRICT-LEVEL SCORES

The statistical significance of these comparisons is based on variability of the scores and on the number of students tested. The table on the following page can be used to assist you in the following ways:

- comparing sub-populations of students within a school or district,
- comparing the scores of two or more schools or districts,
- comparing the scores of a school to the district and/or state,
- comparing the scores of a district to the state, and
- comparing a single school's score across two years.

Caution should be used when making comparisons between schools because even if scores are different they may not be statistically significantly different. It is very unlikely that any two groups will have exactly the same average score. To avoid misinterpretation or over-interpretation of small differences between scores, statistical tests can be conducted to determine the likelihood that the observed difference in scores occurred by chance and that the two groups might actually have the same score.

NECAP scaled scores are represented by a 3-digit number, with the first digit representing the grade level tested; the remaining digits range from 00–80. Although this same scale is used for all three content areas, one cannot accurately compare a school's or district's scaled scores across two content areas since the scaled scores in each content area were determined by separate standard-setting processes. The table on the following page shows the smallest differences in scores that represent a statistically significant difference in performance based on the number of students tested in the school and/or district. If comparing two entities of different sizes, one should use a difference that is approximately the average of the minimally statistically significant difference of each entity. For example, if comparing the average Grade 3 Reading score of a school with 25 students to the average Grade 3 Reading score of a school with 50 students, one should use 4 points as the minimally statistically significant difference.

Number of Scaled Score Points Denoting Minimally Statistically Significant Difference for Average Group Results*

Grade	Subject	Number of Students Tested in Group (Class, School etc.)				
		10	25	50	100	200
3	Reading	6	4	3	2	1
	Mathematics	5	3	2	2	1
4	Reading	6	4	3	2	1
	Mathematics	6	4	3	2	1
5	Reading	6	3	2	2	1
	Mathematics	5	3	2	2	1
	Writing	6	4	3	2	1
6	Reading	6	4	3	2	1
	Mathematics	5	3	2	2	1
7	Reading	6	4	3	2	1
	Mathematics	5	3	2	2	1
8	Reading	6	4	3	2	1
	Mathematics	5	3	2	2	1
	Writing	5	3	2	2	1

*Standard error of the mean difference with assumption n1=n2 and s1=s2

Comparisons across content areas can also be made by comparing the percentage of students at a particular achievement level. But again, since the classification of students into achievement levels carries a small degree of imprecision, small differences in percentages should not be over-interpreted.

The table below shows the minimum differences necessary to determine if the percentage of students in a given achievement level is statistically different from the percentages in another category or between years. For example, if 20% of students are Proficient in 2005, then to observe a statistically significant change for a school of 25 students, there would need to be at least 28% in this category in 2006.

Percentage Difference in Student Achievement Level Classification Denoting Minimally Statistically Significant Differences for Group Results*

Percentages of Students in Category	Number of Students Tested in Group (Class, School etc.)				
	10	**25**	**50**	**100**	**200**
10	9	6	4	3	2
20	13	8	6	4	3
30	14	9	6	5	3
40	15	10	7	5	3
50	16	10	7	5	4
60	15	10	7	5	3
70	14	9	6	5	3
80	13	8	6	4	3
90	9	6	4	3	2

*0.05 level of statistical significance

The table below shows the scaled scores that mark the cut point between achievement levels.

Achievement Level Cut Scores

Grade	Subject	SP/PP*	PP/P*	P/PD*
3	Reading	330 / 331	339 / 340	356 / 357
	Mathematics	331 / 332	339 / 340	352 / 353
4	Reading	430 / 431	439 / 440	455 / 456
	Mathematics	430 / 431	439 / 440	454 / 455
5	Reading	529 / 530	539 / 540	555 / 556
	Mathematics	532 / 533	539 / 540	553 / 554
	Writing	526 / 527	538 / 539	553 / 554
6	Reading	628 / 629	639 / 640	658 / 659
	Mathematics	632 / 633	639 / 640	652 / 653
7	Reading	728 / 729	739 / 740	759 / 760
	Mathematics	733 / 734	739 / 740	751 / 752
8	Reading	827 / 828	839 / 840	858 / 859
	Mathematics	833 / 834	839 / 840	851 / 852
	Writing	827 / 828	838 / 839	854 / 855

*SP = Substantially Below Proficient, PP = Partially Proficient, P = Proficient,
PD = Proficient with Distinction

The following page contains a sample cover of a *NECAP School Results Report* for both "teaching year" and "testing year".

About The New England Common Assessment Program

This report highlights results from the Fall 2005 Beginning of Grade New England Common Assessment Program (NECAP) tests. The NECAP tests are administered to students in New Hampshire, Rhode Island, and Vermont as part of each state's statewide assessment program. NECAP test results are used primarily for school improvement and accountability. Achievement level results are used in the state accountability system required under No Child Left Behind (NCLB). More detailed school and district results are used by schools to help improve curriculum and instruction. Individual student results are used to support information gathered through classroom instruction and assessments.

NECAP tests in reading and mathematics are administered to students in grades 3 through 8 and writing tests are administered to students in grades 5 and 8. The NECAP tests are designed to measure student performance on grade level expectations (GLE) developed and adopted by the three states. Specifically, the tests are designed to measure the content and skills that students are expected to have as they begin the school year in their current grade – in other words, the content and skills which students have learned through the end of the previous grade.

Each test contains a mix of multiple-choice and constructed-response questions. Constructed-response questions require students to develop their own answers to questions. On the mathematics test, students may be required to provide the correct answer to a computation or word problem, draw or interpret a chart or graph, or explain how they solved a problem. On the reading test, students may be required to make a list or write a few paragraphs to answer a question related to a literary or informational passage. On the writing test, students are required to provide a single extended response of 1-3 pages and three shorter responses to questions measuring different types of writing.

This report contains a variety of school- and/or district-, and state-level assessment results for the NECAP tests administered at a grade level. Achievement level distributions and mean scaled scores are provided for all students tested as well as for subgroups of students classified by demographics or program participation. The report also contains comparative information on school and district performance on subtopics within each content area tested.

In addition to this report of grade level results, schools and districts will also receive Summary Reports, Item Analysis Reports, Released Item support materials, and student-level data files containing NECAP results. Together, these reports and data constitute a rich source of information to support local decisions in curriculum, instruction, assessment, and professional development. Over time, this information can also strengthen school's and district's evaluation of their ongoing improvement efforts.

3/15/2006

The map for your state will appear here

Fall 2005
Beginning of Grade 3
NECAP Tests

Grade 2 Students in 2004-2005

School Results

School:
District:
Code:

"Teaching" Year

About The New England Common Assessment Program

This report highlights results from the Fall 2005 Beginning of Grade New England Common Assessment Program (NECAP) tests. The NECAP tests are administered to students in New Hampshire, Rhode Island, and Vermont as part of each state's statewide assessment program. NECAP test results are used primarily for school improvement and accountability. Achievement level results are used in the state accountability system required under No Child Left Behind (NCLB). More detailed school and district results are used by schools to help improve curriculum and instruction. Individual student results are used to support information gathered through classroom instruction and assessments.

NECAP tests in reading and mathematics are administered to students in grades 3 through 8 and writing tests are administered to students in grades 5 and 8. The NECAP tests are designed to measure student performance on grade level expectations (GLE) developed and adopted by the three states. Specifically, the tests are designed to measure the content and skills that students are expected to have as they begin the school year in their current grade – in other words, the content and skills which students have learned through the end of the previous grade.

Each test contains a mix of multiple-choice and constructed-response questions. Constructed-response questions require students to develop their own answers to questions. On the mathematics test, students may be required to provide the correct answer to a computation or word problem, draw or interpret a chart or graph, or explain how they solved a problem. On the reading test, students may be required to make a list or write a few paragraphs to answer a question related to a literary or informational passage. On the writing test, students are required to provide a single extended response of 1-3 pages and three shorter responses to questions measuring different types of writing.

This report contains a variety of school- and/or district-, and state-level assessment results for the NECAP tests administered at a grade level. Achievement level distributions and mean scaled scores are provided for all students tested as well as for subgroups of students classified by demographics or program participation. The report also contains comparative information on school and district performance on subtopics within each content area tested.

In addition to this report of grade level results, schools and districts will also receive Summary Reports, Item Analysis Reports, Released Item support materials, and student-level data files containing NECAP results. Together, these reports and data constitute a rich source of information to support local decisions in curriculum, instruction, assessment, and professional development. Over time, this information can also strengthen school's and district's evaluation of their ongoing improvement efforts.

3/15/2006

The map for your state will appear here

Fall 2005
Beginning of Grade 3
NECAP Tests

Grade 3 Students in 2005-2006

School Results

School:
District:
Code:

"Tested" Year

Grade Level Summary Report

(Second page of the *NECAP School Results Report*)

The second page, titled "Grade Level Summary Report", provides a summary of participation in NECAP and a summary of NECAP results. This page shows the number and percentage of students who were enrolled, tested, and not tested as part of the NECAP test in fall 2005. Students enrolled in a school on or after October 1, 2005 were expected to complete the NECAP test at that school.

