NEW YORK STATE EDUCATION DEPARTMENT

# TECHNICAL PAPER #1-02

# Using the Performance of Students at the Passing Score to Enhance Regents Item Analysis

**OCTOBER, 2002**


The Office of State Assessment

775 EBA * 89 Washington Avenue * Albany, NY 12234

# Using the Performance of Students at the Passing Score to Enhance Regents Item Analysis

October, 2002

## INTRODUCTION

Many schools conduct analyses of student test performance based on a test item-by-item review. By examining the performance of students on test questions they hope to better define particular strengths and weaknesses. One common method for doing this is to analyze the proportions or percentages of students that answered each question correctly. The value of this type of analysis is limited by characteristics of the test questions. For example, this type of analysis fails to separate the weaknesses of the students from the difficulty levels of the questions. This limitation is best addressed by comparing student performance to meaningful achievement milestones on the same test questions which would account for item difficulty. The State Education Department has undertaken to provide such milestones. Schools can then use individual item-analysis data to help evaluate curriculum strengths and weaknesses and identify aspects of the instructional programs that should be changed.

## A TOOL TO HELP

The Office of State Assessment has prepared a series of tables (see Appendices A and B, as well as below) to help districts use Regents examination results to evaluate the academic needs of groups of students. These tables group the questions on specific Regents examinations by the components they address. Table 1 shows three of the seven components or key ideas that the Mathematics A Regents Examination tests. This grouping of test questions reflects meaningful content categories. Educators can see for each test administration, by comparing their students' scores to the achievement milestones, the components in which each student or group of students are:

1. Weak: achieving fewer points than the expected number for a student or group at the scale score of 65;
2. Proficient: achieving as many points as the expected number for a student or group at a scale score of 65, but fewer points than the expected number for a student or group at a scale score of 85;
3. Distinguished: achieving at least as many points as the expected number for a student or group at a scale score of 85.

Table 1
June 2002 Mathematics A Examination Component Information

| Mathematical Reasoning | | | | | Number and Numeration | | | | | Uncertainty | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Expected Points for a Student with a Score of | | | | | Expected Points for a Student with a Score of | | | | | Expected Points for a Student with a Score of | |
| Item | Max Raw Points | Weight | 65 | 85 | Item | Max Raw Points | Weight | 65 | 85 | Item | Max Raw Points | Weight | 65 | 85 |
| 3 | 1 | 2 | 0.72 | 0.87 | 11 | 1 | 2 | 0.70 | 0.86 | 2 | 1 | 2 | 0.96 | 0.99 |
| 21 | 2 | 1 | 0.35 | 0.87 | 24 | 2 | 1 | 0.80 | 1.30 | 34 | 4 | 1 | 0.36 | 2.21 |
| Weighted Total | 4 | | 1.79 | 2.61 | Weighted Total | 4 | | 2.20 | 3.02 | Weighted Total | 6 | | 2.28 | 4.19 |

**Abbreviated Version of Appendix B**

## APPLICATION TO COMPONENT RETESTING

These analyses are important for choosing which component retests to administer to eligible students. By identifying which component(s) the student is relatively weaker in, teachers and administrators can better direct instruction for the student. Students who can be identified as specifically weaker in certain components may be able to take a component retest to receive credit for passing the exam.

## READING THE TABLES

Table 1 shows the expected performance of students on the Mathematics A Regents Examination given in June 2002. Three of the seven key ideas, or component areas, appear here. All seven areas would appear on a table to be used for item analysis. The expected performance is computed using item response theory, which bases expectations on the observed difficulty of each question for students at each scale score level, e.g., 65, 85, 100.

