



# **NYSCOSS Curriculum & Instruction Committee**

## ***ASSESSING THE ASSESSMENTS – 2005***

Walter J. Sullivan, Chair Emeritus  
Superintendent, Skaneateles Central School District

**New York State Council of School Superintendents**

**August 2005**

### **INTRODUCTION**

This document is a synopsis of feedback relative to the recent state assessments compiled by the Skaneateles School's Curriculum Coordinators, Dr. Brenda Weaver, Kathleen DiGennaro, Dr. Margaret Culkowski, John Harmon and staff. Reactions were incorporated from individual members of the Subcommittee on Assessment of the New York State Council of School Superintendents. The report provides a review of the assessments that New York's students took during the 2004-2005 school year. It is meant to be an objective review of New York's assessments as they affect the education and testing of students in New York State.

### **ELEMENTARY LEVEL**

#### **4<sup>th</sup> Grade English Language Arts (ELA) Assessment**

Our past reports spotlighted the Education Department's difficulties in calibrating the readability levels of passages used in the Grade 4 ELA. The 2005 version was an appropriate assessment for 4<sup>th</sup> graders in New York State due to more realistic readability levels of the passages as compared to last year. For five of the passages used in the exam the readability range was 2.4-3.5. Two passages were 7.4 and 8.7. While the exam still includes secondary level reading passages, the majority of the exam was grade appropriate.

It is important to continually monitor the readability of the passages used in this exam.

#### **5<sup>th</sup> Grade Social Studies Assessment**

The 2005 Grade 5 Social Studies Test was a fair, appropriate exam for 5<sup>th</sup> graders: the test was aligned with curriculum and instruction, it assessed what teachers taught, and it did so at a reasonable level of difficulty. Skaneateles teachers generally believed the exam was one of the fairest given. The only area that was somewhat troublesome was that not all of the Document Based Question documents could be used to support the DBQ task. However, this fact did not present a major stumbling block to success on the exam.

#### **4<sup>th</sup> Grade Mathematics Assessment**

##### ***Book I, Session I:***

The 30 multiple-choice questions addressed each of the Seven Key Ideas of the New York State Math Standard with the same number of items for each as in past years. Most students were able to complete the 30 items in the time allowed. The reading required for each

question was not as lengthy as last year's and was on grade level in terms of readability and, therefore, did not hinder the students' understanding of the mathematical tasks.

#### ***Book 2, Session II and Session III:***

The items in Book 2 addressed each of the Seven Key Ideas of the New York State Math Standard with a similar number of items for each as in the past years. As in Book 1, the reading required for each item in Book 2 2005 did not present as an overriding obstacle to the students as in previous years.

Item #39 required the students to create the ubiquitous bar graph. In the Scoring Guide 2005, student sample 39f was given full credit for a graph with bars that may represent those of a histogram. This is consistent with information pertaining to the scoring of question 48 in the Question and Answer Document, Grade 4 Mathematics – 2004 which stated "Q: *In Item 48, why are we accepting bars touching in bar graphs? A: We must remain consistent with how the field test was scored.*" but contrary to the Question and Answer Document, Grade 4 Mathematics – 2003 which stated: "ITEM 37 Q. *If a student makes a histogram rather than a bar graph, is it acceptable? A. Yes. Response 37i in the Scoring Guide is a histogram and it received full credit. However, 2003 is the last year that a histogram will be accepted when the item specifies a bar graph.*"

#### **4<sup>th</sup> Grade Science Assessment**

The format of the new 2005 ELS (Elementary Level Science) is the second year of the present format and is similar to the format of the Grade 8 Intermediate Level Science assessment with a written test and a performance test.

##### ***Part 1***

The written part consisted of 30 multiple-choice questions, each worth one point and 10 open-ended written response questions each worth one to three points for a total of 15 points. Students were allowed unlimited time to complete these items; only a few students required more than the 45 minutes scheduled.