STUDENTS ENROLLED ON OR AFTER OCTOBER 1

The first table in the "Grade Level Summary Report" shows the number of students enrolled in the tested grade. The total number of students reported as enrolled is defined as the number of students tested added to the numbers of students who were not tested.

STUDENTS NOT TESTED IN NECAP

Since students who were not tested did not participate in the NECAP tests, average school scores are not affected by not tested students. These students are included in the calculation of the percent that participated, but are not included in the calculation of scores.

For students who participated in some but not all parts of the NECAP test, their actual score was reported for each content area in which they participated. These reporting decisions were made to support the requirement that all students must participate in the NECAP testing program.

Data is provided for the following groups of students who may not have completed the entire battery of NECAP tests.

- Alternate Assessment—Students in this category completed an alternate assessment for the 2004-2005 school year.
- First Year LEP—Students in this category are defined as being new to the US after October 1, 2004 and were not required to take the NECAP tests in reading and writing. Students in this category were expected to take the mathematics portion of the NECAP.
- Withdrew After October 1—Students withdrawing from a school after October 1, 2005 may have taken some sessions of the NECAP tests prior to their withdrawal from the school.
- Enrolled After October 1— Students enrolling in a school after October 1, 2005 may not have had adequate time to fully participate in all sessions of the NECAP tests.
- Special Consideration—Schools received state approval for special consideration for an exemption for all or part of the NECAP tests for any student whose circumstances are not described by the previous categories, but for whom the school determined that taking the NECAP tests would not be possible.
- Other – Occasionally students will not have completed the NECAP tests for reasons other than those listed above. These "other" categories are considered "not state approved".

NECAP RESULTS

The results portion of the page indicates the number and percentage of students performing at each achievement level in each of the three content areas tested by NECAP. In addition, a Mean Scaled Score is provided for each content area at the school, district, and state levels.

The following page contains a sample of the "Grade Level Summary Report" from a *NECAP School Results Report.*

Fall 2005 - Beginning of Grade 5 NECAP Tests
Grade 5 Students in 2005-2006
Grade Level Summary Report

School:
District:
State:
Code:

Schools and districts administered all NECAP tests to every enrolled student with the following exceptions: students who participated in the alternate assessment for the 2004-05 school year, first year LEP students, students who withdrew from the school after October 1, 2005, students who enrolled in the school after October 1, 2005, students for whom a special consideration was granted through the state Department of Education, and other students for reasons not approved. On this page, and throughout this report, results are only reported for groups of students that are larger than nine (9).

PARTICIPATION in NECAP	Number									Percentage								
	School			District			State			School			District			State		
Students enrolled on or after October 1	95			173			11,589			100			100			100		
	Reading	Math	Writing	Reading	Math	Writing	Reading	Math	Writing	Reading	Math	Writing	Reading	Math	Writing	Reading	Math	Writing
Students tested	**95**	**95**	**95**	**173**	**173**	**173**	**11,404**	**11,476**	**11,371**	**100**	**100**	**100**	**100**	**100**	**100**	**98**	**99**	**98**
Students not tested in NECAP																		
State Approved	0	0	0	0	0	0	176	96	192	0	0	0	0	0	0	2	1	2
Alternate Assessment	0	0	0	0	0	0	53	53	53	0	0	0	0	0	0	0	0	0
First Year LEP	0	0	0	0	0	0	84	0	84	0	0	0	0	0	0	1	0	1
Withdrew After October 1	0	0	0	0	0	0	20	31	40	0	0	0	0	0	0	0	0	0
Enrolled After October 1	0	0	0	0	0	0	15	8	10	0	0	0	0	0	0	0	0	0
Special Consideration	0	0	0	0	0	0	4	4	5	0	0	0	0	0	0	0	0	0
Other	0	0	0	0	0	0	9	17	26	0	0	0	0	0	0	0	0	0

NECAP RESULTS

	School													District						State					
	Enrolled	NT Approved	NT Other	Tested	Level 4		Level 3		Level 2		Level 1		Mean Scaled Score	Tested	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score	Tested	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score
	N	N	N	N	N	%	N	%	N	%	N	%		N	%	%	%	%		N	%	%	%	%	
READING	95	0	0	95	12	13	52	55	24	25	7	7	544	173	10	52	27	12	543	11,404	13	47	25	16	543
MATH	95	0	0	95	13	14	46	48	24	25	12	13	543	173	10	45	31	14	542	11,476	11	41	22	25	540
WRITING	95	0	0	95	11	12	38	40	34	36	12	13	540	173	9	45	35	11	540	11,371	12	43	30	15	539

Level 4 = Proficient with Distinction; Level 3 = Proficient; Level 2 = Partially Proficient; Level 1 = Substantially Below Proficient

Note: Throughout this report, percentages may not total 100 since each percentage is rounded to the nearest whole number.

Page 2 of 8

Content Area Results

(Pages 3, 5, and 7 of the *NECAP School Results Report*)

Content area subscores provide information on performance in specific subtopics of the tested content areas (for example, geometry and measurement within mathematics). Subscore results by content area tested are provided on the following pages of the report:

- page 3—reading,
- page 5—mathematics, and
- page 7—writing.

The purpose of these sections is to help schools determine the extent to which their curricula are effective in helping students achieve the particular standards and benchmarks contained in the *Grade Level Expectations*.

Information about each content area (reading, mathematics and writing) for school, district and state includes:

- the total number of students Enrolled, Not Tested (state-approved reason), Not Tested (other reason), and Tested;
- the total number and percent of students at each achievement level (based on the number in the Tested column); and
- the Mean Scaled Score.

Information about each content area subtopic for reading, mathematics and writing includes:

- The **Total Possible Points** for that category. In order to provide as much information as possible for each category, the total number of points includes both the common items used to calculate scores as well as additional items in each category used for equating the test from year to year.
- A graphic display of the **Percent of Total Possible Points** for the school, state and district. In this graphic display, there are symbols representing school, district and state performance. In addition, there is a line representing the standard error of measurement. This statistic indicates how much a student's score could vary if the student was examined repeatedly with the same test (assuming that no learning occurs between test administrations).

The following page contains a sample "Reading Results" page from a *NECAP School Results Report.*

Fall 2005 - Beginning of Grade 5 NECAP Tests
Grade 5 Students in 2005-2006
Reading Results

School:
District:
State:
Code:

Proficient with Distinction
Student's performance demonstrates an ability to read and comprehend grade-appropriate text. Student is able to analyze and interpret literary and informational text. Student offers insightful observations/assertions that are well supported by references to the text. Student uses range of vocabulary strategies and breadth of vocabulary knowledge to read and comprehend a wide variety of texts.

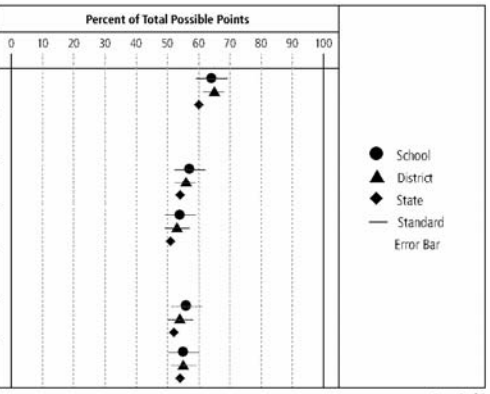
Proficient
Student's performance demonstrates an ability to read and comprehend grade-appropriate text. Student is able to analyze and interpret literary and informational text. Student makes and supports relevant assertions by referencing text. Student uses vocabulary strategies and breadth of vocabulary knowledge to read and comprehend text.

Partially Proficient
Student's performance demonstrates an inconsistent ability to read and comprehend grade-appropriate text. Student attempts to analyze and interpret literary and informational text. Student may make and/or support assertions by referencing text. Student's vocabulary knowledge and use of strategies may be limited and may impact the ability to read and comprehend text.

Substantially Below Proficient
Student's performance demonstrates minimal ability to derive/construct meaning from grade-appropriate text. Student may be able to recognize story elements and text features. Student's limited vocabulary knowledge and use of strategies impacts the ability to read and comprehend text.

	Enrolled	NT Approved	NT Other	Tested	Level 4		Level 3		Level 2		Level 1		Mean Scaled Score
	N	N	N	N	N	%	N	%	N	%	N	%	
SCHOOL													
2005-06	95	0	0	95	12	13	52	55	24	25	7	7	544
2006-07													
2007-08													
Cumulative													
Average													
DISTRICT													
2005-06	173	0	0	173	17	10	90	52	46	27	20	12	543
2006-07													
2007-08													
Cumulative													
Average													
STATE													
2005-06	11,589	176	9	11,404	1,494	13	5,309	47	2,831	25	1,770	16	543
2006-07													
2007-08													
Cumulative													
Average													

Subtopic	Total Possible Points
Word ID/Vocabulary	25
Type of Text	
Literary	56
Informational	49
Level of Comprehension	
Initial Understanding	65
Analysis & Interpretation	40



Page 3 of 8

Disaggregated Content Area Results

(Pages 4, 6, and 8 of the *NECAP School Results Report*)

- page 4—reading,
- page 6—mathematics, and
- page 8—writing.

