

TO: Richard P. Mills

FROM: Math A Panel  
William Brosnan, Stanley Chapman, Gregory Cizek,  
Franco DiPasqua, Andrew Giordano, Lidia Gonzalez,  
Robert Gyles, Daniel Jaye, Sophia Maggelakis,  
Theresa McSweeney, Alfred Posamentier, Katherine Staltare,  
Alan Tucker

SUBJECT: Interim Report

DATE: August 25, 2003

### INTRODUCTION

The Panel would first like to express its appreciation to the Board of Regents and the Commissioner for our appointment. Although we come with different perspectives, we have reflected on several occasions about the positive dynamics of the group. Appreciation is extended to Tom Sheldon for his enormously successful efforts to create a strong Panel.

The Panel would also like to convey its appreciation to the leadership and staff of the New York State Education Department for providing us with substantial amounts of information and materials, and unfettered access to personnel. We are grateful for the full and enthusiastic cooperation we have encountered, which helped make our difficult task manageable and, we believe, successful to date.

This interim report focuses on Elements #4 and #8 of our charge, which are:

#4. Is the June, 2003 Regents Math A exam of the same level of difficulty as prior Math A exams? (That is, in addition to the equating included in question 2, consider the content, cognitive demand, and perceived difficulty of the exam.)

#8. If the June, 2003 Regents Math A exam was not of the same level of difficulty as previous Math A exams, can the results be re-scaled appropriately and used to measure student achievement, and if so, how?

While this report focuses on these elements, our research and findings touch on several other elements of the charge.

**Math A Interim Report****Page 2 of 6**

This interim report is a response to the urgent problem of how June, 2003 Math A test results should be applied to 9<sup>th</sup> and 10<sup>th</sup> graders as they finalize their programs of study for the coming school year. Our work will continue in September when we meet to address the remaining elements of our charge.

**FINDINGS**

A very detailed analysis was conducted of the June, 2003 Math A Regents exam, and an extensive comparison was made of that exam with the June, 2002 Math A Regents exam. We found:

1. An analysis of the Rasch Item difficulty values generated by the Item Response Theory (IRT) method conducted on both exams shows that, while the overall difficulty of the two exams appeared comparable, disaggregation by exam parts showed important differences. While the difficulty of Parts 1 and 2 were close for the two exams (with the June, 2003 exam being slightly easier), Parts 3 and 4 of the 2003 exam were substantially more difficult than the same parts of the 2002 exam. (See Appendix A.)
2. Appendix A also includes a comparison of the probability of success on each item as generated by the IRT analysis, and an aggregation of these probabilities for the June, 2002 exam and the June, 2003 exam. We found:
  - The expected average score for the field test takers of the June, 2002 exam was 51.
  - The expected average score for the field test takers of the June, 2003 exam was 46, five points lower.

(It should be noted that the expected average score for the field test of *both* exams was below the 65 passing level and even below the 55 which was available as a passing score for a local diploma for some students until June, 2003.)

3. While one subgroup of our Panel was working on the above-mentioned item analysis, another subgroup was looking at the wording of the items in Parts 3 and 4, and the order in which the items were presented to students. When the subgroups reconvened together, we were struck by the fact that the conclusions reached by the two subgroups were virtually identical. The findings of the second group are detailed in Appendix B. In addition to finding that the 2003 items were more difficult, this group found that the students were presented at the beginning of Part 3 with several difficult items on a row, which is consistent with anecdotal evidence from the field that some students this year

**Math A Interim Report****Page 3 of 6**

reached Part 3 and faced several tough problems in a row, which caused them to become frustrated and "give up."

4. In addition to the greater difficulty of the June, 2003 items, the Panel identified content coverage concerns in the June, 2003 exam. For example, 3 of the 35 items tested the Pythagorean Theorem; this would dramatically lower the score of a student with a weakness in that area. Yet, trigonometry was not assessed at all (which was a tremendous source of frustration in the field as previous Math A exams contained trigonometry items and many teachers spent substantial time teaching this to their students -- yet there was not one trigonometry question on the exam). In addition to potentially depressing the mean performance on the June, 2003 examination, such imbalances in content representation can also attenuate the content validity of the test.

The above-mentioned statistical and substantive differences lead the Panel to the conclusion that the June, 2003 exam was harder than the June, 2002 exam. In short, students in June, 2003 were held to a higher standard than their counterparts a year earlier.

