The New England Common Assessment Program

Guide to Using the 2009 NECAP Science 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, tkurtz@ed.state.nh.us, 101 Pleasant Street, Concord, NH 03301, www.ed.state.nh.us

Rhode Island Department of Elementary and Secondary Education: Kevon Tucker-Seeley, Assessment Specialist, 401-222-8494, Kevon.Tucker-Seeley@ride.ri.gov, 255 Westminster Street, Providence, RI 02903, www.ride.ri.gov

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

Table of Contents

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 & 11 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 May 2009 administration of the New England Common Assessment Program (NECAP) science tests. This document describes and explains the information included in the following NECAP reports:

- NECAP Tests of Spring 2009: *NECAP Student Report*
- NECAP Tests of Spring 2009: *NECAP Item Analysis Report*
- NECAP Tests of Spring 2009: *NECAP School/District Results Report*
- NECAP Tests of Spring 2009: *NECAP District Summary Report*
- NECAP Tests of Spring 2009: *NECAP District Student-Level Data Files*

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 science 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.

Accessing Reports

NH	School-, district-, and state-level NECAP science reports can be accessed through the NHDOE website homepage: (http://www.ed.state.nh.us/education/) and clicking on the link to the NH School District Profile site. *NECAP Item Analysis Reports* and student-level data files can be accessed using the following URL: http://iservices.measuredprogress.org. Principals and superintendents are able to access the confidential reports and files by selecting New England Common Assessment Program (NECAP) from the drop-down menu, clicking on the NECAP Reporting link, and entering their secure username and password.

RI	All NECAP science reports and data files (confidential and non-confidential) can be accessed using the following URL: http://iservices.measuredprogress.org Principals and superintendents are able to access the reports and files by selecting New England Common Assessment Program (NECAP) from the drop-down menu, clicking on the NECAP Reporting link, and entering their secure username and password. School-, district-, and state-level NECAP science reports can also be accessed through the RIDE website at http://www.ride.ri.gov/assessment/Results.aspx or by clicking on the State Testing and Reporting link on the RIDE homepage.
VT	State- and school-level NECAP science results, as well as results from other assessments, can be accessed on the VT DOE website using the following URL: (http://education.vermont.gov/new/html/pgm_assessment/data.html). To obtain copies of other NECAP science reports and support materials, including grade-level results, contact the local school administrator.

General Guidelines for the Use of NECAP Reports

Alignment of Curriculum and the NECAP Tests

All test items appearing on the NECAP science tests are designed to measure specific *NECAP Science Assessment Targets*. 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 science results are intended to evaluate how well students and schools are achieving the learning targets contained in the *NECAP Science Assessment Targets*. 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 in isolation 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. Schools will receive two copies of the *NECAP Student Report.* The colored copy of the report is for distribution to parents/guardians and the black-and-white copy of the report is for school files. 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.

Details about the NECAP science tests and achievement levels are provided on the cover of the *NECAP Student Report*. Details about the student's performance on the NECAP science 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 science. Achievement level descriptions can be found in Appendix C of this guide and are provided on the reverse side of the report. The *NECAP Student Report* for grades 4, 8, and 11 shows the scaled score earned for science. The scaled score is reported with a score band that indicates the standard error of measurement surrounding the score. The standard error of measurement 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).

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 science. 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 Science Domains

This section of the report shows the student's performance compared to school, district, and statewide performance in a variety of areas. The science areas assessed by NECAP are reported by domains: Physical Science, Earth Space Science, Life Science, and Scientific Inquiry.

Student performance in the science domains 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 science test.

Description of the Inquiry Task

This section of the report includes text that describes inquiry as it relates to the performance task that is included as a part of the NECAP science test. The following paragraph is included on this section of the report:

There are many interesting and essential facts, formulas, and processes that students should know across the content domains of science. But science is more than content. Inquiry skills are skills that all students should have in addition to the content. Inquiry skills are the ability to formulate questions and hypothesize, plan investigations and experiments, conduct their own investigations and experiments, and evaluate their results. These are the broad areas that encompass scientific inquiry. The NECAP Science Inquiry Tasks use content from Physical Science, Earth Space Science, and Life Science as the basis of the task. Student Knowledge of the content is not measured in the inquiry tasks but rather the student's ability to make connections, express ideas, and provide evidence of scientific thinking.

A grade specific paragraph is also included describing the inquiry task that students completed for that grade.

The following two pages contain a sample grade 8 *NECAP Student Report*.

NECAP Student Report - Spring 2009

This report contains results from the Spring 2009 New England Common Assessment Program (NECAP) science 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 standards developed and adopted by the three states. Specifically, the tests are designed to measure the content and skills that students are expected to have at the end of the K–4, 5–8, and 9–11 grade spans.

NECAP science test results are used primarily for program evaluation, school improvement, and public reporting. 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 tested grade. Performance at Proficient or Proficient with Distinction indicates that the student has a level of proficiency necessary to begin working successfully on higher grade content and skills. Performance below Proficient suggests that additional instruction and student work may be needed as the student is introduced to new content and skills at the next 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 440 at grade 4, 840 at grade 8, and 1140 at grade 11 each indicate Proficient performance at that grade level.

Comparisons to Other End of Grade Span 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 state. 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 Science Domains

This section of the report provides information about student performance on sets of items measuring four science domains within the 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 fewer 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.