Students can be grouped according to many characteristics—gender, ethnicity, school programs, etc. The scores provide information on achievement for different groups in a school, males and females for example.

The performance of subgroups is included on the "Disaggregated Content Area Results" pages of the *NECAP School Results Report* for reading, mathematics, and writing. These sections present the relationship between the variables reported and performance in each content area at the school, district and state levels. The tables show the number of students categorized as Enrolled, Not Tested (state-approved reason), Not Tested (other reason), and Tested. The tables also provide the number and percentage of students within the subgroup at each of the four achievement levels, as well as the Mean Scaled Score. The data for achievement levels and Mean Scaled Score is based on the number shown in the Tested column. The data for the reporting categories was provided by information coded on the students' answer booklets by teachers and/or data linked to the student label. Because performance is being reported by categories that can contain relatively low numbers of students, school personnel are advised, under FERPA guidelines, to treat these pages confidentially.

The following page contains a sample "Disaggregated Mathematics Results" page from a *NECAP School Results Report.* Please note that for NH and VT no data appears for 504 Plan in any of the content areas. In addition, for VT, no data appears for Title I in any of the content areas.

Fall 2005 - Beginning of Grade 5 NECAP Tests
Grade 5 Students in 2005-2006
Disaggregated Mathematics Results

School:
District:
State:
Code:

REPORTING CATEGORIES	SCHOOL													DISTRICT						STATE					
	Enrolled	NT Approved	NT Other	Tested	Level 4		Level 3		Level 2		Level 1		Mean Scaled Score	Tested	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score	Tested	Level 4	Level 3	Level 2	Level 1	Mean Scaled Score
	N	N	N	N	N	%	N	%	N	%	N	%	N	N	%	%	%	%	N	N	%	%	%	%	N
All Students	**95**	**0**	**0**	**95**	**13**	**14**	**46**	**48**	**24**	**25**	**12**	**13**	**543**	**173**	**10**	**45**	**31**	**14**	**542**	**11,476**	**11**	**41**	**22**	**25**	**540**
Gender																									
Male	49	0	0	49	6	12	30	61	7	14	6	12	544	88	11	51	25	13	542	5,836	12	41	21	26	540
Female	46	0	0	46	7	15	16	35	17	37	6	13	542	83	10	39	35	17	541	5,533	11	42	23	24	540
Not Reported	0	0	0	0										2						107	3	21	25	50	530
Primary Race/Ethnicity																									
American Indian or Alaskan Native	0	0	0	0										0						69	4	22	39	35	535
Asian	1	0	0	1										2						361	15	48	18	19	542
Black or African American	0	0	0	0										1						953	2	27	24	47	533
Hispanic or Latino	2	0	0	2										3						1,988	3	28	25	44	534
Native Hawaiian or Pacific Islander	0	0	0	0										0						0					
White (non-Hispanic)	92	0	0	92	12	13	46	50	23	25	11	12	543	167	10	46	31	14	542	8,092	15	46	21	18	543
No Primary Race/Ethnicity Reported	0	0	0	0										0						13	0	0	15	85	520
LEP Status																									
Currently receiving LEP services	0	0	0	0										0						624	1	10	18	71	527
Former LEP student - monitoring year 1	0	0	0	0										0						201	4	30	28	37	536
Former LEP student - monitoring year 2	0	0	0	0										0						128	6	33	31	30	538
All Other Students	95	0	0	95	13	14	46	48	24	25	12	13	543	173	10	45	31	14	542	10,523	12	44	22	22	541
IEP																									
Students with an IEP	13	0	0	13	0	0	8	62	1	8	4	31	537	31	3	35	29	32	536	2,248	3	22	23	52	532
All Other Students	82	0	0	82	13	16	38	46	23	28	8	10	544	142	12	46	31	11	543	9,228	14	46	22	19	542
SES																									
Economically Disadvantaged Students	18	0	0	18	0	0	10	56	5	28	3	17	539	39	3	46	31	21	538	4,440	4	30	26	39	535
All Other Students	77	0	0	77	13	17	36	47	19	25	9	12	544	134	13	44	31	13	543	7,036	16	48	19	16	543
Migrant																									
Migrant Students	0	0	0	0										0						0					
All Other Students	95	0	0	95	13	14	46	48	24	25	12	13	543	173	10	45	31	14	542	11,476	11	41	22	25	540
Title I																									
Students Receiving Title I Services	4	0	0	4										22	0	18	55	27	535	3,419	4	28	26	41	535
All Other Students	91	0	0	91	13	14	46	51	21	23	11	12	543	151	12	48	27	13	543	8,057	15	47	20	18	543
504 Plan																									
Students with a 504 Plan	0	0	0	0										0						26	4	38	31	27	540
All Other Students	95	0	0	95	13	14	46	48	24	25	12	13	543	173	10	45	31	14	542	11,450	11	41	22	25	540

Level 4 = Proficient with Distinction; Level 3 = Proficient; Level 2 = Partially Proficient; Level 1 = Substantially Below Proficient

NOTE: Some numbers may have been left blank because fewer than ten (10) students were tested.

Page 6 of 8

Understanding the School and District Summary Reports

Overview

The *NECAP School Summary Report* and the *NECAP District Summary Report* provide NECAP results for schools and districts based on the testing of local students in grades 3 through 8. A separate school report and district report has been produced for each grade level tested.

Although text in this section refers only to the *NECAP School Summary Report*, educators and others who are reviewing the *NECAP District Summary Report* should also refer to this section for applicable information because the data reported, report format, and guidelines for using the reported data are identical for both the school and district reports. The only real difference between the reports is that the *NECAP District Summary Report* includes no individual school data.

The *NECAP School Summary Report* provides details, broken down by content area, about student performance for all grade levels of NECAP that were tested in the school.

The purpose of this summary is to help schools determine the extent to which their students achieve the particular standards and benchmarks contained in the *Grade Level Expectations.*

Information about each content area and grade level for school, district and state includes:

- the total number of students Enrolled, Not Tested (state-approved reason), Not Tested (other reason), and Tested;
- the total number and percent of students at each achievement level (based on the number in the Tested column); and
- the Mean Scaled Score.

The following page contains a sample *NECAP School Summary Report.*

Fall 2005 NECAP Tests
School Summary
2005-2006 Students

School:
District:
State:
Code:

Reading	Enrolled	NT Approved	NT Other	Tested	Achievement Level								
					Level 4		Level 3		Level 2		Level 1		Mean Scaled Score
	N	N	N	N	N	%	N	%	N	%	N	%	
	39	**0**	**1**	**38**	**0**	**0**	**6**	**16**	**9**	**24**	**23**	**61**	
Beginning of Grade 3	3	0	0	3									
Beginning of Grade 4	6	0	1	5									
Beginning of Grade 5	8	0	0	8									
Beginning of Grade 6	7	0	0	7									
Beginning of Grade 7	5	0	0	5									
Beginning of Grade 8	10	0	0	10	0	0	0	0	6	60	4	40	825

Mathematics	Enrolled	NT Approved	NT Other	Tested	Achievement Level								
					Level 4		Level 3		Level 2		Level 1		Mean Scaled Score
	N	N	N	N	N	%	N	%	N	%	N	%	
	39	**0**	**4**	**35**	**0**	**0**	**3**	**9**	**2**	**6**	**30**	**86**	
Beginning of Grade 3	3	0	0	3									
Beginning of Grade 4	6	0	1	5									
Beginning of Grade 5	8	0	0	8									
Beginning of Grade 6	7	0	2	5									
Beginning of Grade 7	5	0	0	5									
Beginning of Grade 8	10	0	1	9									

Writing	Enrolled	NT Approved	NT Other	Tested	Achievement Level								
					Level 4		Level 3		Level 2		Level 1		Mean Scaled Score
	N	N	N	N	N	%	N	%	N	%	N	%	
	18	**0**	**3**	**15**	**0**	**0**	**0**	**0**	**0**	**0**	**15**	**100**	
Beginning of Grade 5	8	0	1	7									
Beginning of Grade 8	10	0	2	8									

Level 4 = Proficient with Distinction; Level 3 = Proficient; Level 2 = Partially Proficient; Level 1 = Substantially Below Proficient

Appendix A

Overview of Assessment Instruments and Procedures
NECAP Tests of 2005

Local Educator Involvement in Test Development

Local educators in all three NECAP states were actively involved in each aspect of the NECAP test development from the beginning of the collaboration among the three states. Educators have been involved in development of *Grade Level Expectations*, review of all passages and items for bias and sensitivity issues, review of all items for purposes of alignment, Depth of Knowledge, age appropriateness and accuracy of content. Local educators were also involved in standard setting and the Technical Advisory Committee.