This conclusion is consistent with evidence across the State about how students performed. One example is from the Marcellus School District, which had the following results on the Math A exam for those who scored a "2" on the Math 8 exam:

- In January, 2002, 11 students passed Math A, out of 12 who had scored a 2 in Math 8.
- In June, 2002, 16 students passed Math A, out of 19 who had scored a 2 in Math 8.
- In January, 2003, 19 students passed Math A, out of 21 who had scored a 2 in Math 8.
- In June, 2003, none (0) of the students passed Math A, out of 15 who had scored a 2 in Math 8.

While this evidence is anecdotal in nature, it is consistent with the experience of the K-12 educators on the Panel in their classes, schools, and districts.

In summary, the Panel is convinced there is compelling evidence that the June, 2003 Math A exam, when compared with the June, 2002 exam, was substantially more difficult for students.

**Math A Interim Report****Page 4 of 6**OPTIONS CONSIDERED

The Panel considered several possible ways of responding to the fact that this year's exam was more difficult than the June, 2002 exam. At the outset, we agreed that any solution recommended would need to meet these criteria:

- the solution must "do no harm" to any student (that is, no student's standing should be made worse by the application of any potential remedy);
- the solution must be understandable (that is, it should be able to be easily explained to parents, educators, policy makers, etc.); and
- the solution must be defensible, as it should have the characteristics of fundamental fairness and be psychometrically sound.

The Panel spent hours exploring possible approaches.

The Panel considered recommending setting aside the June, 2003 examination, as was done (and we believe very appropriately so) for the 11th and 12th graders. Several concerns were expressed about doing this for the 9<sup>th</sup> and 10<sup>th</sup> graders, a major one being that these students would be held to no objective standard. This could allow students to "slip through" and receive diplomas with weaker math skills than needed for success in their futures. Additionally, it was believed that such a precedent could serve to undermine standards-based reform efforts and implementation of important changes to the way mathematics is taught and learned in New York. A significant majority of the Panel was not in favor of this option.

The Panel considered somehow removing the items that were the most troubling. This, also, generated several concerns, the most significant of which was that every item had some students responding correctly; by removing an item and rescaling, those students who had answered the item correctly could end up with a *lower* score. The Panel was unanimous in its thinking that this would violate our "do no harm" criterion for a decision. This was rejected by all.

The Panel considered "curving." That option was quickly rejected because there is no basis for doing that. "Curving" would require determining where the "average" child should score, i.e., a 65 or a 75, etc. There was no basis whatsoever for making such a determination. "Curving" would result in a total disconnect between the meaning of a student's test score with respect to the content standards, replacing this important interpretation with a norm-referenced interpretation.

A number of other adjustments were considered. Ultimately, the Panel focused on the fact that the effect of the anomalies found in this test could be estimated with reasonable precision by looking at comparable groups of 9<sup>th</sup>

**Math A Interim Report****Page 5 of 6**

graders. There are some differences between the June, 2002 population and the June, 2003 population in that, this year, more students who are struggling in math took the Math A exam because the Course I exam is no longer an option. However, struggling math students are almost invariably programmed into 10<sup>th</sup> grade; the 9<sup>th</sup> grade has included, and continues to include, only those students who are strong in math and who the teachers feel can challenge this exam at that early stage of their high school career. Thus, a comparison of the June, 2002 9<sup>th</sup> grade result and the June, 2003 9<sup>th</sup> grade result is valid. The data we have is that, in the sample of 400 school districts for which the State Education Department has data, the June, 2003 9<sup>th</sup> graders scored 9.6 points lower than the June, 2002 9<sup>th</sup> graders. This drop of almost 10 points from June, 2002 to June, 2003 is very consistent with our review of the item statistics, as well as our qualitative review of the Items.

This 9<sup>th</sup> grade comparison led us to our recommendation.

**RECOMMENDATION**

The Panel recommends that the scores on the June, 2003 exam be statistically adjusted, using the 9<sup>th</sup> graders as a basis, so that the June, 2003 students will receive a score similar to what they would have received had they taken the June, 2002 exam. The adjustment the Panel is recommending has these three steps:

1. Draw matched samples of 9<sup>th</sup> graders from the June, 2002 exam and the June, 2003 exam that are as representative as possible of the entire State.
2. Conduct an equal-percentile equating of the two distributions.
3. Establish the conversion of raw scores to scaled scores and put the June, 2003 scores on the June, 2002 scale.