Achievement Level Descriptions

Proficient with Distinction (Level 4) - Students performing at this level demonstrate the knowledge and skills as described in the content standards for this grade span. Errors made by these students are few and minor and do not reflect gaps in knowledge and skills.

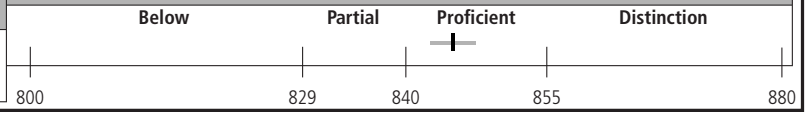
Proficient (Level 3) - Students performing at this level demonstrate the knowledge and skills as described in the content standards for this grade span with only minor gaps. It is likely that any gaps in knowledge and skills demonstrated by these students can be addressed by the classroom teacher during the course of classroom instruction.

Partially Proficient (Level 2) - Students performing at this level demonstrate gaps in knowledge and skills as described in the content standards for this grade span. Additional instructional support may be necessary for these students to achieve proficiency on the content standards.

Substantially Below Proficient (Level 1) - Students performing at this level demonstrate extensive and significant gaps in knowledge and skills as described in the content standards for this grade span. Additional instructional support is necessary for these students to achieve proficiency on the content standards.

Student	Grade	School	District	State
Nicholas G Anderson	8	Demonstration School 1	Demonstration District A	NH

Spring 2009 - Grade 8 NECAP Science Test Results

Achievement Level	Scaled Score	This Student's Achievement Level and Score
Proficient	845	

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 End of Grade 8 Students by School, District, and State

	Student	School	District	State
Proficient with Distinction		<1%	<1%	<1%
Proficient	✓	24%	25%	24%
Partially Proficient		51%	52%	53%
Substantially Below Proficient		24%	23%	23%

This Student's Performance in Science Domains

	Possible Points	Student	Average Points Earned			
			School	District	State	Students at Proficient Level
Physical Science	15	13	7.2	7.3	7.4	8-11.9
Earth Space Science	15	8	7.9	8	8.1	8.3-11.8
Life Science	15	14	7.8	7.9	7.9	8.4-12
Inquiry	18	10	6.7	6.8	6.7	7.2-10.4

Description of the Inquiry Task

There are many interesting and essential facts, formulas, and processes that students should know across the three content domains of science. But science is more than content. Inquiry skills are skills that all students should have in addition to the content. Inquiry skills are the ability to formulate questions and hypothesize, plan investigations and experiments, conduct their own investigations and experiments, and evaluate their results. These are the broad areas that encompass scientific inquiry. The NECAP Science Inquiry Tasks use content from Physical Science, Earth Space Science, and Life Science as the basis of the task. Student knowledge of the content is not measured in the inquiry tasks but rather the student's ability to make connections, express ideas, and provide evidence of scientific thinking.

The grade 8 inquiry task, *Pond Weeds*, had students explore the effect weevils have on the growth of Eurasian Water Milfoil. Students were provided a scenario and authentic data and asked to investigate the relationship among weevils, Eurasian Water Milfoil, and sunfish in control and experimental sites. Students worked independently during this task.

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 science. 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 legend after the last page of data that defines the terms used.

The data used for the *NECAP Item Analysis Report* are the results of the spring 2009 administration of the NECAP science tests. The NECAP science tests are based on the *NECAP Science Assessment Targets* from three grades spans (K–4, 5–8, and 9–11). For example, the Grade 8 NECAP science test, administered in the spring of eighth grade, is based on the grades 5–8 *NECAP Science Assessment Targets*. Every student who participated in the NECAP science tests will be represented in a "testing year" school report, because NECAP science testing takes place near the end of the school year, there are no "teaching year" school reports for science.

For 2009, the state Departments of Education chose to release the inquiry tasks as stand-alone documents so schools could more easily conduct the tasks in the classroom throughout the year. As a result, the top portion of the *NECAP Item Analysis Report* was split to reflect both the Released Items and the Released Inquiry Task.

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

- The first row lists the **Item Number** (For the Released Items, this is not the position of the item in the actual student test booklet, but for the Released Inquiry Task, it is.)
- The second row lists the **Science Domain** for the item.
- The third row lists the **Assessment Target** or **Inquiry Construct** 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 section 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 followed by each student's state assigned student ID number. The column after the released items shows Total Test Results, broken into several categories. Domain Points Earned columns report the points the student earned in each science domain. The Total Points Earned column is a summary of all of the points earned on the science test. 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.

The *NECAP Item Analysis Reports* are 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 grade 4.