Grade Level Expectation Development

The NH, RI and VT Departments of Education have developed a common set of grade level expectations, known as the *New England Common Assessment Program Grade Level Expectations* (GLE), and test specifications in mathematics, reading, and writing. These expectations were developed in response to the requirements of the federally mandated No Child Left Behind Act of 2001 to test all students, beginning in the 2005-2006 academic year, in each of grades 3 through 8 in mathematics and reading/language arts. Although these sets of GLEs were developed for this purpose, the partner states were committed to building coherent sets of expectations that would focus, not narrow, the curricula; would support good instruction; and would be aligned with each state's standards. Throughout the development process, each of the NECAP partners has relied upon the expertise of educators in their states. These educators have helped guide the development of these documents and have made numerous insightful contributions in an effort to help support meaningful instruction in mathematics and reading/language arts.

Item Review Committee

During the item review process, a committee of local educators is convened to review all of the items developed for NECAP. Committee member comments are solicited for each item. Each item is evaluated on the following four criteria:

- alignment with the GLE being measured;
- accurate Depth of Knowledge coding;
- appropriateness for grade-level; and
- content accuracy.

Bias and Sensitivity Committee

A committee of local educators also meets to review all reading passages and individual test items. Committee members determine if the passages and items are likely to place a particular group of students at an advantage or disadvantage for non-educational reasons; and if so, whether the passage or item should be revised or removed.

Technical Advisory Committee

A committee of nationally recognized test and measurement experts and local educators has been established and meets regularly to ensure the technical integrity of NECAP tests.

Test Design

TYPES OF ITEMS ON NECAP

In order to provide a valid assessment of students' attainment of the *Grade Level Expectations*, a variety of item types needed to be used. Therefore, multiple-choice items, short-answer items constructed-response items and extended-response writing prompts were used as follows.

Multiple choice (one point)

Multiple-choice items are efficient for testing a broad array of content in a relatively short time span.

Short answer (one point and two point)

These open-ended items ask students to generate a short response to a question.

Constructed response (four points)

This is a more complex item type that requires students to give a longer response to items related to a reading passage or solve multi-step mathematics problems.

Extended-response writing prompts (twelve points)

These are topics or questions designed to prompt students to respond in writing. Students compose a response to the writing prompt.

COMMON AND MATRIX-SAMPLED ITEMS

There are multiple versions, or forms, of the NECAP tests; nine forms were created for each grade level tested in reading and mathematics. Approximately half of the items in each of the NECAP test forms were the same in every form, or were "common" to all forms of the test. All individual student results (achievement levels, scaled scores, content area subscores) and school results are based on only common items. The other half of the items in each form were matrix sampled. "Matrix sampling" means distributing a large number of items among the different forms of the test. This approach allows for field testing of new items for subsequent years' tests and also allows some items to be administered in successive years for purposes of equating the tests from year to year.

The writing test is entirely common, meaning that every student in grade 5 took the same writing test, and each student in grade 8 took the same writing test.

A portion of common items is publicly released following each year's test administration to inform local curriculum and instruction. Released common items are replaced each year with some of the items from the previous year's matrix-sampled section.

Content Knowledge and Skills Tested on NECAP

All items appearing on the NECAP test were designed to measure a specific GLE. The GLE documents for each content area can be found at each state Department of Education website (see page 1 for DOE web addresses).

READING OVERVIEW

The NECAP reading test consists of 42 multiple-choice items and 9 constructed- response items at all grades.

The reading passages on the NECAP test are broken down into the following categories:

- Literary passages representing a variety of forms—modern narratives; diary entries; drama; poetry; biographies; essays; excerpts from novels; short stories; and traditional narratives such as fables, tall tales, myths, and folktales.
- Informational passages, which are factual texts and often deal with the areas of science and social studies. These passages are taken from sources such as newspapers, magazines, and excerpts from books. Informational text also includes directions, manuals, or recipes.

The passages are authentic texts—selected from grade-level appropriate reading sources— that students would be likely to experience in both classroom and independent reading. None of the passages are written specifically for the assessment, but instead are collected from published works.

The items on the NECAP test are categorized by both the type of passage associated with the item and also whether the item measured lower or higher level comprehension. The level of comprehension is designated as either "Initial Understanding" or "Analysis and Interpretation".

Word identification and vocabulary skills are tested, primarily through multiple-choice items, at each grade level.

Reading Distribution of Emphasis

	2 (3)	3(4)	4(5)	5(6)	6(7)	7(8)
Word Identification Sills and Strategies	20%	15%	10%	0%	0%	0%
Vocabulary Strategies/Breadth of Vocabulary	20%	20%	20%	20%	25%	25 %
Initial Understanding of Literary Text	20%	20%	20%	20%	15%	15%
Initial Understanding of Informational Text	20%	20%	20%	20%	20%	20%
Analysis and Interpretation of Literary Text	10%	15%	15%	20%	20%	20%
Analysis and Interpretation of Informational Text	10%	10%	15%	20%	20%	20%
	100%	100%	100%	100%	100%	100%

MATHEMATICS OVERVIEW

The NECAP mathematics test at grades 3 and 4 consists of 44 multiple-choice items, 13 one-point short-answer items, and 13 two-point short-answer items.

The NECAP mathematics tests at grades 5 through 8 consists of 41 multiple-choice items, 9 one-point short-answer items, 9 two-point short-answer items, and 6 constructed-response items.

The content standards in mathematics identify four major strands.

- Numbers and Operations
- Geometry and Measurement
- Functions and Algebra
- Data, Statistics, and Probability

In addition, problem solving, reasoning, connections and communication are embedded throughout the GLEs.

Mathematics Distribution of Emphasis

	2(3)	3(4)	4(5)	5(6)	6(7)	7(8)
Number and Operations	55%	50%	50%	45%	30%	20%
Geometry and Measurement	15%	20%	20%	25%	25%	25 %
Algebra and Functions	15%	15%	15%	15%	30%	40%
Data, Statistics, and Probability	15%	15%	15%	15%	15%	15%
	100%	100%	100%	100%	100%	100%

WRITING OVERVIEW

The NECAP writing test at grades 5 and 8 consists of 12 multiple-choice items, 3 constructed-response items, 3 one-point short-answer items, and one extended-response writing prompt.

NECAP assesses students' writing skills directly through the use of writing prompts, or topics, to which students respond. In the 2005 tests, all students were administered one extended writing prompt. Students also completed constructed-response items. Each of the constructed-response items and the writing prompt addressed a different genre of writing.

The content standards in writing identify four major genres.

- Writing in Response to Literary Text
- Writing in Response to Informational Text
- Narratives
- Informational Writing (Report/Procedure at grade 5 and Persuasive at grade 8)

Each year, all four genres of writing are assessed in the writing portion of the NECAP test. In addition, structures and conventions of language are assessed through multiple-choice items and throughout the student's writing.

Writing Distribution of Emphasis

	2 (3)	3(4)	4(5)	5(6)	6(7)	7(8)
Structures			Less emphasis			Less emphasis
Responsive to Text			Greater emphasis			Greater emphasis
Narrative			Greater emphasis			Less emphasis
Informational Writing			Greater emphasis (Reports)			Greater emphasis (Reports or Persuasive)
Conventions			Less emphasis			Less emphasis
			100%			100%

Administration Procedures for NECAP

Guidelines for test scheduling, student participation, and test security, as well as detailed administration manuals, were provided to districts and schools prior to the October 2005 testing period. Training on test administration procedures was provided through five or more Test Administration Workshops held in each of the three states three weeks prior to testing.

Student Participation
All students were to participate in the assessment in one of the following three ways:

- the general assessment without accommodations,
- the general assessment with accommodations, or
- state-specific alternate assessment.

The decision about how a student with disabilities would participate using accommodations was made at the local level. Guidance in making these decisions was available through each state's Department of Education and through use of the *NECAP Accommodations Training Guide*, available at the DOE website for each state.

Test Scheduling
The NECAP Reading and Mathematics tests were designed to be administered in six separate sessions. For students in grades 5 and 8, two additional writing sessions were administered. The guidelines for scheduling test sessions were based on an estimate that each session would require approximately forty-five minutes and all students were allowed up to ninety minutes per session. Administrators were instructed to allow extra time for any students who required test accommodations that could not be made during the regular test sessions. For scheduling purposes, each session was treated as an intact unit. That is, once students started a session of the test they had to finish it within the time allotted; also, under no circumstances were they allowed to go back to an earlier session once they had moved on to another session.

Scoring

In November 2005, more than 26 million responses were processed and scored at Measured Progress. The scoring activities that were used to produce the results for the NECAP reports are described below.

Scoring was separated into the following four major tasks:

- scoring of responses to multiple-choice items,
- scoring of responses to short-answer items,
- scoring of responses to constructed-response items, and
- scoring of extended-response writing prompts.

SCORING OF MULTIPLE-CHOICE ITEMS

Multiple-choice items were machine-scored using digital scanning equipment. Correct responses were assigned a score of one point each; incorrect or blank responses were assigned a score of zero points each.