##### ***Multiple Choice (#1-30)***

The items on this test were similar in content and specificity to the questions on the 2004 sampler, and 2004 ELS.

##### ***Open-Ended Questions (#31-40)***

There were 10 open-ended written response questions this year with seven worth one point, two worth three points, and one question worth two points.

##### ***Part 2, Performance Stations (#1-3)***

The construction of this portion of the test was exactly the same as last year. The performance session of the test was administered on a separate day, during which students were required to complete three independent tasks and related constructed response questions worth 1-3 point for a total of 26 points. The students were allowed 15 minutes for each task, which proved to be more than enough time for the students to comfortably finish each task.

## **MIDDLE LEVEL**

### **8<sup>th</sup> Social Studies in Grade Assessment**

The test itself is a fair assessment of student progress, measuring what teachers taught and scored at a reasonable level of difficulty. The biggest issue with this exam is the enormous burden it places on teachers and administrators to rate the exam. In a school of 400-450 students (total enrollment, grades 6-8), it takes five teachers devoting 4-12 hours each to rate the exam. In addition to that time, the teachers had to be trained for the scoring. Finally, there are three different forms for the partial scores each student receives in the exam. All of these score reports have to be filled out, reconciled, and then finally transferred on to the final answer sheet. This paperwork alone takes hours and hours.

Perhaps the biggest burden, though, is on the students. They are simply not ready, in 8<sup>th</sup> grade, to take a three-hour exam. In fact, the test is two 90-minute sections. There continues to be a disregard for the developmental nature and tasks appropriate for middle school students

### **8<sup>th</sup> Grade English Language Arts Assessment**

Although the test itself is fairly well designed, it is still far too long. 8<sup>th</sup> grade students should not be expected to sit for two lengthy sessions on two consecutive days. There is most certainly a reliable and valid way to assess students in Language Arts using no more than a single three-hour session. Once again, the readings for the assessment come from disciplines other than literature. Although this is an ELA exam, there is very little of the type of literature typically studied in English classes. It is heavily weighted in favor of nonfiction.

Again, the scoring of this exam is a huge burden. Even in a relatively small school (400-450 students) with seven teachers participating in the scoring, each one had to miss a day in the classroom for both training and rating. This rating becomes yet another burden on the middle school. All of this preparation, training, and scoring results in a final score for the student of 1-2-3 or 4. These scores do not tell us a single thing about the students' ability in *language arts* that we didn't know before the exam.

Since the inception of the test, SED has made significant changes in how Item Response Theory is applied in scoring. Previously, consideration was given to which specific questions a student answered correctly. Now, only the percentage of correct answers is taken into account. One result of the change is that, in the middle range, an incorrect answer on a multiple choice question moves a student's score by one or two points, while in the upper range, a single incorrect answer could move a student's score by 30 points! The test is to be used to make judgments about students and schools, but a scoring system that produces such drastic swings is spurious. The score reports are also confusing to parents. Items from the multiple-choice are rather arbitrarily assigned to one of the four state standards. In reality, these standards are interconnected and not at all discrete. The result is that if a student misses two or more questions linked to that one standard, then that student is assigned an artificially low score for that standard. It is difficult explaining to parents that these sub-scores are meaningless.

As our study indicates and the District Superintendent's study reflects, there is little correlation between the results on the ELA 8 and the Regents English 11 except to note that students with a score of 1 or a low 2 in ELA 8 will have difficulty on the 11<sup>th</sup> grade English Regents (a graduation requirement).

### **8<sup>th</sup> Grade Mathematics Assessment**

This year's assessment was, for the most part, a fair one. One question that students found difficult was #33: *"In the space below, use prime factorization to find the least common multiple of 120 and 252."* The numbers 120 and 252 are quite large. That is probably why students confused least common multiple (LCM) with greatest common factor (GCF). The fact that students had to use a particular method of finding the LCM led them down the path of finding the GCF. Prime factorization is the method that students use to find the GCF. In an era where we celebrate different methods of finding the answer to questions, it seems incongruous that we dictate the method that students must use to solve a problem, and it is the method that is used to find the GCF. Only one percent of students in some schools earned full credit on this question.

