From: Susan Pareigis
Sent: Friday, December 02, 2011 12:54 PM
To: Robinson, Gerard
Cc:
Subject: Florida Council of 100 issues re: FCAT 2.0 and Alg I EOC standard setting
Attachments: FCAT_2 0_VS .xlsx

Importance: High

Commissioner Robinson:

After participating in two Reactor panels, viewing the SBE workshop on November 14, and attempting to obtain answers from DOE and Hillsborough County staff to multiple questions about the standard-setting process, methodologies, and results, the Council of 100 firmly believes that DOE must address the following data points and issues before providing the State Board of Education with a set of final cut score recommendations

Due to the lack of specific answers, to our specific questions, we were unable to vote on Tuesday as recorded and subsequently issued by DOE.
It would be very beneficial to receive the answers in writing today, or over the weekend, and before the SBOE workshop on Monday. Please feel free to share the answers with the full board in your response to the Council of 100.

Thank you in advance for your consideration,
Susan

Overarching Principles

1. DOE must ensure that the methodology for setting the FCAT 2.0 standards is quantitative, defensible (both statistically and from a policy perspective), and applied consistently to all aspects of the process. Further, the methodology must be transparent and readily explainable to students, parents, and state policy makers.

2. In this standard-setting process, every percentage point – every partial percentage point -- counts. Thus, precision is important, as are descriptions of the variability around every average and in every statistical analysis. DOE must ensure that it reports and explains such variability to the SBE.

3. The fact that Florida will be switching to PARCC in a few years is no justification for maintaining low standards or lowering standards for the current cohorts of students who will not be affected by PARCC. Neither those students nor the colleges that will be educating them and/or the businesses that will be hiring them can afford anything but the highest of standards to drive their achievement. Yes, changing performance expectations for these students mid-stream might be temporarily painful, but does a parent withhold an antibiotic from a sick child because it might taste nasty for a little while, thus prolonging the illness and exposing more family members and innocent bystanders to the contagion? And consider this – what if something unexpected happens and PARCC never comes to fruition in 3 years, either falling apart, getting watered down, or taking much longer to be put in force? Lastly, if there are concerns about the domino effect the cut score decisions will have on the school grading process, those issues should be dealt with within the framework of that process.

4. Inertia should never be a barrier to improving policy. If a rule needs to be changed, the SBE should change it. If statute needs to be changed, DOE should actively pursue legislation to make the necessary changes.

Reading Cut Scores

1. Sharon Koon told Steven Birnholz that the theta value analysis should be controlling over the cut score line "smoothing" and the impact analysis (although one could argue that the theta value analysis and the impact analysis are two sides of the same coin). Based on her explanation, here's how to read the attached theta value charts provided by DOE – (1) The midpoint of the distribution is at a theta of 0. For Grade 3 Reading, that mean is at 200. However, for this grade as well as the others, there is a standard deviation of about 20. That means, for example, that for Grade 3 Reading, the expected score was 198 +/- 20 (10%). The point is that there is variability in those predicted theta values that should be explained to the SBE. (2) As the theta values decrease, it means that less ability is required to pass the test and, thus, more students would be expected to pass the test. Currently, all the Reading theta values are below 0 (the midpoint of the distribution), rather than clustering around the midpoint or being higher than the midpoint (i.e., incenting higher performance than the "average" student). (Math cut score theta values are also below 0, except for Grade 6 which is at 0.) What is the policy rationale for opting for a clustering of cut scores below the midpoint of the distribution?

2. As voiced by both DOE staff and participants at Reactor Panel #1 and at the November SBE workshop, a key goal of the FCAT 2.0 cut-score-setting process is to ensure that Florida students are ready for the PARCC exams in 3 years. As a result, beginning with Reactor Panel #1, FC100 has been asking for DOE to provide a competitive benchmark, based on Massachusetts (the lead PARCC state and generally recognized top state educationally in the nation) or the nation as a whole, against which Florida could assess its current performance status and determine the FCAT 2.0 cut scores needed to ensure a smooth transition to PARCC cut scores.