SCORING OF SHORT-ANSWER AND CONSTRUCTED-RESPONSE ITEMS

Short-answer and constructed-response items were scored by scorers employed by Measured Progress, the testing contractor. Short-answer items were given a score from zero to one or zero to two. Constructed-response items were given a score from zero to four. Zeros are employed when a student produces some work, but the work is totally wrong or irrelevant or if he or she leaves the item blank. For purposes of aggregating item results, blanks and zeros both count as zero points towards a student's score.

The work in preparation for scoring student responses included:

- development of scoring guides (rubrics) by content specialists from the NH, RI and VT Departments of Education and Measured Progress's test developers, and
- selection of "benchmark" responses—examples of student work at different score points for each item—that were used in training and continuous monitoring of scorer accuracy.

Scorer training consisted of:

- review of each item and its related content and performance standard,
- review and discussion of the scoring guide and multiple sets of benchmark responses for each score point, and
- qualifying rounds of scoring in which scorers needed to demonstrate a prescribed level of accuracy.

SCORING OF EXTENDED RESPONSES

Extended-response items (writing test only) were given a score from zero to six. Zeros are employed when a student produces some work, but the work is totally wrong or irrelevant or if he or she leaves the item blank. For purposes of aggregating item results, blanks and zeros both count as zero points towards a student's score. All NECAP extended response items are 100% double-blind scored. Double-blind scoring refers to the method of scoring whereby two scorers score the same response and neither scorer has any indication as to what score the other person has given. If there is a difference in reader scores that is greater than 1 score point, then the response goes into an arbitration queue. Senior scoring staff members score all arbitration responses without knowing the scores given by the two previous readers. The state Departments of Education defined how resolution should be reached if discrepant scores arise.

Preparation for scoring extended-response items included the selection of benchmark responses that mirrored the work described on page 34 for scoring short-answer and constructed-response items. Scorers were trained by grade level and content area in large sessions by senior scoring staff for that grade.

Setting Standards for Performance on the NECAP Tests

Standard setting is the process of determining the minimum or "threshold" score for each achievement level, grade, and subject for which results are reported. The multi-step process of setting standards for the NECAP tests began with creation of achievement level descriptions.

In January 2006, the state Departments of Education in NH, RI and VT convened panels of educators to participate in the standard-setting process for NECAP. A challenging aspect of standard setting is that many methods exist to set standards and establish cut points. With this in mind, the Departments of Education, in consultation with the NECAP Technical Advisory Committee and Measured Progress, determined that three kinds of judgments would be employed for setting standards on the NECAP tests.

Upon completion of the data gathering phases of standard setting, the Commissioners of each state met on January 20, 2006 to approve the recommended cut points.

TEACHER JUDGMENTS

The judgments captured by this standard-setting method are not based on the actual NECAP tests; rather, the standards rely on teacher judgments in matching students to the same achievement descriptors established for NECAP. Teachers were asked to judge how well their students were prepared to succeed in each content area in the upcoming school year by reviewing the achievement level descriptions and considering each student. The achievement level descriptors were given insignificant labels (i.e. 1, 2, 3, 4) to encourage teachers to focus on the descriptors and not on labels such as Proficient.

Once this information was collected from teachers, preliminary data were established by calculating the percentage of students teachers rated in each category for each content area and grade level.

BOOKMARK STANDARD-SETTING PROCESS

One standard setting panel was convened for each grade level (3 through 8) in mathematics and another in reading. More than 100 panelists participated in two days of meetings to set the standards for each content area. Panelists were local educators invited to participate by the NH, RI and VT Departments of Education.

The Bookmark method of standard setting is a multi-step process. First, participants took the NECAP test as though they were students. Then, as a group, the panels reviewed the achievement level descriptions, paying special attention to differentiating between knowledge, skills and abilities typically associated with students described as being on the borderline between achievement levels. Panelists then looked at "Ordered Item Booklets", which show each common item on the test in order from easiest to hardest. The "Ordered Item Booklet" also includes actual student work samples for each score point for short-answer and constructed-response items. Participants made decisions about which items would differentiate between students at each achievement level and placed a "bookmark" between those items to represent the cut point between achievement levels. In addition, panelists were provided with Score Profiles (in intervals of 20-25 points) that show average or typical item scores for a student whose total score is at the center of that interval. Panelists looked at the profiles that would fall

into each achievement level category based on their cut points and determined if they thought their cut point placements made sense. Small- and large-group discussions followed regarding the knowledge, skills and abilities associated with the items around each cut point. Participants had the opportunity to change their placement of the "bookmark" based on these discussions. Finally, panelists had the opportunity to provide feedback on the achievement level descriptions.

BODY OF WORK STANDARD-SETTING PROCESS

One standard setting panel was convened for each grade level (5 and 8) in writing. Approximately 30 panelists participated in two days of meetings to set the standards for writing. Panelists were local educators invited to participate by the NH, RI and VT Departments of Education.

The Body of Work method of standard setting was used for writing. Similar to the Bookmark method described above, participants took the NECAP test, reviewed achievement level descriptions, and discussed knowledge, skills and abilities typically associated with "borderline" students. Participants were then provided with identical sets of student work (including responses to all relevant item types) from approximately 25 different, anonymous students distributed along the raw score continuum. Panelists were asked to independently (without discussion) categorize the student sets as Proficient with Distinction, Proficient, Partially Proficient, or Substantially Below Proficient based on their match to the achievement level. Panelists were also asked to write brief comments rationalizing each of their ratings and to help refresh their memories when discussing their ratings. After completing their categorizations, panelists reviewed and discussed all of the sets of student work, and the facilitator was instructed to encourage discussion of ratings where there was a noticeable split among panelists. After the discussion of each student's work, panelists were provided the opportunity to change any of their ratings. Second ratings were recorded in a column immediately adjacent to the first ratings on each panelist's rating form. Finally, panelists had the opportunity to provide feedback on the achievement level descriptions.

Reporting

The NECAP tests were designed to measure student performance against the learning goals described in the *Grade Level Expectations*. Consistent with this purpose, primary results on the NECAP tests are reported in terms of achievement levels that describe student performance in relation to these established state standards. There are four achievement levels: Proficient with Distinction, Proficient, Partially Proficient, and Substantially Below Proficient. Students receive a separate achievement-level classification (based on total scaled score) in each content area in which they complete a test. There is no overall classification of student performance across content areas. School- and district-level results are reported as the number and percentage of students attaining each achievement level at each grade level tested.

In addition to achievement levels, NECAP results are also reported as scaled scores. The major purpose of including scaled scores in NECAP reports is to enhance the level of feedback provided to students, parents, and teachers. Each of the four achievement levels encompasses a range of student performance. A student whose test performance is just above Substantially Below Proficient and a student whose level of performance is slightly below Proficient are both classified as Partially Proficient. However, scaled-score results are more precise since they pinpoint a student's performance (score) on the continuum of scores within the achievement levels. The additional information provided by scaled scores is critical in forming the most accurate impression of performance possible.

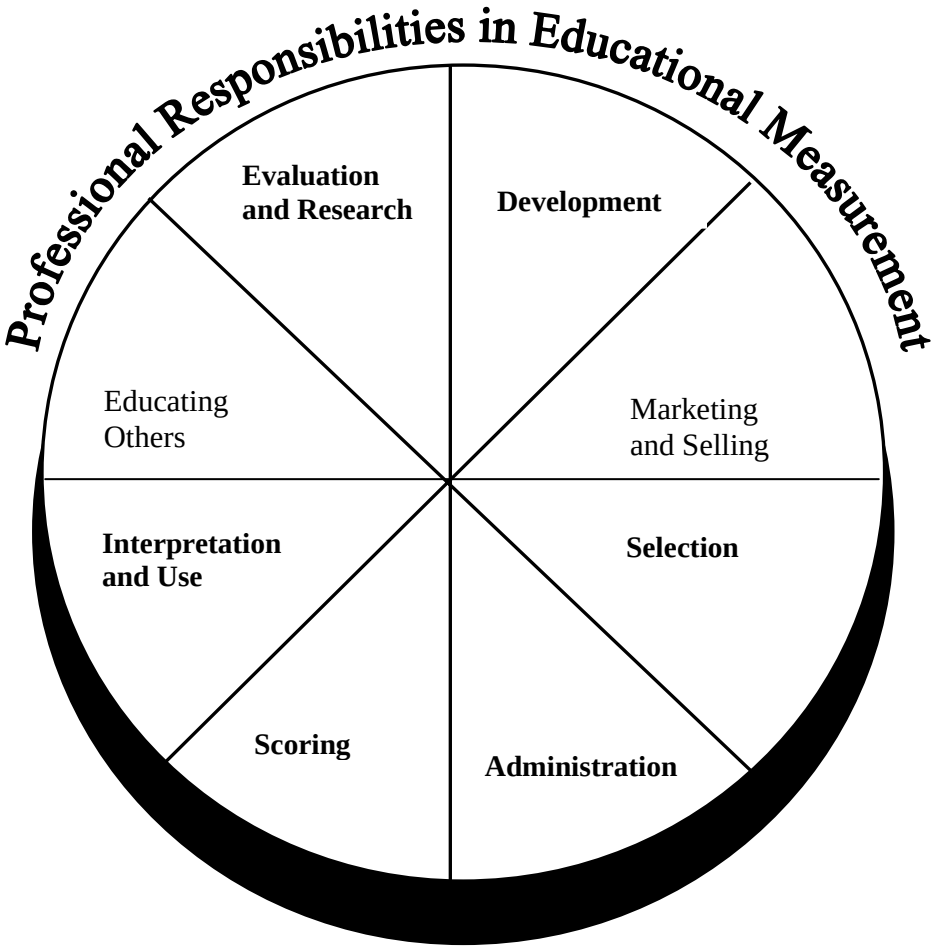
Translating Raw Scores to Scaled Scores and Achievement Levels

