

Understanding Question (Item) Level of Difficulty

New York State assessments are, by design, composed of items of varying difficulty. That is done to identify student performance levels. A process called bookmarking is used when developing the assessments. NYSED has a well-defined and rather involved procedure to do that, but it boils down to this:

1. Questions are arranged in order of difficulty using the p-values measured in field testing.
2. Teacher evaluation committees determine the levels 2, 3, and 4 cut points by identifying the most difficult question that two-thirds of the students at each proficiency level would be expected to answer correctly.
3. Cut scores are calculated by adding up all the points from the easiest question (the one with the highest p-value) to the question identified to be at each of the cut points (from step 2).

That is all done before the fact, i.e., prior to the assessment administration. NYSED does not indicate the question levels in any of its documentation, but we can look at aggregated student score results to **reverse-engineer the level-setting**.

Most RICs have a Regional Benchmark Cutpoint Report that will allow one to go through this process. WNYRIC's looks like this:

Regional Benchmark (Cutpoints) - Regents - MC Analysis

This report shows the percentage of students who got each question correct at various cutpoint levels, for the selected school year and assessment.

2017		Assessment: Regents Phy Set/Earth Sci - Jun										
Key Idea	Performance Indicator	Question	District Success Rate	Regional Success Rate (N=33098)	Regional High 1 SS=54 RS=41 (N=269)	Regional Low 2 SS=55 RS=42 (N=293)	Regional High 2 SS=64 RS=50 (N=360)	Regional Low 3 SS=65 RS=51 (N=487)	Regional High 3 SS=84 RS=74 (N=981)	Regional Low 4 SS=85 RS=78 (N=789)	Regional High 4 SS=100 RS=101 (N=44)	
The Earth and celestial phenomena can be described by principles of relative motion and perspective.	Describe current theories about the origin of the universe and solar system.	A-09	77%	68%	47%	50%	54%	55%	75%	74%	100%	
		A-12	85%	76%	52%	53%	64%	66%	85%	88%	100%	
		A-26	64%	47%	14%	13%	21%	20%	54%	57%	100%	
		B1-39	63%	54%	32%	38%	36%	39%	56%	58%	100%	
		B1-41	65%	53%	23%	28%	36%	36%	58%	63%	100%	
		B1-42	62%	44%	19%	18%	16%	17%	49%	53%	100%	

We're mainly concerned with three columns in this thing; the item number, the low 3 cut values, and the low 4 cut values. (If you have an urge to figure out the difference between Level 1 and Level 2 question, you could use the low 2 cut values as well, but I'm not certain that really provides much particularly useful information.)

In the low 4 column, highlight all of the values that are LOWER than .67. Those will be your Level 4 questions — questions intended to be challenging for those top students. Do the same thing with the low 3 column — don't change the questions already marked as Level 4 questions. This second set of questions, the Level 3 questions, will be challenging for students at Level 3. Everything that's not highlighted is either a Level 2 or a Level 1 question. Proficient students should be able to answer most of the Level 1 and Level 2 questions without too much difficulty. I've included a sample on the back of this sheet.

(slightly) **Technical description of the process**

The basis for this slight-of-hand is the premise that a student at the cut-point will have a better than two out of three chance of correctly answering all of those questions to and including the question at the cut-point. This report shows the aggregated results of all of the students that scored exactly at the cut-points. While it's not exactly the same thing, we're looking at the population of students scoring at the cut-point and identifying the questions that two-thirds of them answered correctly.

To create your own (potentially far more attractive) version of this report, you could use as a dataset the following columns from your RIC's Regional Benchmark Cutpoint Report: Item number, Regional p-value, Low 2, Low 3, and Low 4 (and another for Low 5 if you're looking at a common core regents exam).

1. Combine CR and MC into the same chart, sorted by item number (just for tidiness...it really makes no difference)
2. Identify all items in the Low 4 column with values <.67; color them blue. Those are the Level 4 items.
3. Identify all items in the Low 3 column with values <.67, excluding those already identified as Level 4 items in step 2; color those green. Those are the Level 3 items.
4. Identify all items in the Low 2 column with values <.67, excluding those previously identified as Level 3 or Level 4 items. Color those a different color. That takes care of the Level 2 items.
5. Everything else is a Level 1 item.

Understanding Question (Item) Level of Difficulty

Regional Benchmark (Cutpoints) - Regents - MC Analysis

This report shows the percentage of students who got each question correct at various cutpoint levels, for the selected school year and assessment.

2017		Assessment: Regents Phy Set/Earth Sci - Jun									
Key Idea	Performance Indicator	Question	District Success Rate	Regional Success Rate (N=33098)	Regional High 1 SS=54 RS=41 (N=269)	Regional Low 2 SS=55 RS=42 (N=293)	Regional High 2 SS=64 RS=50 (N=360)	Regional Low 3 SS=65 RS=51 (N=487)	Regional High 3 SS=84 RS=74 (N=981)	Regional Low 4 SS=85 RS=78 (N=789)	Regional High 4 SS=100 RS=101 (N=44)
The Earth and celestial phenomena can be described by principles of relative motion and perspective.	Describe current theories about the origin of the universe and solar system.	A-09	77%	68%	47%	50%	54%	55%	75%	74%	100%
		A-12	85%	76%	52%	53%	64%	66%	85%	88%	100%
		A-26	64%	47%	14%	13%	21%	20%	54%	57%	100%
		B1-39	63%	54%	32%	38%	36%	39%	56%	58%	100%
		B1-41	65%	53%	23%	28%	36%	36%	58%	63%	100%
		B1-42	62%	44%	19%	18%	16%	17%	49%	53%	100%
		A-06	79%	70%	50%	51%	54%	54%	76%	80%	100%
		A-07	58%	49%	31%	32%	33%	34%	51%	48%	100%
		A-11	78%	64%	28%	32%	42%	41%	77%	77%	100%
		A-14	80%	64%	29%	35%	40%	43%	74%	78%	100%
Explain complex phenomena, such as tides, variations in day length, solar insolation, apparent motion of the planets, and annual traverse of the constellations.		A-32	55%	48%	25%	24%	37%	35%	51%	54%	100%
		B1-36	84%	76%	59%	63%	65%	67%	82%	82%	100%
		B1-37	77%	69%	38%	40%	51%	47%	79%	80%	100%
		B1-38	71%	56%	27%	32%	37%	36%	61%	63%	100%
		A-15	95%	87%	63%	60%	83%	81%	97%	98%	100%
		A-17	83%	64%	30%	31%	42%	45%	77%	78%	100%
		A-21	78%	64%	30%	30%	42%	42%	80%	77%	100%
		A-24	75%	62%	41%	38%	46%	47%	69%	71%	100%
		A-25	86%	78%	52%	59%	65%	66%	87%	87%	100%
		A-28	84%	76%	55%	55%	67%	67%	83%	84%	100%
The observations made while testing proposed explanations, when analyzed using conventional and invented methods, provide new insights into phenomena.	The observations made while testing proposed explanations, when analyzed using conventional and invented methods, provide new insights into phenomena.	A-29	79%	65%	42%	43%	50%	46%	75%	78%	100%
		A-33	70%	56%	31%	29%	35%	35%	64%	69%	100%
		A-34	75%	62%	38%	31%	40%	38%	73%	72%	100%
		B1-43	60%	45%	11%	11%	19%	21%	56%	58%	100%
		B1-45	87%	75%	43%	45%	56%	61%	89%	93%	100%