Spring 2009 - Grade 4 NECAP Tests
Grade 4 Students in 2008-2009
Item Analysis Report
Science

School: Demonstration School 2
District: Demonstration District A
State: Rhode Island
Code: DEMOA-DEMO2

		Released Items										Released Inquiry Task								Total Test Results						
Item Number		1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	Domain Points Earned				Total Points Earned	Scaled Score	Achievement Level
Science Domain		PS	PS	PS	ESS	ESS	ESS	LS	LS	LS	LS	INQ	INQ	INQ	INQ	INQ	INQ	INQ	INQ	Physical Science	Earth Space Science	Life Science	Inquiry			
Assessment Target/Inquiry Construct		1-2	1-3	3-8	1-2	1-3	1-5	1-3	3-7	4-8	2-6	8	7	10	12	6	1	12	13							
Depth of Knowledge Code		1	2	2	2	2	2	1	2	2	2	2	1	2	2	2	3	3	3							
Item Type		MC	MC	MC	MC	MC	MC	MC	MC	MC	CR	CR	SA	SA	SA	SA	CR	SA	SA							
Correct MC Response		D	B	C	D	D	B	B	D	A																
Name/Student ID	Total Possible Points	1	1	1	1	1	1	1	1	1	4	3	2	2	2	2	3	2	2	**15**	**15**	**15**	**18**	**63**		
Thacker, Kevin D	F057091000	B	A	+	+	B	A	+	+	B	1	1	1	0	0	1	1	0	1	5	7	9	5	26	428	2
Thibeault, Cody J	E903751000	A	D	+	+	+	D	C	C	+	2	1	1	1	0	0	2	0	1	4	7	8	6	25	426	1
Thompson, Anna	F790401000	+	A	A	+	+	+	D	+	+	3	3	2	1	1	0	3	1	1	9	11	9	12	41	442	3
Thornton, Traci A	F000481000	A	C	A	+	+	C	+	+	+	1	0	2	0	0	1	1	0	1	0	9	6	5	20	421	1
Trimble, Cody J	F349361000	+	+	+	+	A	+	D	+	+	3	3	1	1	0	1	1	1	1	12	13	9	9	43	444	3
True, Haley	F028701000	C	+	A	B	A	D	C	+	+	2	3	2	1	1	0	0	0	1	5	7	8	8	28	430	2
Vansteemburg, Emily	F773071000	+	A	+	+	+	D	+	+	+	1	2	2	0		1	1	0	1	7	9	7	7	30	431	2
Vettese, Mariah K	E831241000	+	+	+	+	+	+	+	C	+	3	2	2	2	1	1	2	1	1	12	13	12	12	49	451	3
Vogel, Jameson E	E971101000	A	+	+	+	C	A	A	C	+	2	0	2	1	1	0	1	1	1	7	8	8	7	30	431	2
Whitmore, Crysta	E917911000	+	+	+	+	+	+	+	+	+	2	2	1	1	1	1	2	0	2	10	11	10	10	41	442	3
Williams, Colton A	F103291000	+	+	+	+	C	+	+	+	+	2	2	2	1	1	0	1	1	1	11	8	11	9	39	439	2
Wilson, Connor A	D957581000	A	A	D	+	B	D	D	+	+	1	0	1	0	0	0	1	0	1	7	7	7	3	24	426	1
Woodward, Mark C	E949531000	+	A	+	+	C	+	+	+	+	2	2	2	1	2	1	1	1	1	13	12	11	11	47	448	3
Young, Claire J	E964881000	A	+	A	+	+	A	+	C	+	2	1	2	0	1	1	2	1	0	9	13	10	8	40	441	3
Item Number		1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8							
Percent Correct/Average Score: School		61	71	47	84	59	52	48	76	83	2.0	1.6	1.6	0.8	0.7	0.6	1.3	0.4	0.9	8.7	9.6	8.9	7.8			
Percent Correct/Average Score: District		66	75	52	85	60	55	53	72	80	2.0	1.7	1.6	0.8	0.7	0.6	1.3	0.4	0.9	8.8	9.8	9.1	8.1			
Percent Correct/Average Score: State		61	74	53	82	62	56	51	70	79	2.0	1.7	1.6	0.9	0.7	0.6	1.3	0.5	0.9	8.6	9.8	8.9	8.1			

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 science testing of local students in grades 4, 8 & 11. 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. 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 district and school 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 guide), and represent the aggregate of individual student scores (achievement level results and scaled scores).

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 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 the *NECAP Science Assessment Targets* 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 tables on the following pages can be used to assist 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, and
- comparing the scores of a district to the state.

These tables provide figures that can be used to make approximate comparisons between scores. Similar to the score band provided in the *NECAP Student Report*, the figures in the tables are estimates of one standard error around the score or difference between scores. For those interested in making more exact comparisons or learning more about the statistical methods used to make comparisons, a list of references is provided in Appendix D *Reference Materials* on page 44 of this guide.

Caution should be used when making any of the comparisons listed above 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 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.

SCALED SCORES

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. NECAP scaled scores for grade 11 are represented by a 4-digit number, with the first two digits representing the grade; the remaining digits also range from 00-80.

The table on the following page shows the smallest differences in scaled scores that represent a statistically significant difference in performance based on the number of students tested in the school and/or district. When comparing the scores of two groups of different sizes, one should use a difference that is approximately the average of the minimally statistically significant difference of each group. For example, when comparing the average grade 8 science scaled scores of a school with 25 students and a school with 100 students one should use two points as the minimally statistically significant difference. Two points is the average of the values in the table for a school of 25 students (3 points) and a school of 100 students (1 point). If the difference in scaled scores between the two groups is at least two points, then the difference is statistically significant. If the difference in scaled scores between the two groups is fewer than two points, the difference is not statistically significant.