A concept that still gives students trouble is the difference between an expression and an equation.

Next year's mathematics assessment has been scheduled for two months earlier in the school year, moving to March (from May). In addition to the different schedule, there will be a new curriculum statewide for grade 8 math (and K-7, as well). Comparing results from May 2005 to March 2006 will be an interesting exercise for the statisticians, replete with design flaws.

### **8<sup>th</sup> Grade Science Assessment**

As has been noted in previous years, the valuable portion of this assessment is the Performance Test, which was given in January. In the New York State assessment shuffle, this test will be moved to April and May of next year. It clearly prepares students for the laboratory skills they need for success in high school science.

The format of the written test needs improvement. There is no reason that teachers should have to score 32 one-point questions and then pencil in those 32 bubbles for each student on the answer sheet. These questions often require one-word answers. They could be written as multiple-choice questions and spare the teachers the drudgery of scoring and bubbling. This written test was originally conceived as being composed of short answer and extended responses. It has evolved into 45 multiple-choice questions on one side of the answer sheet, and 35 "extended response" questions on the reverse side of the answer sheet. Each of these extended response questions has to be scored and bubbled by a teacher. Of the 35 questions, all but three were one-point questions! Each of these questions could have been written as a very good multiple-choice question. It may have taken longer to write the question, but would have saved valuable teacher time later on. The New York State assessment shuffle has this test scheduled for May next year, one month earlier. It will be a challenge to teach all of eighth grade science and review four years of science (grades 5-8) prior to May.

## **HIGH SCHOOL LEVEL**

### **English Regents**

This test continues to be far too long. Why is English the only Regents exam that has to be given in two three-hour sittings? Certainly a three-hour exam could be designed which would effectively and reliably assess students reading and writing. Perhaps part of a student's Regents English score could be derived from writing completed during the school year. After

all, that practice is already well established in the Foreign Language exams, as well as the science exams, where students are given credit for work done before the exam.

The content for this June's exam was, once again, only marginally literary. The listening section was fairly engaging, a biographical discussion of Nellie Bly, the newspaper reporter who traveled around the world. The informational passage was stunningly boring, requiring the students to read a tedious discussion of pedestrian safety, accompanied by a requisite chart. Of all the sections, this part of the exam needs to completely disappear. Students' scores would vary by only a tiny fraction if this section were removed. The task adds no new information to a student's overall score. As a brief recap of this section for the past four years we have: pedestrian safety, the griots of Africa, global warming and forest fires. Again, how many schools include such readings in their English curriculum?

Once again, the issue for this exam is the huge amount of time and human resources required to rate the exam. It took eight teachers five full days to train for and rate these exams. Even in a relatively small school (600-650 students in grades 9 through 12) there were nearly 1300 essay readings.

The scoring grid for this June's exam was really quite fair.

### **Global History**

The exam and its results are very similar to those of the last few years. Because this exam also includes extensive student writing, scoring becomes a huge task, requiring several teachers and many hours.

The writers for the exam persist in developing essay prompts with the phrase "and/or" in them. Any lawyer will advise that such a term should never be used in a high-stakes situation; it is far too ambiguous. For example, if a student were asked to write about the "social, political, and/or religious" influences of a historical event, how many from that list should the student choose? The word "or" clearly implies that the student could choose one. If a person is asked, "Paper *or* plastic?" that person will receive only one, not both. Yet the scoring guide (which the students don't see) requires at least two from the list.

### **United States History**

This exam is also quite similar to last year's. Just as with the Global History exam, the rating of the essays becomes an incredibly huge task. In fact, in many schools, substitute teachers need to be hired to assist in dealing with the paper load.