Each area has five columns: Item, Max. Raw Points, Weight, Expected Points for a Student with a Score of 65, and Expected Points for a Student with a Score of 85. At the bottom of each column is the "Weighted Total" row. This row presents the criterion totals for each area. Thus, a *just passing* student, that is, a student who achieves a scale score of 65 on the Mathematics A Examination in June 2002, has the greatest likelihood of achieving a weighted score of 1.79 on the Mathematical Reasoning component of the examination. The reader will note that the questions that comprise each key idea are given in the scoring guide for each administration of Mathematics A. The process for computing each student's performance, then includes:

1. Identifying the raw total achieved on each question in a component;
2. Multiplying that raw total for each question by that question's weight, as given in Appendix B;

3. Summing these products over all of the questions that comprise the component.

Comparing the actual total points derived from this process for each student to the expected raw value of 1.79 for students at the scaled score of 65 tells educators whether each particular student is stronger than *just passing* or weaker than *just passing* for those questions. Comparing a student's actual total point to expected total for students at a scale score of 85 tells educators whether this particular student was stronger or weaker than students at the level of distinction in this area. This comparison enables educators to locate each student's performance in each component as to whether the performance is:

- Below the expected performance of a just passing student;
- Between the expected performance of a just passing student and a student passing with distinction and closer to the expected performance of a just passing student;
- Between the expected performances of a just passing student and a student passing with distinction and closer to the expected performance of a student passing with distinction;
- At or above the expected performance of a student passing with distinction.

## HOW TO INTERPRET THE QUESTION WEIGHTING

The weighting represents the weights that the teacher committees gave to each content area to be represented on the Regents examinations. The Expected Points for a Student with a Score of 65 and the Expected Points for a Student with a Score of 85 represent the relative difficulty of each test question listed for *just passing* and *just passing with distinction* students, respectively. These values are unweighted to provide educators with information about how relatively difficult each question is. To derive the weighted total, multiply each expected point value by the weight. A profile analysis can then be made for each student, by summing these products as described above. For example, a student performs above the *passing with distinction* level for all groups of items except the Key Idea of Operations on the June 2002 Mathematics A Regents Examination. On that section, the student scores below *just passing.* This information shows weaknesses in the student's understanding of the Mathematics A curriculum. Specifically, this student's performance reflects a relative weakness in understanding of Operations.

## APPROPRIATE ITEM ANALYSIS

The tables and information provided by the State Education Department will enable educators to perform item analyses for groups of students. The analyses will identify the Key Ideas in which the students are weak and in which they are strong. Analyses of individual questions can follow this component level analysis; which provide the reliability that is missing from simple analyses of each test question. Teachers or

administrators may go through their students' examinations and fill in the appropriate category in which each student's total raw score falls.

Table 2 (on the following page) shows a hypothetical class of 35 students who took the June 2002 Mathematics A Regents Examination. Teachers tallied each student's score in each of the Key Ideas and made a “tick” mark in the appropriate column. Based on these tallies this population of students appears relatively weak in the area of Uncertainty and Measurement. Teachers could use this information to begin evaluating the curriculum and program they are currently using, to identify interventions and strategies for improving performance.

The State Education Department intends to provide similar information for each test administration. This will enable educators to identify consistent patterns of strengths and weaknesses in a reliable manner and in a way that continually corrects for the normal fluctuations in the intrinsic difficulties of test questions.

## LIMITATIONS OF THIS ANALYSIS

Item analyses of this type are time consuming. It requires a person to organize the test questions into groups according to the item maps provided in the scoring guides. A sum of the point values achieved for the questions in each component must be computed. Then, those sums must be compared to the values provided for the appropriate examinations in the tables. Educators must balance the demands of these analyses on their time against the information they provide. Many schools, not realizing the problems of the unreliability of other types of item analyses, may be misled.

## INAPPROPRIATE ITEM ANALYSIS

Simple analyses of performance on test items present two major and related dangers. Most often, such analyses provide the proportions or percentages of students who answer each question correctly. For each item the proportions are then compared to each other and conclusions are made concerning student strengths and weaknesses. This is dangerous because it confounds student strengths and weaknesses with the intrinsic difficulties of the questions. Since test questions themselves vary in difficulty, what may appear to be a strength might in fact merely be a simple question. Not taking item difficulty into account precludes an analysis of strengths and weaknesses.