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

Grade	Number of Students Tested in Group (Class, School etc.)				
	10	25	50	100	200
4	5	3	2	2	1
8	4	3	2	1	1
11	4	3	2	1	1

*Standard error of the mean difference assuming equal number of students and standard deviation

ACHIEVEMENT LEVELS

Comparisons of group performance can also be made by comparing the percentages of students scoring at or above a particular achievement level. But again, small differences in percentages should not be over-interpreted. Because, unlike scaled scores, achievement level results are reported as percentages, a slightly different procedure is used to make comparisons between the performance of two groups or between a group and a fixed point. To compare percentages, an interval estimation approach similar to a margin of error or the score band reported on the *NECAP Student Report* can be used.

With percentages, the statistical significance of differences is impacted by both the size of the group and the percentage of students in the category of interest (for example, Proficient or above on the Grade 4 science test). The table on the following page shows the size of the confidence interval that should be drawn around a score for selected percentages and school sizes. For example, if 60% of the students in a school of 50 students are Proficient or above, a confidence interval of ±7 percentage points, from 53% to 67%, would be drawn around the score of 60%. If the school's performance were being compared to a fixed percentage of 65% of students Proficient or above, the conclusion would be that the school score was not significantly different because the 53%-67% confidence interval includes 65%.

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

Percentages of Students in Achievement Level(s)	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

*Standard error of a percentage

The previous example compared the performance of a relatively small school to a fixed point (for example, a very large group such as the state). When two relatively small groups are compared, a confidence interval should be drawn around each score using the appropriate values from the table based on the size and performance of each group. If the two confidence intervals do not overlap, then the conclusion is that the difference between the two groups is statistically significant. If the two confidence intervals do overlap, then the difference in performance between the two groups is too small to be considered statistically significant. The distance between the two confidence intervals or their degree of overlap also provides a visual indication of the probability that the two scores are significantly different.

SCIENCE DOMAIN SUBSCORES

Science domain subscores cannot be directly compared from one year to the next even within a grade level. Unlike achievement levels and scaled scores, these scores are reported as raw scores and have not been linked across years and placed on the same scale. Differences in subscores from one year to the next in the total number of points earned by a student or in the percent of total possible points earned by a school or district may simply reflect either a small difference in the number of possible points in the reporting category or a slight difference in the difficulty of items within a particular reporting category. The process of equating that accounts for these differences to produce scaled scores and achievement levels for the total content area is not applied to individual reporting categories. There is not a sufficient number of points within each reporting category to equate these subscores from one year to the next.

There are, however, comparisons that can be made with science domain subscores to assist schools in the evaluation of their curricula and instructional programs. For each science domain subscore, normative information is provided describing performance in comparison to the school, district, state, and at the student level, students scoring at the Proficient threshold. Across years, this information can be used to determine whether progress has been made relative to one of the comparison groups. Even more than with scaled scores and achievement levels, it is important not to over-interpret small changes from one year to the next.

ACHIEVEMENT LEVEL CUT SCORES

The table below shows the scaled scores that identify the cut point between the four achievement levels: Proficient with Distinction, Proficient, Partially Proficient, and Substantially Below Proficient. The achievement level cut scores for science were the result of the standard setting process that was completed in August 2008 and will remain consistent year to year.

Achievement Level Cut Scores

Grade	Subject	SBP/PP*	PP/P*	P/PD*
4	Science	426 / 427	439 / 440	462 / 463
8	Science	828 / 829	839 / 840	854 / 855
11	Science	1129 / 1130	1139 / 1140	1151 / 1152

*SBP = Substantially Below Proficient
PP = Partially Proficient
P = Proficient
PD = Proficient with Distinction

Grade Level Summary Report

(Page 2 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 science tests in spring 2009. Students enrolled in a school on or after May 11, 2009 were expected to complete the NECAP science tests at that school.

STUDENTS ENROLLED ON OR AFTER MAY 11

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 science test, 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 sessions of the NECAP science test, their actual score was reported. 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 NECAP science test.

- Withdrew After May 11—Students withdrawing from a school after May 11, 2009 may have taken some sessions of the NECAP science test prior to their withdrawal from the school.
- Enrolled After May 11—Students enrolling in a school after May 11, 2009 may not have had adequate time to fully participate in all sessions of the NECAP science test.
- Special Consideration—Schools received state approval for special consideration for an exemption for all or part of the NECAP science test for any student whose circumstances are not described by the previous categories, but for whom the school determined that taking the NECAP science test would not be possible.
- Other—Occasionally students will not have completed the NECAP science test for reasons other than those listed above. These "other" categories are considered "not state approved".

Note: First Year LEP students who were new to the U.S. after October 1, 2008 were required to take the NECAP science test.

NECAP Results

The results portion of the page indicates the number and percentage of students performing at each achievement level in science tested by NECAP. In addition, a mean scaled score is provided for science in grades 4, 8 & 11 at the school, district, and state levels.