Publicly, DOE has asserted that, based on USDOE research "Mapping State Proficiency Standards Onto the NAEP Scales: Variation and Change in State Standards for Reading and Mathematics, 2005-2009," that, because Florida is second in the nation in terms of the rigor of its Grade 8 Reading standards (higher than MA), the past Grade 8 FCAT Reading performance should be used as a standard and an "anchor" for determining other FCAT cut scores, especially in Reading. Because of this connection, DOE noted at Reactor Panel #1 that the then-upcoming release of 2011 NAEP scores would be critical. Thus, if what DOE said is true, then it is important to note that Florida's national performance rankings dropped in all categories (Grades 4 and 8, Reading and Math).

In fact, the average Florida NAEP Grade 8 Reading score dropped from 264 to 262 (30th in the nation to 34th in the nation), while the U.S. average increased from 262 to 264 and the MA average increased from 274 to 275. That's a net change from the conditions under which DOE made its Reactor Panel 1 recommended linkage of 1.5% versus the U.S. average and 1.1% versus the MA average. Further, the % Proficient in Florida dropped from 29% to 27% while it increased from 28% to 29% for the U.S. and from 37% to 40% for MA. And the % Below Basic in Florida increased from 24% to 27% while it decreased from 26% to 25% for the U.S. and from 17% to 16% for MA. What this indicates is two-fold: (1) Our students are increasingly unprepared for the rigors of NAEP, and, thus, we need to toughen our standards – i.e., make it harder to get a higher score. (2) Based on the NAEP score decrease, one could argue that the 243 cut score needs to be increased anywhere from:

- 0.8% (245 -- the drop in FL average score)
- 2.0% (248 – the drop in FL % Proficient)
- 3.0% (250 -- the increase in FL % Below Basic)

- 1.5% (247 -- the net loss in average score, FL vs. US)
- 3.0% (250 – the net loss in % Proficient, FL vs. US)
- 4.0% (253 – the net loss in % Below Basic, FL vs. US)

- 1.1% (246 -- the net loss in average score, FL vs. MA)
- 4.0% (253 – the net loss in % Below Basic, FL vs. MA)
- 5.0% (255 -- the net loss in % Proficient, FL vs. MA)

That being said, at Reactor Panel #2, Sharon Koon told Susan Pareigis and Steven Birnholz that she had run a regression analysis looking at the correlation between the Mapping Study's NAEP scale equivalent scores to the state standards and the actual state NAEP performance scores and found little correlation. When asked about this by Susan during the meeting, Sharon said that Susan must have been mistaken. FC100 then ran its own regression analysis on the data and confirmed Sharon's findings. In fact, when looking at 2009 performance data for Grade 8 Reading, r-squared was only 0.035 (scale of 0 to 1, with 1 indicating the strongest predictive power), and, when looking at 2011 performance data, r-squared was only 0.025.

Additionally, FC100's review of the Study itself found that, while the relative error around Florida's Grade 8 Reading NAEP scale equivalency score is less than 0.5, it should be noted that the Study's methodology (equipercentile mapping) "could be applied to any set of numbers, whether or not they are meaningfully related. Additional data, beyond the percentage meeting the standard in the state and the distribution of NAEP score—the only data used in the computation—are needed to test the validity of the mapping." This appears to contribute to the Study's disclaimer in its "Cautions in Interpretation" section that,

> "As the earlier mapping reports pointed out (McLaughlin et al. 2008a, 2008b; National Center for Education Statistics 2007; Bandeira de Mello, Blankenship, and McLaughlin 2009), the mapping methodology has several caveats that need to be noted….This report is not an evaluation of state assessments. State assessments and NAEP are developed for different purposes and have different goals and they may vary in format and administration. Findings of different standards, different trends, and different gaps are presented without suggestion that they be considered as deficiencies either in state assessments or in NAEP. The analyses in this report do not address questions about the content, format, exclusion criteria, or conduct of state assessments, as compared to NAEP. State assessments and their associated proficiency standards are designed to provide pedagogical information about individual students to their parents and teachers, whereas NAEP is designed to provide performance information at an aggregate level. Also, the analyses do not address any change in states' assessments or proficiency standards that may have occurred after 2009. Mapping the various state proficiency standards on the NAEP scale and comparing the standards with NAEP achievement levels gives context to the discussion, but it does not imply that the NAEP achievement levels are more valid than the state standards or that states should emulate NAEP standards. There is a wide range of policy considerations involved in setting achievement standards, and what is appropriate for NAEP may not be the best fit for a given state. NAEP's achievement levels are used to interpret the meaning of the NAEP scales. NCES has determined (as indicated by NAEP's authorizing legislation) that NAEP achievement levels should continue to be used on a trial basis and should be interpreted with caution."