Item-by-item comparisons against a group which has similar, identified skills, e.g., all *just passing* students, addresses this concern. The following hypothetical situation illustrates why item difficulty needs to be incorporated into the analysis. Two groups of students are compared. One group is the students in a science class (this could be any population of students), and the other group is all of the students in the State who *just passed* this examination, e.g., who earned a score of 65. On a question concerning bedrock structure, 55 percent of the students from the science class answered the question correctly, while only 40 percent of the *just passing* students answered the

**Table 2.**
**June 2002 Mathematics A Examination Component Information**
Hypothetical Class Breakdown of 35 Students in Six of the Seven Key Ideas

| **Mathematical Reasoning** | | | | **Number and Numeration** | | | | **Operations** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| For Items | Number of Students whose Total Raw Scores Were: | | | For Items | Number of Students whose Total Raw Scores Were: | | | For Items | Number of Students whose Total Raw Scores Were: | | |
| | Below 1.79 | Between ←→ | Above 2.61 | | Below 2.20 | Between ←→ | Above 3.02 | | Below 11.03 | Between ←→ | Above 14.5 |
| 3 | IIIII | IIIII IIIII | IIIII | 11 | IIIII III | IIIII IIIII | IIIII | 6 | IIIII | IIIII IIIII | IIIII |
| 21 | | IIIII IIIII | IIIII | 24 | | IIIII IIII | III | 7 | II | IIIII IIIII II | I |
| | | | | | | | | 9 | | | |
| | | | | | | | | 16 | | | |
| | | | | | | | | 17 | | | |
| | | | | | | | | 18 | | | |
| | | | | | | | | 33 | | | |
| **Totals** | **5** | **20** | **10** | **Totals** | **8** | **19** | **8** | **Totals** | **7** | **22** | **6** |
| | | | | | | | | | | | |
| **Measurement** | | | | **Uncertainty** | | | | **Patterns and Functions** | | | |
| For Items | Number of Students whose Total Raw Scores Were: | | | For Items | Number of Students whose Total Raw Scores Were: | | | For Items | Number of Students whose Total Raw Scores Were: | | |
| | Below 11.77 | Between ←→ | Above 16.69 | | Below 2.28 | Between ←→ | Above 4.19 | | Below 8.19 | Between ←→ | Above 12.8 |
| 4 | IIIII | IIIII IIIII | IIIII | 2 | IIIII II | IIIII IIIII | II | 14 | IIIII | IIIII IIIII | IIIII |
| 5 | III | IIIII II | IIIII | 34 | IIIII IIIII | IIIII I | | 25 | I | IIIII IIIII III | I |
| 15 | | | | | | | | 29 | | | |
| 20 | | | | | | | | 32 | | | |
| 22 | | | | | | | | 35 | | | |
| 23 | | | | | | | | | | | |
| 28 | | | | | | | | | | | |
| 31 | | | | | | | | | | | |
| | | | | | | | | | | | |
| **Totals** | **8** | **17** | **10** | **Totals** | **17** | **16** | **2** | **Totals** | **6** | **23** | **6** |

bedrock structure question correctly. If the 55-percent-correct rate of the science class is looked at by itself, the conclusion might be that the class is deficient in its understanding of bedrock structure. However, if the 55-percent-correct rate is compared to the 40-percent-correct rate of the students at the *just passing* level, the students in the class can be seen as performing above the *just passing* level of all the State's students. These data indicate that the item was very difficult.

On a question concerning air currents, 70 percent of the students in the science class answered the question correctly, while 85 percent of the *just passing* students answered the same question correctly. These two pieces of information indicate that the science class does have some deficiencies in its understanding of air currents because the students performed below the level of the State's *just passing* students on that item. Having the achievement milestone (65) of the *just passing* students to compare against, schools now have some reference against which they can gauge their own students' achievement.