The following page contains a sample of a grade 4 "Grade Level Summary Report" from a *NECAP School Results Report.*

Spring 2009 - Grade 4 NECAP Science Test

Grade Level Summary Report

School: Demonstration School 1
District: Demonstration District A
State: Vermont
Code: DEMOA-DEMO1

Schools and districts administered all NECAP tests to every enrolled student with the following exceptions: students who participated in the alternate assessment for the 2008-09 school year, students who withdrew from the school after May 11, 2009, students who enrolled in the school after May 11, 2009, 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 May 11	**168**	**290**	**6,429**	**100**	**100**	**100**
	Science			Science		
Students tested	**167**	**289**	**6,395**	**99**	**100**	**99**
Students not tested in NECAP						
State Approved	0	0	0	0	0	0
Alternate Assessment	0	0	0	0	0	0
Withdrew After May 11	0	0	0	0	0	0
Enrolled After May 11	0	0	0	0	0	0
Special Consideration	0	0	0	0	0	0
Other	1	1	34	1	0	1

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	%	%	%	%	
SCIENCE	168	0	1	167	1	1	86	51	68	41	12	7	440	289	1	49	42	8	440	6,395	1	51	38	10	440

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.

Science Results

(Page 3 of the *NECAP School Results Report*)

The science results page of the report provides information on performance in specific domains of science (for example, Earth Space Science). The purpose of this section 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 *NECAP Science Assessment Targets*.

Information about science for school, district and state includes:

- the total number of students Enrolled, NT Approved (not tested for a state-approved reason), NT Other (not tested for other reasons), 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 information listed above is provided for both the current testing year (2008-09) in bold as well as the previous testing year (2007-08). This page of the report also includes a location for scores for 2009-10 so that next year a cumulative total over the three years can be reported. This information is only included for each year where the number of students tested at a grade level was at least 10.

For this year, because there are only two years of data available, these scores are combined into a "Cumulative Total" row. The two years of data are summed for the Enrolled, Not Tested Approved, Not Tested Other, and Tested columns. For the achievement levels, the two years of data in the "N" columns are summed while the percentages of students are calculated by dividing the cumulative total of the number of students in the achievement level by the cumulative total number of students tested. The Mean Scaled Score is calculated by summing the product of the mean scaled score and tested N for each year, and dividing the sum by the tested N from the cumulative total row (weighted average).

Information about each science domain includes:

- The **Total Possible Points** for that domain. In order to provide as much information as possible for each domain, the total number of points includes both the common items used to calculate scores as well as additional items in each domain used for equating the test from year to year.
- A graphic display of the **Percent of Total Possible Points** for the school, district, and state. 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 grade 4 "Science Results" page from a *NECAP School Results Report.*

Spring 2009 - Grade 4 NECAP Science Test

Science Results

School: Demonstration School 1
District: Demonstration District A
State:
Code: DEMOA-DEMO1

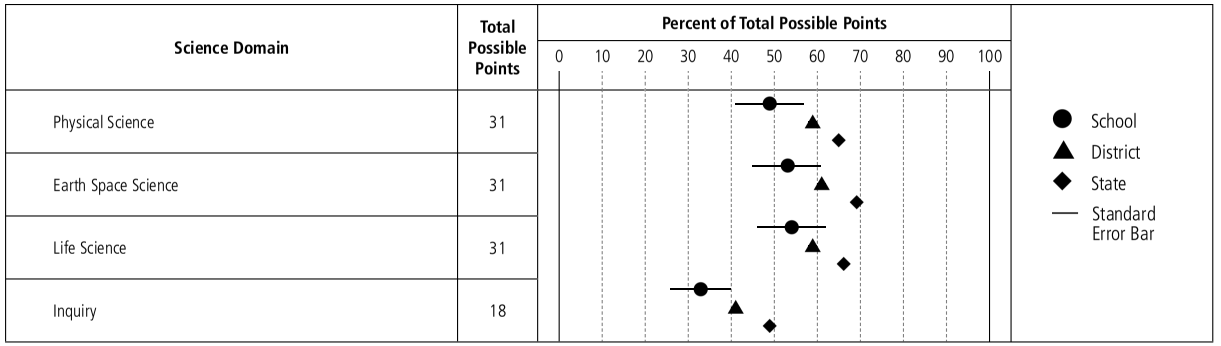
Proficient with Distinction (Level 4)
Students performing at this level demonstrate the knowledge and skills as described in the content standards for this grade span. Errors made by these students are few and minor and do not reflect gaps in knowledge and skills.

Proficient (Level 3)
Students performing at this level demonstrate the knowledge and skills as described in the content standards for this grade span with only minor gaps. It is likely that any gaps in knowledge and skills demonstrated by these students can be addressed by the classroom teacher during the course of classroom instruction.

Partially Proficient (Level 2)
Students performing at this level demonstrate gaps in knowledge and skills as described in the content standards for this grade span. Additional instructional support may be necessary for these students to achieve proficiency on the content standards.

Substantially Below Proficient (Level 1)
Students performing at this level demonstrate extensive and significant gaps in knowledge and skills as described in the content standards for this grade span. Additional instructional support is necessary for these students to achieve proficiency on the content standards.