(On the other hand, the Study also notes that, "A measure of the appropriateness of the mapping is the correlation coefficient showing the relationship between the percentages reported for schools by the state and those estimated from the NAEP scale equivalents: the two assessments must agree on which schools are high achieving and which are not. With a correlation coefficient of 0.7, Florida is one of 22 states (for Grade 8 reading) that had state assessment results that were highly correlated with NAEP.)

Lastly, DOE and Hillsborough County staff are now privately asserting that there are limitations on the value of NAEP scores for gauging Florida performance (e.g., participation rates by state, ELL and ESE participation rates by state, only testing grades 4 and 8, little-to-no correlation between proficiency standard ratings and student performance).

Unfortunately, this leaves (1) a cleavage between DOE's public and private positions on the value of NAEP as a benchmarking tool, and (2) If NAEP cannot be a valid, accurate, and reliable benchmark, no meaningful way to determine whether, and/or the degree to which, the proposed FCAT 2.0 cut scores will help bridge the gap with PARCC and make Florida more competitive vis-à-vis Massachusetts or the nation as a whole. Now, the only tie to an external test score is to the SAT concordance threshold currently in rule (which is problematic in and of itself – see below), i.e., Florida is still comparing our performance to itself. Further, at least during this process, there hasn't been any analysis looking at concordance between Florida's and other states' performance.

3. Based on Reactor Panel #2, DOE is putting a great emphasis on the "PSAT to FCAT Linking Strategy" as a key way to determine the Grade 10 Reading Level 3 cut score. However, there are several concerns with this approach:

- With all the talk of using concordance tables (e.g., FCAT to SAT-10, PSAT to FCAT, FCAT to SAT, FCAT to ACT) to set FCAT cut scores, Susan asked in Reactor Panel #2 why the state shouldn't just require the PSAT, SAT, ACT, and SAT-10 since they concorded so closely with the FCAT. The response was that the PSAT, SAT, ACT, and SAT-10 are very different types of tests than the FCAT and test very different things. Thus, the question is, <u>if those tests are so different from the FCAT, why is the state's primary way to set FCAT cut scores based on concording to those tests? While such concordance might be able to be done via a bunch of statistical machinations, it doesn't make sense from a policy perspective.</u>

- This concordance approach is not being used with regard to Math cut scores.

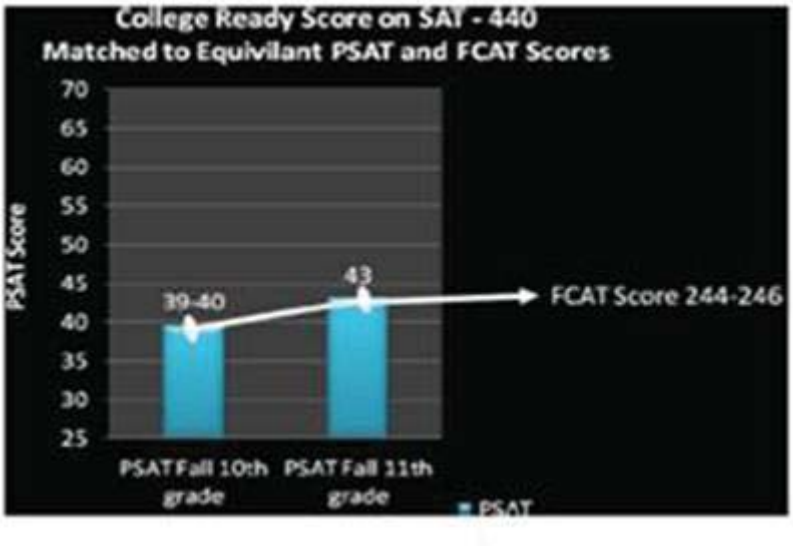
- The Grade 10 Reading concordance target being advocated by DOE and Hillsborough County is an SAT of 440. However, an SAT of 440 is only the 30th percentile of national performers. While that might be the concordance

target currently in rule for "college readiness" (relating to common placement and PERT, for which DOE says it will be 2-3 years before there's good data for evaluation), is the 30th percentile really what we want our students to aim for? Additionally, there was much discussion at Reactor Panel #2 about the many different definitions (formal and informal) of "college readiness" used by the state and how such discrepancies are weakening the state's ability to set high and consistent achievement targets. For example, it was generally agreed at Reactor Panel #2 that an SAT of at least 500 (51st percentile) is necessary to have success in college rather than to just get into college.