The second difficulty with a simple item analysis, as described in the above illustration, is that item-by-item analysis is simply not reliable enough to be a basis for any conclusions. Very simply, as the illustration shows, particular questions are themselves so variable that it is difficult to draw conclusions about students' strengths and weaknesses based on performance on individual questions. Rather, it is much better to evaluate performance on whole groupings of questions first and, only then, based on the picture that emerges from those groupings, to examine the performance on the questions that comprise the groupings, giving particular attention to the features of the questions themselves that might make them more or less difficult for the students.

## WHAT'S NEXT FOR ITEM ANALYSIS?

Programs are currently being piloted that could significantly decrease the pen-and-paper aspects of item analysis. The State Education Department is investigating the possibility of using uniform, scannable answer forms, which could help make this process much easier. Such forms would enable all of a school's student answer forms to be scanned in one location and the student responses to each question could be easily entered into the table shown in the "Appropriate Item Analysis" section. Once enough schools use the same scannable answer forms, a program to do this could be developed to automate the analytical process.

**Table A-1.**

**Regents Comprehensive Examination in June 2002**
**Component English Information**

| Session 1 | | | | | Session 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Item | Max Raw Points | Weight | Expected Points for a Student with a Score of | | Item | Max Raw Points | Weight | Expected Points for a Student with a Score of | |
| | | | 65 | 85 | | | | 65 | 85 |
| 1 | 1 | 1 | 0.81 | 0.96 | 1 | 1 | 1 | 0.34 | 0.76 |
| 2 | 1 | 1 | 0.97 | 1.00 | 2 | 1 | 1 | 0.73 | 0.94 |
| 3 | 1 | 1 | 0.90 | 0.98 | 3 | 1 | 1 | 0.72 | 0.94 |
| 4 | 1 | 1 | 0.79 | 0.96 | 4 | 1 | 1 | 0.44 | 0.83 |
| 5 | 1 | 1 | 0.85 | 0.97 | 5 | 1 | 1 | 0.47 | 0.85 |
| 6 | 1 | 1 | 0.92 | 0.99 | 6 | 1 | 1 | 0.59 | 0.90 |
| 7 | 1 | 1 | 0.93 | 0.99 | 7 | 1 | 1 | 0.79 | 0.96 |
| 8 | 1 | 1 | 0.99 | 1.00 | 8 | 1 | 1 | 0.74 | 0.95 |
| 9 | 1 | 1 | 0.98 | 1.00 | 9 | 1 | 1 | 0.44 | 0.83 |
| 10 | 1 | 1 | 0.96 | 0.99 | 10 | 1 | 1 | 0.74 | 0.95 |
| 11 | 1 | 1 | 0.87 | 0.98 | Essay 3 | 6 | 2 | 1.85 | 3.22 |
| 12 | 1 | 1 | 0.81 | 0.96 | Essay 4 | 6 | 2 | 3.25 | 4.33 |
| 13 | 1 | 1 | 0.95 | 0.99 | | | | | |
| 14 | 1 | 1 | 0.92 | 0.99 | | | | | |
| 15 | 1 | 1 | 0.94 | 0.99 | | | | | |
| 16 | 1 | 1 | 0.83 | 0.97 | | | | | |
| Essay 1 | 6 | 2 | 3.98 | 4.91 | | | | | |
| Essay 2 | 6 | 2 | 3.90 | 4.35 | | | | | |
| | | | | | | | | | |
| Weighted Total | 40.00 | | 30.18 | 34.21 | Weighted Total | 34.00 | | 16.19 | 23.99 |