NECAP scores in each content area are reported on a scale that ranges from 00 to 80. Scaled scores supplement the NECAP achievement-level results by providing information about the position of a student's results within an achievement level. School- and district-level scaled scores are calculated by computing the average of student-level scaled scores. Students' raw scores, or total number of points, on the NECAP tests are translated to scaled scores using a data analysis process called scaling. Scaling simply converts raw points from one scale to another. In the same way that the same temperature can be expressed on either the Fahrenheit or Celsius scales and the same distance can be expressed either in miles or kilometers, student scores on the NECAP tests could be expressed as raw scores (i.e., number right) or scaled scores.

It is important to note that converting from raw scores to scaled scores does not change the students' achievement-level classifications. Given the relative simplicity of raw scores, it is fair to question why scaled scores are used in NECAP reports instead of raw scores. Foremost, scaled scores offer the advantage of simplifying the reporting of results across content areas, grade levels, and subsequent years. Because the standard-setting process typically results in different cut scores across content areas on a raw score basis, it is useful to transform these raw cut scores to a scale that is more easily interpretable and consistent. For NECAP, a score of 40 is the cut score between the Partially Proficient and Proficient achievement levels. This is true regardless of the content area, grade, or year with which one may be concerned. If one were to use raw scores, the raw cut score between Substantially Below Proficient and Partially Proficient might, for example, be 35 in mathematics at grade 3, but 33 in mathematics at grade 7, or 36 in writing at grade 8. Using scaled scores greatly simplifies the task of understanding how a student performed.

Appendix B

CODE OF PROFESSIONAL RESPONSIBILITIES IN EDUCATIONAL MEASUREMENT



Prepared by the **NCME Ad Hoc Committee on the Development of a Code of Ethics:**
Cynthia B. Schmeiser, ACT – Chair
Kurt F. Geisinger, State University of New York
Sharon Johnson-Lewis, Detroit Public Schools
Edward D. Roeber, Council of Chief State School Officers
William D. Schafer, University of Maryland

©1995 National Council on Measurement in Education
Any portion of this Code may be reproduced and disseminated for educational purposes

CODE OF PROFESSIONAL RESPONSIBILITIES IN EDUCATIONAL MEASUREMENT

PREAMBLE AND GENERAL RESPONSIBILITIES

As an organization dedicated to the improvement of measurement and evaluation practice in education, the National Council on Measurement in Education (NCME) has adopted this Code to promote professionally responsible practice in conduct that arises from either the professional standards of the field, general ethical principles, or both.

The purpose of the Code of Professional Responsibilities in Educational Measurement, hereinafter referred to as the Code, is to guide the conduct of NCME members who are involved in any type of assessment activity in education. NCME is also providing this Code as a public service for all individuals who are engaged in educational assessment activities in the hope that these activities will be conducted in a professionally responsible manner. Persons who engage in these activities include local educators such as classroom teachers, principals, and superintendents; professionals such as school psychologists and counselors; state and national technical, legislative, and policy staff in education; staff of research, evaluation, and testing organizations; providers of test preparation services; college and university faculty and administrators; and professionals in business and industry who design and implement educational and training programs.

This Code applies to any type of assessment that occurs as part of the educational process, including formal and informal, traditional and alternative techniques for gathering information used in making educational decisions at all levels. These techniques include, but are not limited to, large-scale assessments at the school, district, state, national, and international levels; standardized tests; observational measures; teacher-conducted assessments; assessment support materials; and other achievement, aptitude, interest, and personality measures used in and for education.

Although NCME is promulgating this Code for its members, it strongly encourages other organizations and individuals who engage in educational assessment activities to endorse and abide by the responsibilities relevant to their professions. Because the Code pertains only to uses of assessment in education, it is recognized that uses of assessments outside of educational contexts, such as for employment, certification, or licensure, may involve additional professional responsibilities beyond those detailed in this Code.

The Code enumerates professional responsibilities in eight major areas of assessment activity. Specifically, the Code presents the professional responsibilities of those who:

1) Develop Assessments
2) Market and Sell Assessments
3) Select Assessments
4) Administer Assessments
5) Score Assessments
6) Interpret Use, and Communicate Assessment Results
7) Educate About Assessment
8) Evaluate Programs and Conduct Research on Assessments.

Although the organization of the Code is based on the differentiation of these activities, they are viewed as highly interrelated, and those who use this Code are urged to consider the Code in its entirety. The index following this Code provides a listing of some of the critical interest topics within educational measurement that focus on one or more of the assessment activities.

The professional responsibilities promulgated in this Code in eight major areas of assessment activity are based on expectations that NCME members involved in educational assessment will:

1) protect the health and safety of all examinees;

2) be knowledgeable about, and behave in compliance with, state and federal laws relevant to the conduct of professional activities;

3) maintain and improve their professional competence in educational assessment;

4) provide assessment services only in areas of their competence and experience, affording full disclosure of their professional qualifications;

5) promote the understanding of sound assessment practices in education;

6) adhere to the highest standards of conduct and promote professionally responsible conduct within educational institutions and agencies that provide educational services; and

7) perform all professional responsibilities with honesty, integrity, due care, and fairness.

Responsible professional practice includes being informed about and acting in accordance with the Code of *Fair Testing Practices in Education* (joint Committee on Testing Practices, 1988), the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, NCME, 1985), or subsequent revisions as well as all applicable state and federal laws that may govern the development, administration, and use of assessment. Both the *Standards for Educational and Psychological Testing* and the *Code of Fair Testing Practices in Education* are intended to establish criteria for judging the technical adequacy of tests and the appropriate uses of tests and test results. The purpose of this Code is to describe the professional responsibilities of those individuals who are engaged in assessment activities. As would be expected, there is a strong relationship between professionally responsible practice and sound educational assessments, and this Code is intended to be consistent with the relevant parts of both of these documents.

It is not the intention of NCME to enforce the professional responsibilities stated in the Code or to investigate allegations of violations to the Code. Since the Code provides a frame of reference for the evaluation of the appropriateness of behavior, NCME recognizes that the Code may be used in legal or other similar proceedings

Responsibilities of Those Who Develop Assessment Products and Services

SECTION 1

Those who develop assessment products and services, such as classroom teachers and other assessment specialists, have a professional responsibility to strive to produce assessments that are of the highest quality. Persons who develop assessments have a professional responsibility to:

1.1 ensure that assessment products and services are developed to meet applicable professional, technical, and legal standards.

1.2 develop assessment products and services that are as free as possible from bias due to characteristics irrelevant to the construct being measured, such as gender, ethnicity, race, socioeconomic status, disability, religion, age, or national origin.

1.3 plan accommodations for groups of test takers with disabilities and other special needs when developing assessments.

1.4 disclose to appropriate parties any actual or potential conflicts of interest that might influence the developers' judgment or performance.

1.5 use copyrighted materials in assessment products and services in accordance with state and federal law.

1.6 make information available to appropriate persons about the steps taken to develop and score the assessment, including up-to-date information used to support the reliability, validity, scoring and reporting processes, and other relevant characteristics of the assessment.

1.7 protect the rights to privacy of those who are assessed as part of the assessment development process.

1.8 caution users, in clear and prominent language, against the most likely misinterpretations and misuses of data that arise out of the assessment development process.

1.9 avoid false or unsubstantiated claims in test preparation and program support materials and services about an assessment or its use and interpretation.

1.10 correct any substantive inaccuracies in assessments or their support materials as soon as feasible.

1.11 develop score reports and support materials that promote the understanding of assessment results.

SECTION 2 — Responsibilities of Those Who Market and Sell Assessment Products and Services

The marketing of assessment products and services, such as tests and other instruments, scoring services test preparation services, consulting, and test interpretive services, should be based on information that is accurate, complete, and relevant to those considering their use. Persons who market and see assessment products and services have a professional responsibility to:

2.1 provide accurate information to potential purchasers about assessment products and services and their recommended uses and limitations.

2.2 not knowingly withhold relevant information about assessment products and services that might affect an appropriate selection decision.

2.3 base all claims about assessment products and services on valid interpretations of publicly available information.

2.4 allow qualified users equal opportunity to purchase assessment products and services.

2.5 establish reasonable fees for assessment products and services.