	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													
2007-08	71	2	0	69	0	0	21	30	31	45	17	25	434
2008-09	**44**	**5**	**0**	**39**	**0**	**0**	**5**	**13**	**21**	**54**	**13**	**33**	**430**
2009-10													
Cumulative Total	115	7	0	108	0	0	26	24	52	48	30	28	433
DISTRICT													
2007-08	1,128	9	1	1,118	5	<1	304	27	557	50	252	23	434
2008-09	**1,064**	**24**	**2**	**1,038**	**1**	**<1**	**339**	**33**	**494**	**48**	**204**	**20**	**435**
2009-10													
Cumulative Total	2,192	33	3	2,156	6	<1	643	30	1,051	49	456	21	434
STATE													
2007-08	14,967	117	20	14,830	254	2	7,325	49	5,877	40	1,374	9	440
2008-09	**14,641**	**172**	**27**	**14,442**	**60**	**<1**	**7,614**	**53**	**5,537**	**38**	**1,231**	**9**	**441**
2009-10													
Cumulative Total	29,608	289	47	29,272	314	1	14,939	51	11,414	39	2,605	9	440



Disaggregated Science Results

(Page 4 of the *NECAP School Results Report*)

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 science results page of the *NECAP School Results Report*. This section presents the relationship between the variables reported and performance in science at the school, district, and state levels. The table shows the number of students categorized as Enrolled, NT Approved (not tested for a state-approved reason), NT Other (not tested for other reasons), and Tested. The table also provides 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 Science Results" page from a *NECAP School Results Report.* Please note that for NH and VT no data appears for 504 Plan or Title I. NH will have Title I data for science in future years.

Spring 2009 - Grade 4 NECAP Science Test

Disaggregated Science Results

School: Demonstration School 1
District: Demonstration District A
State: Rhode Island
Code: DEMOA-DEMO1

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	%	%	%	%	
All Students	**236**	**7**	**5**	**224**	**1**	**<1**	**92**	**41**	**94**	**42**	**37**	**17**	**438**	**388**	**1**	**40**	**40**	**19**	**437**	**9,648**	**<1**	**40**	**40**	**20**	**437**
Gender																									
Male	133	5	5	123	1	1	53	43	51	41	18	15	438	212	1	42	38	19	437	5,005	<1	39	40	21	436
Female	101	2	0	99	0	0	38	38	42	42	19	19	437	173	0	38	43	18	437	4,638	<1	41	40	19	437
Not Reported	2	0	0	2										3						5					
Primary Race/Ethnicity																									
American Indian or Alaskan Native	1	0	0	1										1						69	0	20	52	28	432
Asian	14	0	0	14	1	7	7	50	5	36	1	7	441	18	6	56	33	6	441	311	1	37	48	14	437
Black or African American	25	2	1	22	0	0	2	9	12	55	8	36	430	40	0	8	48	45	428	879	0	17	43	40	430
Hispanic or Latino	41	0	0	41	0	0	7	17	22	54	12	29	432	70	0	17	49	34	432	1,738	0	15	45	40	429
Native Hawaiian or Pacific Islander	0	0	0	0										0						0					
White (non-Hispanic)	153	5	4	144	0	0	75	52	53	37	16	11	440	256	<1	51	37	12	439	6,646	<1	50	38	12	440
No Primary Race/Ethnicity Reported	2	0	0	2										3						5					
LEP Status																									
Currently receiving LEP services	14	0	0	14	0	0	0	0	5	36	9	64	425	29	0	3	41	55	427	576	0	6	38	56	424
Former LEP student - monitoring year 1	0	0	0	0										1						63	0	17	51	32	431
Former LEP student - monitoring year 2	1	0	0	1										3						62	0	21	53	26	432
All Other Students	221	7	5	209	1	<1	91	44	89	43	28	13	438	355	1	44	40	16	438	8,947	<1	42	40	17	437
IEP																									
Students with an IEP	37	6	1	30	0	0	5	17	13	43	12	40	428	55	0	15	44	42	428	1,531	<1	16	39	45	428
All Other Students	199	1	4	194	1	1	87	45	81	42	25	13	439	333	1	45	40	15	439	8,117	<1	44	40	15	438
SES																									
Economically Disadvantaged Students	98	4	2	92	0	0	16	17	49	53	27	29	431	160	0	16	53	32	431	4,130	<1	19	47	34	431
All Other Students	138	3	3	132	1	1	76	58	45	34	10	8	442	228	1	58	32	10	441	5,518	1	56	34	9	441
Migrant																									
Migrant Students	0	0	0	0										0						0					
All Other Students	236	7	5	224	1	<1	92	41	94	42	37	17	438	388	1	40	40	19	437	9,648	<1	40	40	20	437
Title I																									
Students Receiving Title I Services	95	3	2	90	0	0	17	19	48	53	25	28	432	152	0	18	50	32	431	3,814	<1	20	45	34	431
All Other Students	141	4	3	134	1	1	75	56	46	34	12	9	442	236	1	55	34	11	441	5,834	1	53	36	10	440
504 Plan																									
Students with a 504 Plan	7	1	0	6										13	0	54	31	15	441	233	1	48	42	9	440
All Other Students	229	6	5	218	1	<1	90	41	91	42	36	17	438	375	1	40	41	19	437	9,415	<1	40	40	20	436

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.

Understanding the District Summary Report

Overview

The *NECAP District Summary Report* provides NECAP results for districts based on the testing of local students in grades 4, 8 & 11.

The *NECAP District Summary Report* provides details, broken down by content area, about student performance for all grade levels of NECAP that were tested in the district. Please note that the *NECAP District Summary Report* includes no individual school data.