- If Hillsborough-based data is going to be the key driver of cut scores, then that data should be adjusted to reflect the student characteristics the state as a whole. For example, in 2011 Grade 10 Reading FCAT, Hillsborough's mean scale score (306) was 3 points (1%) lower than the state average (309); its % passing (57%) was 3 percentage points below the state average (60%); and its "3 or higher" percentage (37) was 2 percentage points lower than the state average (39). Additionally, for 10th graders taking the 2011 Reading FCAT, here are some of the demographic differences:

	State	Hillsborough
ELL	5.1%	6.7%
Migrant	0.5%	0.9%
Section 504	2.0%	3.9%
Free or Reduced Lunch	48.7%	47.7%
Total ESE Other Than Gifte	10.7%	10.0%

- The Hillsborough-based methodology apparently uses imprecise steps such as one person's "rules of thumb" [e.g., (PSAT score + 1) x 10 = SAT predicted score]. Thus, FC100 has asked DOE to provide us with the full statistical analysis (in electronic format) that resulted in the following slide, including all related variability information. This analysis should include a detailed description of each step in the analysis as well the results of each step and not round any number involved to fewer than 2 decimal points.



- Based on the great emphasis DOE is putting on the following "FLDOE FCAT/SAT Concordance Table" as a key way to determine the Grade 10 Reading Level 3 cut score, FC100 is concerned that the interpretation of the tables is not as precise as it could be. Specifically, while (1) a 244 on FCAT 2.0 might equate to a range of 307-310 on FCAT 1.0; and (2) the cumulative percentage at the 310 mark is 31.7706, which falls below the cumulative percentage of 34.42543 at the SAT score of 440; in fact, the cumulative percentage on the FCAT scale that most closely falls below the cumulative percentage of 34.42543 at the SAT score of 440 is 34.30824, i.e., the cumulative percentage for an FCAT 1.0 score of 313. Thus, we have asked DOE to provide the equivalent FCAT 2.0 score for an FCAT 1.0 score of 313, and asked that they not round any number involved to fewer than 2 decimal points.

FCAT 2009 Concordance Study

FCAT SSS Reading

	Frequency	Percent	Valid Perce	Cumulative Percent
300	982	0.697443	0.697443	23.99787
301	1004	0.713068	0.713068	24.71094
302	992	0.704545	0.704545	25.41548
303	1010	0.71733	0.71733	26.13281
304	1101	0.78196	0.78196	26.91477
305	1113	0.790483	0.790483	27.70526
306	1135	0.806108	0.806108	28.51136
307	1167	0.828835	0.828835	29.3402
308	1090	0.774148	0.774148	30.11435
309	1124	0.798295	0.798295	30.91264
310	1208	0.857955	0.857955	31.7706
311	1201	0.852983	0.852983	32.62358
312	1135	0.806108	0.806108	33.42969
313	1237	0.878551	0.878551	34.30824
314	1216	0.863636	0.863636	35.17188
315	1229	0.872869	0.872869	36.04474
316	1239	0.879972	0.879972	36.92472
317	1236	0.877841	0.877841	37.80256
318	1233	0.87571	0.87571	38.67827
319	1252	0.889205	0.889205	39.56747
320	1304	0.926136	0.926136	40.49361
321	1237	0.878551	0.878551	41.37216
322	1278	0.90767	0.90767	42.27983
323	1257	0.892756	0.892756	43.17259
324	1277	0.90696	0.90696	44.07955
325	1280	0.909091	0.909091	44.98864
326	1274	0.90483	0.90483	45.89347
327	1301	0.924006	0.924006	46.81747
328	1323	0.939631	0.939631	47.7571

SAT Scores

	Frequency	Percent	Valid Perce	Cumulative Percent
420	5496	3.903409	3.903409	26.95241
430	4284	3.042614	3.042614	29.99503
440	6238	4.430398	4.430398	34.42543
450	5614	3.987216	3.987216	38.41264
460	5368	3.8125	3.8125	42.22514
470	6583	4.675426	4.675426	46.90057
480	5031	3.573153	3.573153	50.47372