### **Earth Science**

No major problems on the exam or with the scaled score chart. Great job by the SED on this exam.

### **Living Environment:**

There were no major problems with the exam. However, question #3, a multiple-choice question about the relationships between the function of human organs and the organelles of a single cell, was puzzling. It took science teachers quite awhile to figure out what it was asking. There was probably a better way to get at the concept than the way it was presented. Also, in question #35, the word *acrosome*, is not a Regents word. It is fair to use words that are not in the Core, but although it was labeled on the sperm diagram, it caused some students to stumble on that question since it also appeared in one of the answer choices. The scaled score chart for the Living Environment exam was generous.

**Physics**

For the last two years some schools have made the Regents Physics exam optional for students. Instead, students took local exams which have four parts patterned after the Regents science exams. Some local exam includes a Part D based on labs that each student was required to complete. There is no Part D for the Regents Physics exams because although it is a goal of the SED to create a lab-based Part D, it has not been done.

Physics is still an elective in New York State schools and the SED needs to make sure that everything possible is done to encourage students to challenge themselves in order to participate in this rigorous curriculum. Many schools are watching to see if the actual physics exam will be fair.

**Chemistry**

In each of the last three editions of this report, we have noted that problems with the score conversion chart for the Regents Chemistry Exam make it hard for students to earn mastery level scores (85 or better). Those problems persist. The contrast with this year’s Math B Regents Exam is striking. The mathematics community identified a problem, contacted SED, and a new score conversion chart was issued within 24 hours.

Students don’t necessarily fail the Chemistry Regents exam, though. It is just that the scaled score chart makes it difficult for a student to earn mastery. *Every score between 85 and 100 exists on all science and math exams, except Chemistry. Why? Why is it not possible for a student to earn a 99, 97, 95, 92, 89, or even an 84 in Chemistry? Why do all those scores exist on every other test but Chemistry? What is so unique about the Regents Chemistry exam that makes the scaled score chart so different from every other math and science Regents exam?* Students continue to pass the exam, unlike the June 2003 Regents Math A exam, unlike the June 2002 and June 2003 Physics exam, and unlike the June 2005 Math B exam, all of which were issued “revised” scaled score charts so that more students would pass. But there is still a problem.

Chances of Earning "Mastery" Grade (85 or better) On Regents Science Exams				
Raw Score	Percentage of Points Earned	Scaled Score for Living Environment	Scaled Score for Physics	Scaled Score for Chemistry
85	100	100	100	100
84	99	99	99	98
83	98	98	98	96
82	96.5	97	97	94
81	95	96	96	93
80	95	95	95	91
79	93	94	94	90
78	92	93	93	88
77	91	93	92	87
76	89	92	91	86
75	88	91	90	<b>85</b>
74	87	90	89	83
73	86	90	89	82
72	85	89	88	81
71	84	88	87	80
70	82	88	86	79
69	81	87	<b>85</b>	78
68	80	86	84	77
67	79	<b>85</b>	83	76
66	78	<b>85</b>	83	75

In Living Environment, there are 20 different scores that will earn a student "Mastery". In Physics, there are 17. In Chemistry, there are 11.

**Math A**

The test was fair in its alignment with curriculum and instruction and the scaled score chart was generous, particularly compared to 2003.

**Math B**

There were problems in numerous districts in New York State on Math B. The Math B curriculum is difficult to teach in three semesters. Juniors often are not ready for the Math B

exam, and that is why some districts wait until January of the senior year to administer the Math B Regents Exam to students. Districts may want to think about waiting. Some districts had little problem with this test and perhaps other districts who had difficulty may wish to follow the model described above. Perhaps something can be learned from utilizing this new model. Except for accelerated students, a typical student takes the Math B Regents after the first semester of pre-calculus. Results generally have been good and perhaps districts will see better results using this time, structure and preparation. In districts that have used this model, they have not eliminated or reduced any regular pre-calculus topics. Topics can be shifted between junior and senior year to make it better for students.