2.6 communicate to potential users, in advance of any purchase or use, all applicable fees associated with assessment products and services.

2.7 strive to ensure that no individuals are denied access to opportunities because of their inability to pay the fees for assessment products and services.

2.8 establish criteria for the sale of assessment products and services, such as limiting the sale of assessment products and services to those individuals who are qualified for recommended uses and from whom proper uses and interpretations are anticipated.

2.9 inform potential users of known inappropriate uses of assessment products and services and provide recommendations about how to avoid such misuses.

2.10 maintain a current understanding about assessment products and services and their appropriate uses in education.

2.11 release information implying endorsement by users of assessment products and services only with the users' permission.

2.12 avoid making claims that assessment products and services have been endorsed by another organization unless an official endorsement has been obtained.

2.13 avoid marketing test preparation products and services that may cause individuals to receive scores that misrepresent their actual levels of attainment.

SECTION 3 — Responsibilities of Those Who Select Assessment Products and Services

Those who select assessment products and services for use in educational settings, or help others do so, have important professional responsibilities to make sure that the assessments are appropriate for their intended use. Persons who select assessment products and services have a professional responsibility to:

3.1 conduct a thorough review and evaluation of available assessment strategies and instruments that might be valid for the intended uses.

3.2 recommend and/or select assessments based on publicly available documented evidence of their technical quality and utility rather than on unsubstantiated claims or statements.

3.3 disclose any associations or affiliations that they have with the authors, test publishers or others involved with the assessments under consideration for purchase and refrain from participation if such associations might affect the objectivity of the selection process.

3.4 inform decision makers and prospective users of the appropriateness of the assessment for the intended uses, likely consequences of use, protection of examinee rights, relative costs, materials, and services needed to conduct or use the assessment, and known limitations of the assessment, including potential misuses and misinterpretations of assessment information.

3.5 recommend against the use of any prospective assessment that is likely to be administered, scored, and used in an invalid manner for members of various groups in our society for reasons of race, ethnicity, gender, age, disability, language background, socioeconomic status, religion, or national origin.

3.6 comply with all security precautions that may accompany assessments being reviewed.

3.7 immediately disclose any attempts by others to exert undue influence on the assessment selection process.

3.8 avoid recommending, purchasing, or using test preparation products and services that may cause individuals to receive scores that misrepresent their actual levels of attainment.

Those who prepare individuals to take assessments and those who are directly or indirectly involved in the administration of assessments as part of the educational process, including teachers, administrators, and assessment personnel, have an important role in making sure that the assessments are administered in a fair and accurate manner. Persons who prepare others for and those who administer, assessments have a professional responsibility to:

4.1 inform the examinees about the assessment prior to its administration, including its purposes, uses; and consequences; how the assessment information will be judged or scored; how the results will be kept on file; who will have access to the results; how the results will be distributed; and examinees rights before, during, and after the assessment.

4.2 administer only those assessments for which they are qualified by education, training, licensure, or certification.

4.3 take appropriate security precautions before, during, and after the administration of the assessment.

4.4 understand the procedures needed to administer the assessment prior to administration.

4.5 administer standardized assessments according to prescribed procedures and conditions and notify appropriate persons if any nonstandard or delimiting conditions occur.

4.6 not exclude any eligible student from the assessment.

4.7 avoid any conditions in the conduct of the assessment that might invalidate the results.

4.8 provide for and document all reasonable and allowable accommodations for the administration of the assessment to persons with disabilities or special needs.

4.9 provide reasonable opportunities for individuals to ask questions about the assessment procedures or directions prior to and at prescribed times during the administration of the assessment.

4.10 protect the rights to privacy and due process of those who are assessed.

4.11 avoid actions or conditions that would permit or encourage individuals or groups to receive scores that misrepresent their actual levels of attainment.

Responsibilities of Those Who Score Assessments

SECTION 5

The scoring of educational assessments should be conducted properly and efficiently so that the results are reported accurately and in a timely manner. Persons who score and prepare reports of assessments have a professional responsibility to:

5.1 provide complete and accurate information to users about how the assessment is scored, such as the reporting schedule, scoring process to be used, rationale for the scoring approach, technical characteristics, quality control procedures, reporting formats, and the fees, if any, for these services.

5.2 ensure the accuracy of the assessment results by conducting reasonable quality control procedures before, during, and after scoring.

5.3 minimize the effect on scoring of factors irrelevant to the purposes of the assessment.

5.4 inform users promptly of any deviation in the planned scoring and reporting service or schedule and negotiate a solution with users.

5.5 provide corrected score results to the examinee or the client as quickly as practicable should errors be found that may affect the inferences made on the basis of the scores.

5.6 protect the confidentiality of information that identifies individuals as prescribed by state and federal law.

5.7 release summary results of the assessment only to those persons entitled to such information by state or federal law or those who are designated by the party contracting for the scoring services.

5.8 establish, where feasible, a fair and reasonable process for appeal and rescoring the assessment.

The interpretation, use, and communication of assessment results should promote valid inferences and minimize invalid ones. Persons who interpret, use, and communicate assessment results have a professional responsibility to:

6.1 conduct these activities in an informed objective, and fair manner within the context of the assessment's limitations and with an understanding of the potential consequences of use.

6.2 provide to those who receive assessment results information about the assessment, its purposes, its limitations, and its uses necessary for the proper interpretation of the results.

6.3 provide to those who receive score reports an understandable written description of all reported scores, including proper interpretations and likely misinterpretations.

6.4 communicate to appropriate audiences the results of the assessment in an understandable and timely manner, including proper interpretations and likely misinterpretations.

6.5 evaluate and communicate the adequacy and appropriateness of any norms or standards used in the interpretation of assessment results.

6.6 inform parties involved in the assessment process how assessment results may affect them.

6.7 use multiple sources and types of relevant information about persons or programs whenever possible in making educational decisions.

6.8 avoid making, and actively discourage others from making, inaccurate reports, unsubstantiated claims, inappropriate interpretations, or otherwise false and misleading statements about assessment results.

6.9 disclose to examinees and others whether and how long the results of the assessment will be kept on file, procedures for appeal and rescoring, rights examinees and others have to the assessment information, and how those rights may be exercised.

6.10 report any apparent misuses of assessment information to those responsible for the assessment process.

6.11 protect the rights to privacy of individuals and institutions involved in the assessment process.

SECTION 7 — Responsibilities of Those Who Educate Others about Assessment

The process of educating others about educational assessment, whether as part of higher education, professional development, public policy discussions, or job training, should prepare individuals to understand and engage in sound measurement practice and to become discerning users of tests and test results. Persons who educate or inform others about assessment have a professional responsibility to:

7.1 remain competent and current in the areas in which they teach and reflect that in their instruction.

7.2 provide fair and balanced perspectives when teaching about assessment.

7.3 differentiate clearly between expressions of opinion and substantiated knowledge when educating others about any specific assessment method, product, or service.

7.4 disclose any financial interests that might be perceived to influence the evaluation of a particular assessment product or service that is the subject of instruction.

7.5 avoid administering any assessment that is not part of the evaluation of student performance in a course if the administration of that assessment is likely to harm any student.

7.6 avoid using or reporting the results of any assessment that is not part of the evaluation of student performance in a course if the use or reporting of results is likely to harm any student.

7.7 protect all secure assessments and materials used in the instructional process.

7.8 model responsible assessment practice and help those receiving instruction to learn about their professional responsibilities in educational measurement.

7.9 provide fair and balanced perspectives on assessment issues being discussed by policymakers, parents and other citizens.

Conducting research on or about assessments or educational programs is a key activity in helping to improve the understanding and use of assessments and educational programs. Persons who engage in the evaluation of educational programs or conduct research on assessments have a professional responsibility to:

8.1 conduct evaluation and research activities in an informed, objective, and fair manner.

8.2 disclose any associations that they have with authors, test publishers, or others involved with the assessment and refrain from participation if such associations might affect the objectivity of the research or evaluation.

8.3 preserve the security of all assessments throughout the research process as appropriate.

8.4 take appropriate steps to minimize potential sources of invalidity in the research and disclose known factors that may bias the results of the study.

8.5 present the results of research, both intended and unintended, in a fair, complete, and objective manner.

8.6 attribute completely and appropriately the work and ideas of others.

8.7 qualify the conclusions of the research within the limitations of the study.

8.8 use multiple sources of relevant information in conducting evaluation and research activities whenever possible.

8.9 comply with applicable standards for protecting the rights of participants in an evaluation or research study, including the rights to privacy and informed consent.

Afterword

As stated at the outset, the purpose of the *Code of Professional Responsibilities in Educational Measurement* is to serve as a guide to the conduct of NCME members who are engaged in any type of assessment activity in education. Given the broad scope of the field of educational assessment as well as the variety of activities in which professionals may engage, it is unlikely that any code will cover the professional responsibilities involved in every situation or activity in which assessment is used in education. Ultimately, it is hoped that this Code will serve as the basis for ongoing discussions about what constitutes professionally responsible practice. Moreover, these discussions will undoubtedly identify areas of practice that need further analysis and clarification in subsequent editions of the Code. To the extent that these discussions occur, the Code will have served its purpose.