Excerpt from FLDOE produced FCAT/SAT/ACT concordance score table distributed by FLDOE Assessment Office 10/31/11

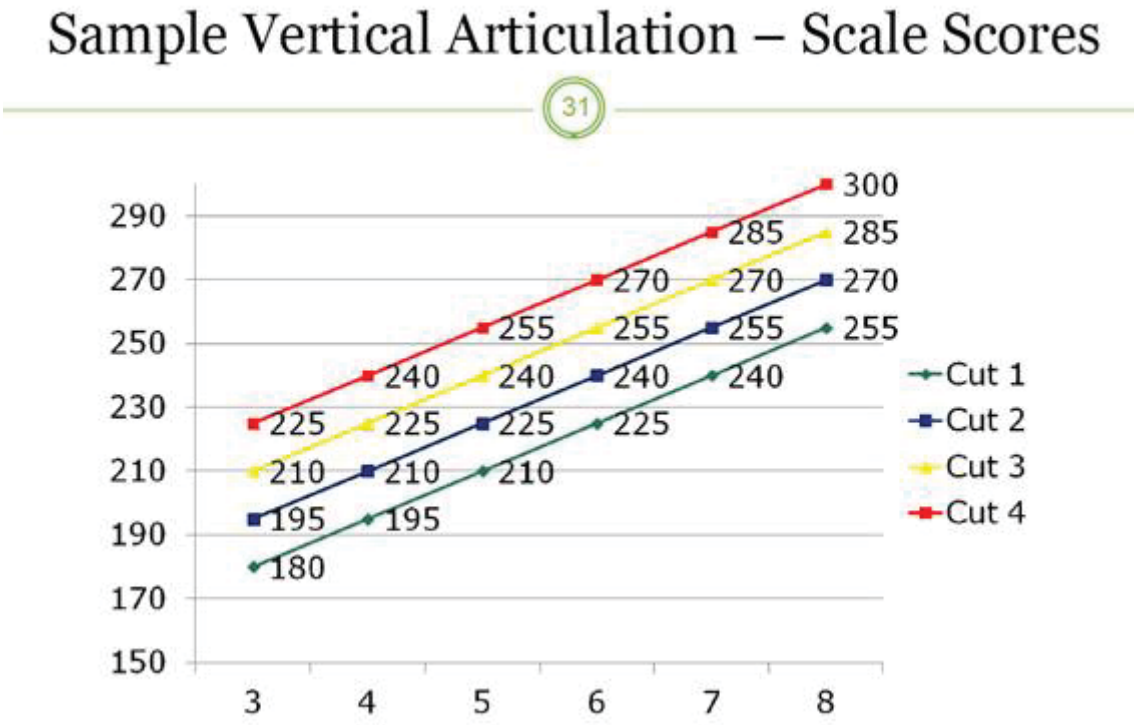
4. If a Grade 10 Reading cut score other than 243 is selected, according to DOE pronouncements, the cut scores for other grades (especially Grades 8 and 9) should be adjusted to ensure "consistency." According to Kris Ellington, Sharon says "consistency" means that "a student who maintains the same ability level over time and across grade levels will maintain their standing on the vertical scale and in relationship to achievement level cut scores." Sharon told Steven that the theta level analyses (attached) best demonstrate this concept. For example, in Reading, the yellow boxes are clustered around a couple of theta levels, meaning the cut scores are fairly consistent across grade levels. (Ironically, the yellow boxes in the Math chart are not as tightly clustered, and thus the cut scores are less consistent across grade levels. In fact, Sharon writes, "If anything, math could use a few adjustments.") Sharon has confirmed to Susan and Steven that, based on the principle of "consistency," a change in the proposed Grade 10 Reading cut score should, in theory, be followed by appropriate adjustments to the cut scores of other grades in order to maintain the current level of theta-level clustering.

DOE has placed a premium on ensuring "consistency" among the cut scores. In fact, the Commissioner adjusted a certain Reactor Panel #1-recommended Grade 8 Reading cut score simply to "achieve consistency." If "consistency" continues to be a primary methodological driver, then such principles should be rigorously applied to all subjects, grade levels, and achievement levels.