The purpose of this summary is to help districts determine the extent to which their students achieve the particular standards and benchmarks contained in the *NECAP Science Assessment Targets.*

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

- the total number of students Enrolled, NT Approved (not tested for a state-approved reason), NT Other (not tested for other reasons), 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 District Summary Report.*

District Summary
2008-2009 Students

District: Demonstration District A
State: Rhode Island
Code: DEMOA

Science	Enrolled	NT Approved	NT Other	Tested	Achievement Level								
	N	N	N	N	Level 4		Level 3		Level 2		Level 1		Mean Scaled Score
					N	%	N	%	N	%	N	%	
Demonstration District A	**1928**	**38**	**41**	**1849**	**11**	**1**	**398**	**22**	**838**	**45**	**602**	**33**	
Grade 4	404	10	6	388	2	1	157	40	156	40	73	19	437
Grade 8	737	17	9	711	3	<1	104	15	333	47	271	38	831
Grade 11	787	11	26	750	6	1	137	18	349	47	258	34	1132

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

District Student-Level Data Files

In addition to all of the reports, districts are also able to access and download student-level data files from the NECAP reporting website for each grade of students tested within their district.

The student-level data files list students alphabetically within each school and contain all of the demographic information that was provided by the state for each student. Student records contain the scaled score, achievement level, and subscores earned by the student for science. In addition, the student records contain each student's actual performance on each of the released items as well as the student's responses to the student questionnaire.

The file layout of the student-level data files that lists all of the field names, variable information, and valid values for each field is also available to districts on the NECAP reporting website. Schools must contact their district office to obtain copies of their student-level data files and the file layout.

Optional Report Data

The data collected from the optional reports field, which was coded by schools on page two of the student answer booklets, are also available for each student in the student-level data file. The optional reports field was provided to allow schools the option of grouping individual students into additional categories (for example, by class or by previous year's teacher). This allows schools to make comparisons between subgroups that are not already listed on the disaggregated results pages of the school and district results reports.

Appendix A

Overview of Assessment Instruments and Procedures
NECAP Tests of 2009

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 *Assessment Targets,* review of all inquiry tasks 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.

NECAP Science Assessment Target Development

The NH, RI, and VT Departments of Education have developed a common set of content standards, known as the *New England Common Assessment Program Science Assessment Targets*, and test specifications in science. These targets 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 2007–2008 academic year, in at least one grade each in the K–4, 5–8, and 9–12 grade spans in science. Although these sets of targets 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 science.

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 NECAP Science Assessment Target 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 inquiry tasks and individual test items. Committee members determine if the tasks 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 task 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 all NECAP tests.

Test Design

TYPES OF ITEMS ON NECAP SCIENCE

In order to provide a valid assessment of students' attainment of the *NECAP Science Assessment Targets*, a variety of item types needed to be used. Therefore, multiple-choice items, short-answer items, and constructed-response items 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 (two points)

These open-ended items ask students to write a short answer in response to a question on the inquiry task.

Constructed response (three and four points)

This item type is used to measure a students' ability to solve a complex multi-step problem. This item type is used to measure a students' ability to apply science content to a unique situation or scientific inquiry skills on the inquiry task.

COMMON AND MATRIX-SAMPLED ITEMS

There are multiple versions, or forms, of the NECAP tests; for science, four forms were created for each grade level tested. The majority 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.

All students at grades 4, 8 & 11 take the same common inquiry task for their grade level.

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. The inquiry tasks for grades 4, 8, and 11 are also released each year.

Content Knowledge and Skills Tested on NECAP

All items appearing on the NECAP science tests were designed to measure a specific NECAP Science Assessment Target.

SCIENCE OVERVIEW

The NECAP science tests at grades 4, 8 & 11 consist of 33 multiple-choice items, 3 four-point constructed-response items (one each for Physical Science, Life Science, and Earth Space Science), and 8 inquiry task items (two-point short-answer items and three-point constructed-response items) that are common for a total of 63 possible raw score points.

The content standards in science identify four domains.

- Physical Science
- Life Science
- Earth Space Science
- Scientific Inquiry

Science Distribution of Emphasis

	4	8	11
Physical Science	24%	24%	24%
Earth Space Science	24%	24%	24%
Life Science	24%	24%	24%
Scientific Inquiry	28%	28%	28%
	100%	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 May 2009 testing period. Training on test administration procedures was provided through Test Administration Workshops held in each of the three states three to four 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, Guidelines, and Procedures: Administrator's Training Guide,* available on the DOE website for each state.

Test Scheduling
The NECAP science tests for grades 4, 8 & 11 were designed to be administered in three separate sessions. The guidelines for scheduling test sessions were based on an estimate that sessions 1 and 2 would require approximately 45 minutes and all students were allowed up to 90 minutes for those sessions. Session 3 varied by grade. Grade 4 required approximately 75 minutes to complete the session, but all students were allowed up to 120 minutes due to the hands-on nature of the inquiry task. Grade 8 and grade 11 students performed a paper-and-pencil inquiry task analysis for session 3 and required approximately 60 minutes to complete the task.

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 July 2009, more than 6.2 million NECAP student 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 three major tasks:

- scoring of responses to multiple-choice items,
- scoring of responses to short-answer items, and
- scoring of responses to constructed-response items

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 two. Constructed-response items were given a score from zero to three or 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.

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 content area for which results are reported. The multi-step process of setting standards for the NECAP tests began with creation of achievement level descriptions.

In August 2008, the state Departments of Education in NH, RI and VT convened panels of educators to participate in the standard-setting process for NECAP science grades 4, 8 & 11. For more detailed information on standard setting, see the *NECAP Science Standard Setting Report,* which is available on the Department of Education website of each state.