To assist in the ongoing refinement of the Code, comments on this document are most welcome. Please send your comments and inquiries to:

Dr. William J. Russell
Executive Officer
National Council on
Measurement in Education
1230 Seventeenth Street, NW
Washington, DC 20036-3078

The following list of resources is provided for those who want to seek additional information about codes of professional responsibility that have been developed and adopted by organizations having an interest in various aspects of educational assessment.

American Association for Counseling and Development (now American Counseling Association). (1988). Ethical *standards of the American Counseling Association.* Alexandria, VA: Author.

American Association for Counseling and Development (now American Counseling Association) & Association for Measurement and Evaluation in Counseling and Development (now Association for Assessment in Counseling). (1989) *Responsibilities of users of standardized tests;* RUST *statement revised.* Alexandria, VA: Author.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC: Author.

American Educational Research Association. (1992). Ethical standards of the American Educational Research association. *Educational Researcher,* 21 (7), 23-26.

American Federation of Teachers, National Council on Measurement in Education, & National Education association. (1990). Standards *for teacher competence in educational assessment of students.* Washington, DC: Author.

American Psychological Association. (1992). *Ethical principles of psychologists and code of conduct.* Washington, DC: Author.

American Psychological Association President's Task Force on Psychology in Education. (In press). *Learner-centered psychological principles: Guidelines for school redesign and reform.* Washington, DC: Author.

Joint Advisory Committee. (1993). *Principles for fair assessment practices for education in Canada.* Edmonton, Alberta: Author.

Joint Committee on Testing Practices. (1988). Code *of fair testing practices in education.* Washington, DC: Author.

Joint Committee on Standards for Educational Evaluation. (1988). The *personnel evaluation standards: How to assess systems for evaluating educators.* Newbury Park, CA: Sage.

Joint Committee on Standards for Educational Evaluation. (1 The *program evaluation standards: How to assess evaluatio educational programs.* Thousand Oaks, CA: Sage.

National Association of College Admission Counselors. (1988). Statement *of principles of good practice.* Alexandria, VA: Author.

Index to the Code of Professional Responsibilities in Educational Measurement

This index provides a list of major topics and issues addressed by the responsibilities in each of the eight sections of the Code. Although this list is not intended to be exhaustive, it is intended to serve as a reference source for those who use this Code.

Topic **Responsibility**

Advertising.....1.9, 1.10, 2.3, 2.11, 2.12

Bias……1.2, 3.5, 4.5, 4.7, 5.3, 8.4

Cheating……... 4.5, 4.6, 4.11

Coaching and Test Preparation…….. 2.13, 3.8, 4.11

Competence…… 2.10, 4.2, 4.4, 4.5, 5.2, 5.5, 7.1, 7.8, 7.9, 8.1, 8.7

Conflict of Interest…….. 1.4, 3.3, 7.4, 8.2

Consequences of Test Use……... 3.4, 6.1, 6.6, 7.5, 7.6

Copyrighted Materials, Use of…….. 1.5, 8.6

Disabled Examinees, Rights of …….1.3, 4.8

Disclosure……1.6, 2.1, 2.2, 2.6, 3.3, 3.7, 4.1, 5.1, 5.4, 6.2, 6.3, 6.4, 6.6, 6.9, 8.2, 8.4, 8.5

Topic **Responsibility**

Due Process…… 4.10, 5.8, 6.9

Equity ……..1.2, 2.4, 2.7, 3.5, 4.6

Fees …….. 2.5, 2.6, 2.7

Inappropriate Test Use ……..1.8, 2.8, 2.9, 3.4, 6.8, 6.10

Objectivity…… 3.1, 3.2, 3.3, 6.1, 6.5, 7.2, 7.3 7.9, 8.1, 8.2, 8.5, 8.7

Rights to Privacy ……1.7, 3.4, 4.10, 5.6, 5.7, 6.11, 8.9

Security ……..3.6, 4.3, 7.7, 8.3

Truthfulness ……1.10, 2.1, 2.2, 2.3, 2.11, 2.12, 3.2, 4.6, 7.3

Undue Influence ……..3.7

Unsubstantiated Claims …….1.9, 3.2, 6.8

Appendix C

NECAP Achievement Level Descriptions

General Achievement Level Descriptions	
Proficient with Distinction (Level 4)	Students performing at this level demonstrate the prerequisite knowledge and skills needed to participate and excel in instructional activities aligned with the GLE at the current grade level. Errors made by these students are few and minor and do not reflect gaps in prerequisite knowledge and skills.
Proficient (Level 3)	Students performing at this level demonstrate minor gaps in the prerequisite knowledge and skills needed to participate and perform successfully in instructional activities aligned with the GLE at the current grade level. It is likely that any gaps in prerequisite knowledge and skills demonstrated by these students can be addressed during the course of typical classroom instruction.
Partially Proficient (Level 2)	Students performing at this level demonstrate gaps in prerequisite knowledge and skills needed to participate and perform successfully in instructional activities aligned with the GLE at the current grade level. Additional instructional support may be necessary for these students to meet grade level expectations.
Substantially Below Proficient (Level 1)	Students performing at this level demonstrate extensive and significant gaps in prerequisite knowledge and skills needed to participate and perform successfully in instructional activities aligned with the GLE at the current grade level. Additional instructional support is necessary for these students to meet grade level expectations.

Reading Achievement Level Descriptions	
Proficient with Distinction	Student's performance demonstrates an ability to read and comprehend grade-appropriate text. Student is able to analyze and interpret literary and informational text. Student offers insightful observations/assertions that are well supported by references to the text. Student uses range of vocabulary strategies and breadth of vocabulary knowledge to read and comprehend a wide variety of texts.
Proficient	Student's performance demonstrates an ability to read and comprehend grade-appropriate text. Student is able to analyze and interpret literary and informational text. Student makes and supports relevant assertions by referencing text. Student uses vocabulary strategies and breadth of vocabulary knowledge to read and comprehend text.
Partially Proficient	Student's performance demonstrates an inconsistent ability to read and comprehend grade-appropriate text. Student attempts to analyze and interpret literary and informational text. Student may make and/or support assertions by referencing text. Student's vocabulary knowledge and use of strategies may be limited and may impact the ability to read and comprehend text.
Substantially Below Proficient	Student's performance demonstrates minimal ability to derive/construct meaning from grade-appropriate text. Student may be able to recognize story elements and text features. Student's limited vocabulary knowledge and use of strategies impacts the ability to read and comprehend text.

Mathematics Achievement Level Descriptions	
Proficient with Distinction	Student's problem solving demonstrates logical reasoning with strong explanations that include both words and proper mathematical notation. Student's work exhibits a high level of accuracy, effective use of a variety of strategies, and an understanding of mathematical concepts within and across grade level expectations. Student demonstrates the ability to move from concrete to abstract representations.
Proficient	Student's problem solving demonstrates logical reasoning with appropriate explanations that include both words and proper mathematical notation. Student uses a variety of strategies that are often systematic. Computational errors do not interfere with communicating understanding. Student demonstrates conceptual understanding of most aspects of the grade level expectations.
Partially Proficient	Student's problem solving demonstrates logical reasoning and conceptual understanding in some, but not all, aspects of the grade level expectations. Many problems are started correctly, but computational errors may get in the way of completing some aspects of the problem. Student uses some effective strategies. Student's work demonstrates that he or she is generally stronger with concrete than abstract situations.
Substantially Below Proficient	Student's problem solving is often incomplete, lacks logical reasoning and accuracy, and shows little conceptual understanding in most aspects of the grade level expectations. Student is able to start some problems but computational errors and lack of conceptual understanding interfere with solving problems successfully.

Writing Achievement Level Descriptions	
Proficient with Distinction	Student's writing demonstrates an ability to respond to prompt/task with clarity and insight. Focus is well developed and maintained throughout response. Response demonstrates use of strong organizational structures. A variety of elaboration strategies is evident. Sentence structures and language choices are varied and used effectively. Response demonstrates control of conventions; minor errors may occur.
Proficient	Student's writing demonstrates an ability to respond to prompt/task. Focus is clear and maintained throughout the response. Response is organized with a beginning, middle and end with appropriate transitions. Details are sufficiently elaborated to support focus. Sentence structures and language use are varied. Response demonstrates control of conventions; errors may occur but do not interfere with meaning.
Partially Proficient	Student's writing demonstrates an attempt to respond to prompt/task. Focus may be present but not maintained. Organizational structure is inconsistent with limited use of transitions. Details may be listed and lack elaboration. Sentence structures and language use are unsophisticated and may be repetitive. Response demonstrates inconsistent control of conventions.
Substantially Below Proficient	Student's writing demonstrates a minimal response to prompt/task. Focus is unclear or lacking. Little or no organizational structure is evident. Details are minimal and/or random. Sentence structures and language use are minimal or absent. Frequent errors in conventions may interfere with meaning.