One question that was very unfair on the Math B was the relatively obscure proof (#33), a six-point question. Inequality proofs are not stressed in Math B. There was disappointment throughout the state in the negative scaled score chart on the Math B Regents exam. This chart was changed and a new chart was published on the afternoon of June 24<sup>th</sup>. The new chart was more generous, and every score does exist between 85 and 100. This last-minute creation of this "revised" new chart was disruptive and confusing for both teachers and particularly students. However, we compliment the SED in taking quick action to assist with this troublesome exam.

### **"STEP" DATA REPORTING SYSTEM**

---

Many districts have reported significant problems with the state's STEP (System for Tracking Education Performance) data reporting system. In many cases, the data that is initially transmitted to the SED from the Regional Information Centers (RICs) is incorrect. In some cases there is confusion over data entry at the district level and in other cases, it involves the transferring of data from the RICs to the State Education Department.

One obvious need for change is that if corrections are necessary to have valid and correct data, the SED has to have the capability to make those corrections even if they are needed after a so-called due date. There has been a tremendous amount of confusion throughout New York State with regard to information incorrectly provided to the press and the media. There has to be an ability for school districts to make ongoing changes into the system. Each case is individual to the student and the school and the school district needs the opportunity to update the SED state files before they are disseminated to the media and the public..

### **CONCLUSION**

---

The state assessments this year in many areas seem to have remarkably improved.

1. It is true that the state assessment on the ELA 4 (being the oldest of the assessments) has gradually improved. There continues to be a need to double check on readability because it is a much more important issue than often is acknowledged.
2. It is important for the public, teachers and schools to recognize that these are accountability tests, not diagnostic in nature. The diagnostic and prescriptive tests that can be used to better inform instruction need to be created particularly with a barrage of new tests under NCLB coming online.

3. Math B was troublesome this year primarily because of the score conversion chart. However, chemistry continues to provide unique problems with its chart. There was no reason that what was done for Math B could not be done for chemistry as well. This has been a long-standing problem. It appears that because the problem occurs at the mastery level, it is ignored. However, we do appreciate that the Math B chart was revised almost immediately.
4. The lack of correlation between the ELA 8 and the High School English Regents exam, which is a graduation requirement, still exists.
5. As we had mentioned on numerous occasions, students who take the very rigorous AP exams or baccalaureate exams could also be given Regent's credit in subjects where there is obviously a relationship. For example, the student who takes advanced placement English could also receive Regents credit in English rather than duplicating the time spent on testing. (*Note: NYSCOSS has called for (1) the establishment of an appeals process to permit students who fall just short of passing a required exam to submit alternative evidence that they have met state standards; (2) creating an option for students to complete "project-based assessments" similar to those employed in the highly respected International Baccalaureate program; and (3) adopting more user friendly component re-testing – permitting students to re-take portions of a test they have failed when they and their teachers feel they are ready, instead of only at one set time each year as is currently done. The Board of Regents has since authorized creation of an appeals process*).
6. STEP: Data Reporting System -- There has to be a capacity to make changes and corrections to district data as corrections become necessary particularly before it goes to the public.

Overall, the improvement is noticeable and we compliment the State Education Department. The challenge will continue to be what happens with the continuing testing that is "robbing" instructional time from teachers as well as requiring tremendous amounts of time for scoring and grading. What will happen when these additional tests required by No Child Left Behind come online starting next year?

The New York State  
Council of School Superintendents  
7 Elk Street, Third Floor  
Albany, New York 12207-1002  
(518) 449-1063 — (518) 426-2229 Fax

[www.nyscoss.org](http://www.nyscoss.org)

The New York State Council of School Superintendents is a professional organization of school district leaders providing member services, influencing educational policy and asserting the leadership role of superintendents.