The Panel is recommending this adjustment because it holds this year's students to the same standards as the ones to which the June, 2002 students were held; it essentially corrects for the anomalies of the June, 2003 exam. Any adjustment less than this would be unfair because it would not correct for the full effect of the anomalies found in the exam. Any adjustment greater than this would result in an overcorrection, which means the students who took the June, 2003 exam would end up being held to a lower standard than last year's students; the majority of the Panel felt this would not be a fair result.

The recommendation for the score adjustment detailed above is endorsed by the entire membership of this Panel present for the two day session on August 19 and 20.

**Math A Interim Report****Page 6 of 6**THE PANEL'S WORK CONTINUES

This Panel has been asked to review in great depth the status of Math A. As we have explored the June, 2003 exam, we have identified a number of issues about which we have serious concerns, including, but not limited to, the content standards, the performance standards, the lag in full implementation of curriculum and instruction aligned with the Math A frameworks, and the infrastructure to support the move to higher standards. We view the adjustment we are recommending for the June, 2003 exam as an interim step. We intend to consider the broader issues at our upcoming meetings on September 10, 11 and 19, and to make recommendations to the Board of Regents on October 8. It will be our hope that these broader recommendations, if accepted, will result in changes beginning with the January, 2004 exam. The Panel sees the recommendation made in this report as an interim step to take care of the students who took the exam last June and to enable them to be appropriately scheduled for the fall semester.

Test Date	Part I		Part II		Part III		Part IV		Total
	<del><math>\theta</math></del>	E(x)	<del><math>\theta</math></del>	E(x)	<del><math>\theta</math></del>	E(x)	<del><math>\theta</math></del>	E(x)	E(x)
June 2003	-.88	23.5	0.30	4.65	0.74	4.34	1.43	2.72	35.21
June 2002	-.64	22.8	0.60	3.66	0.16	6.65	1.01	4.62	37.73

	<u>Expected</u> <u>Raw Scores</u>	<u>Scale Scores</u>
June 2003	35	46
June 2002	38	51

Comparative of Difficulty of Form (June 2002-June 2003)

1) Grouping of items by difficulty

June 03	III	#26	#28	#30	#27	#29
	0 =	.52	.53	.76	.76	1.11
June 02	III	#26	#27	#30	#29	#28
	0 =	-.22	.18	.51	.54	.85
June 03	IV	#34	#31	#33	#35	#32
	0 =	.96	1.14	1.24	1.79	2.1
June 02	IV	#32	#31	#33	#35	#34
	0 =	.48	.74	.97	1.27	1.61

Easiest → Difficult by theta values

The effect of failing to order in increasing difficulty can account for a depressive performance in subsequent items.

2) Improper balance of content standards

a) Overrepresentation

↳ 7 key ideas

↳ 32 performance indicators

↳ 103 bullets

35 items on test

#8 & #17	key idea 1, PI A,	bullet #1	Part I
#25 & #32	key idea 4, PI B,	bullet #1	Part II & IV
#33 & #35	key idea 7, PI B,	bullet #1	Part IV
#31 & #34	key idea 5, PI A,	bullet #3	Part IV
	↳ #27 key to idea 4, PI A,	bullet #1	(5A3-volume of solids)
	↳ #30 key idea 5, PI A,	bullet #4	
	↳ #34 key idea 5, PI A,	bullet #4	
#5 & #29	key idea 5, PI B,	bullet #2	#6 & #15 KI2, PIC, B1

b) Omission of themes (examples)

5E - Trig

4E - Systems of equations and inequalities

- 3) "Inequity issues"
  - a) Graphing calculator issue #33, unclear directions about sketching
  - b) "Reading" Issues
    - \* #35 "breakeven" - economics term
    - \* A comparison between Pts. III & IV; June 2002-June 2003, the problems in the June 03 exam were clearly more "wordy" and might have become a disadvantage where the students' reading weakness may have affected their mathematical performance
  - c) Figures not stated "not drawn to scale."  
#31 & #34
- 4) Lack of Clear Direction
  - a) #27 change of measurement not italicized as in previous exams
  - b) #32 unclear directions of what students needed to show
  - c) #22 5:4 ratio was not indicated as corresponding parts
- 5) Rubrics
  - a) #29 - the length of the question, two parts  
large numbers - 3 points
  - b) #24 - conceptual and computational both = 1 point
  - c) #21 - simple mental arithmetic, penalized for not showing  
- stem and leaf plot not mathematically correct