**Table B-1.**
**June 2002 Mathematics Regents Examination Component Information**

| Mathematical Reasoning | | | | | Number and Numeration | | | | | Operations | | | | | Modeling/Multiple Representation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | Max Raw Points | Weight | Expected Points for a Student with a Score of | | Item | Max Raw Points | Weight | Expected Points for a Student with a Score of | | Item | Max Raw Points | Weight | Expected Points for a Student with a Score of | | Item | Max Points | Weight | Expected Points for a Student with a Score of | |
| | | | 65 | 85 | | | | 65 | 85 | | | | 65 | 85 | | | | 65 | 85 |
| 3 | 1 | 2 | 0.72 | 0.87 | 11 | 1 | 2 | 0.70 | 0.86 | 6 | 1 | 2 | 0.87 | 0.94 | 1 | 1 | 2 | 0.60 | 0.79 |
| 21 | 2 | 1 | 0.35 | 0.87 | 24 | 2 | 1 | 0.80 | 1.30 | 7 | 1 | 2 | 0.85 | 0.94 | 8 | 1 | 2 | 0.94 | 0.98 |
| | | | | | | | | | | 9 | 1 | 2 | 0.94 | 0.97 | 10 | 1 | 2 | 0.75 | 0.88 |
| | | | | | | | | | | 16 | 1 | 2 | 0.87 | 0.95 | 12 | 1 | 2 | 0.84 | 0.93 |
| | | | | | | | | | | 17 | 1 | 2 | 0.57 | 0.77 | 13 | 1 | 2 | 0.62 | 0.81 |
| | | | | | | | | | | 18 | 1 | 2 | 0.85 | 0.93 | 19 | 1 | 2 | 0.65 | 0.83 |
| | | | | | | | | | | 33 | 4 | 1 | 1.13 | 3.54 | 26 | 3 | 1 | 2.53 | 2.84 |
| | | | | | | | | | | | | | | | 27 | 3 | 1 | 2.10 | 2.58 |
| | | | | | | | | | | | | | | | 30 | 3 | 1 | 1.55 | 2.74 |
| Weighted Total | 4.00 | | 1.79 | 2.61 | Weighted Total | 4.00 | | 2.20 | 3.02 | Weighted Total | 16.00 | | 11.03 | 14.54 | Weighted Total | 21.00 | | 14.98 | 18.60 |

| Measurement | | | | | Uncertainty | | | | | Patterns/Functions | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | Max Raw Points | Weight | Expected Points for a Student with a Score of | | Item | Max Raw Points | Weight | Expected Points for a Student with a Score of | | Item | Max Raw Points | Weight | Expected Points for A Student with a Score of | |
| | | | 65 | 85 | | | | 65 | 85 | | | | 65 | 85 |
| 4 | 1 | 2 | 0.88 | 0.95 | 2 | 1 | 2 | 0.96 | 0.99 | 14 | 1 | 2 | 0.88 | 0.95 |
| 5 | 1 | 2 | 0.58 | 0.78 | 34 | 4 | 1 | 0.36 | 2.21 | 25 | 2 | 1 | 0.82 | 1.53 |
| 15 | 1 | 2 | 0.85 | 0.94 | | | | | | 29 | 3 | 1 | 2.57 | 2.89 |
| 20 | 1 | 2 | 0.81 | 0.91 | | | | | | 32 | 4 | 1 | 2.40 | 3.71 |
| 22 | 2 | 1 | 1.18 | 1.76 | | | | | | 35 | 4 | 1 | 0.64 | 2.81 |
| 23 | 2 | 1 | 1.36 | 1.79 | | | | | | | | | | |
| 28 | 3 | 1 | 0.95 | 2.56 | | | | | | | | | | |
| 31 | 4 | 1 | 2.04 | 3.42 | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| Weighted Total | 19.00 | | 11.77 | 16.69 | Weighted Total | 6.00 | | 2.28 | 4.19 | Weighted Total | 15.00 | | 8.19 | 12.84 |

This paper was produced by the Office of State Assessment.

**Primary author(s):** Justin Stuck and Gerald DeMauro

Office of State Assessment
775 EBA
89 Washington Ave
Albany, NY 12234

For technical questions regarding this paper, please contact Gerald DeMauro at:

(518) 474-5902.

## OBTAINING COPIES OF OSA TECHNICAL PAPERS

This report may be accessed on the Internet at: www.emsc.nysed.gov/ciai/testing/pubs.html. Hard copies may be obtained by writing to the Office of State Assessment at the address above.