Reporting

The NECAP tests were designed to measure student performance against the learning goals described in the *NECAP Science Assessment Targets*. Consistent with this purpose, primary results on the NECAP science 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. 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. 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 science results for grades 4, 8 & 11 are also reported as scaled scores.

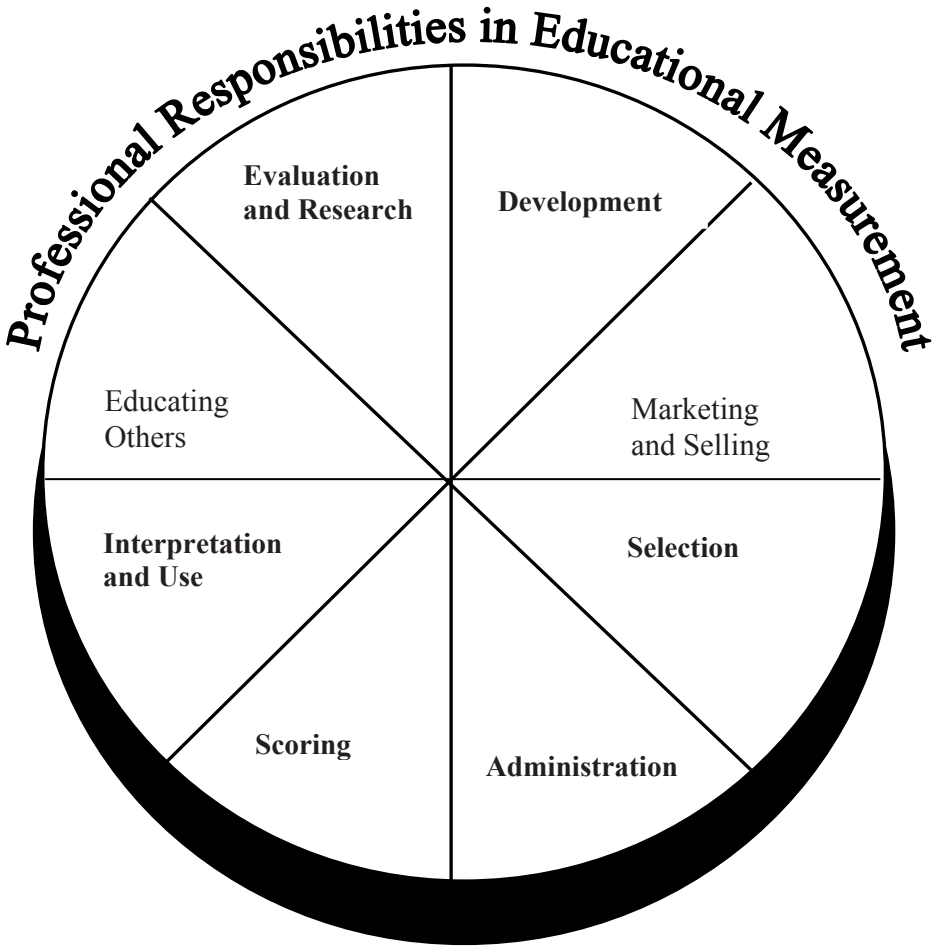
TRANSLATING RAW SCORES TO SCALED SCORES AND ACHIEVEMENT LEVELS

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 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 science at grade 4, but 33 in science at grade 8, or 36 in mathematics 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

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.

Responsibilities of Those Who Administer Assessments

SECTION 4

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.

SECTION 6 Responsibilities of Those Who Interpret, Use, and Communicate Assessment Results

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.

Responsibilities of Those Who Evaluate Educational Programs & Conduct Research on Assessments

SECTION 8

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.

Appendix C

NECAP Achievement Level Descriptions

General Achievement Level Descriptions	
Proficient with Distinction (Level 4)	Students performing at this level demonstrate the knowledge and skills as described in the content standards for this grade span. Errors made by these students are few and minor and do not reflect gaps in knowledge and skills.
Proficient (Level 3)	Students performing at this level demonstrate the knowledge and skills as described in the content standards for this grade span with only minor gaps. It is likely that any gaps in knowledge and skills demonstrated by these students can be addressed by the classroom teacher during the course of classroom instruction.
Partially Proficient (Level 2)	Students performing at this level demonstrate gaps in knowledge and skills as described in the content standards for this grade span. Additional instructional support may be necessary for these students to achieve proficiency on the content standards.
Substantially Below Proficient (Level 1)	Students performing at this level demonstrate extensive and significant gaps in knowledge and skills as described in the content standards for this grade span. Additional instructional support is necessary for these students to achieve proficiency on the content standards.

Appendix D

Reference Materials

Coladarci, T, Cobb, C.D., Minimum, E.W., & Clarke, R.C. (2004). *Fundamentals of statistical reasoning in education*. Hoboken, NJ: John Wiley & Sons, Inc. (ISBN: 0471069728)

Glass, G.V. & Hopkins, K.D. (1996). *Statistical methods in education and psychology* (3rd edition). Boston: Allyn & Bacon. (ISBN: 0205142125)

Shavelson, R.J. (1996). *Statistical reasoning for the behavioral sciences* (3rd edition). Boston: Allyn & Bacon. (ISBN: 020518460X)