Another situation for which the issue of "consistency" crops up is when someone pronounces that the decisions of the Educator Panel, made-up of "true experts," should be religiously followed. However, if that were the case, then that would necessitate adopting the Educator Panel's recommendations lock, stock, and barrel – including the "inconsistent" cut scores that Reactor Panel #1 was strongly pushed to "smooth out." (Note: According to DOE staff, there is no quantified definition of the term "smoothing." In fact, at the Reactor Panel #1, smoothing was accomplished by the group looking at the cut score lines on an overhead project screen and moving the data points up and down until the lines looked straight.)

Lastly, DOE has yet to clearly explain, in lay-language, why the slopes of the Reading cut score lines decrease starting in Grade 8, or what that tells us about the rigor of the standards, the difficulty of the tests, and/or the ability of the students beginning in Grade 8. Note that the same slope change is not seen in the Math cut score lines. For this reason, below is a reproduction of the previously submitted slope analysis –

The Reactor Panel was shown the following slide on a screen at the end of the room and then proceeded to adjust the cut scores from the Educator Panel to reflect the same "consistency," i.e., slope on the example slide. The process was done based on "eye-balling." In short, while the slopes of the cut score trend lines appeared to consistent based on the images shown, they're not. Below, are some graphs demonstrating this and some charts providing an estimate of how far the proposed cut scores differ than the proposed Grades 3-7 trend lines.

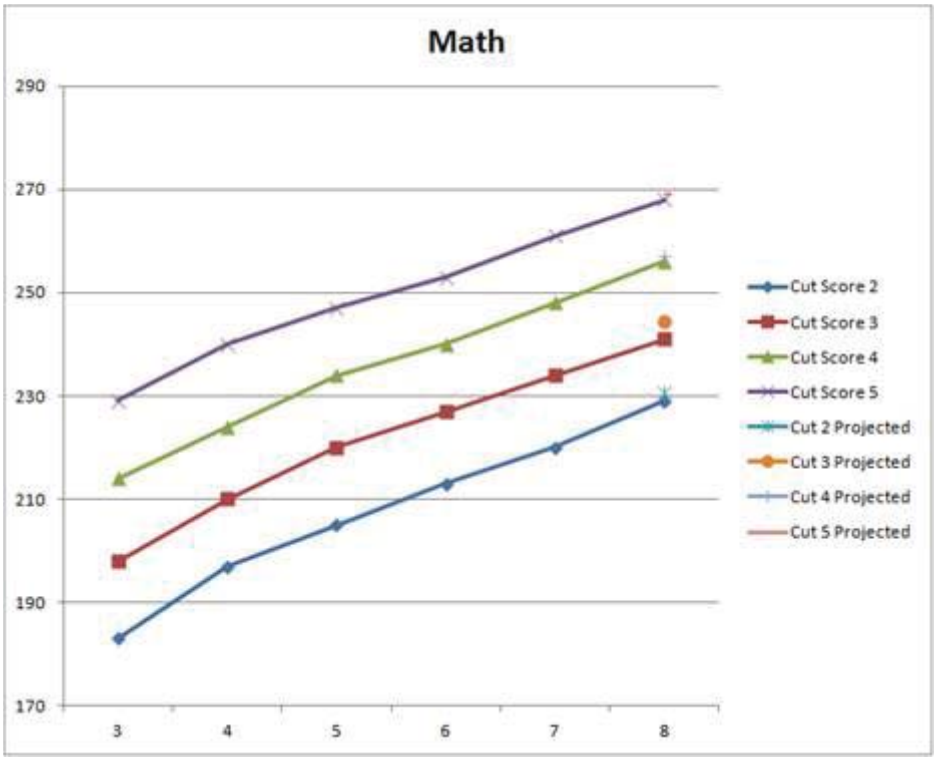


Source: Florida Department of Education, Office of Assessment, "2011 FCAT 2.0 Reading, Mathematics, and Algebra 1 End-of-Course Assessment Standard Setting," Rule Development Workshops, October 10-12, 2011.

It appears that the Math adjustments held pretty close to the "ideal" shown above. The "projected" Grade 8 cut scores are close to the "actual," as evidenced by the raw score difference and the graph showing the projected and actual 8th grade scores virtually on top of each other.

MATH

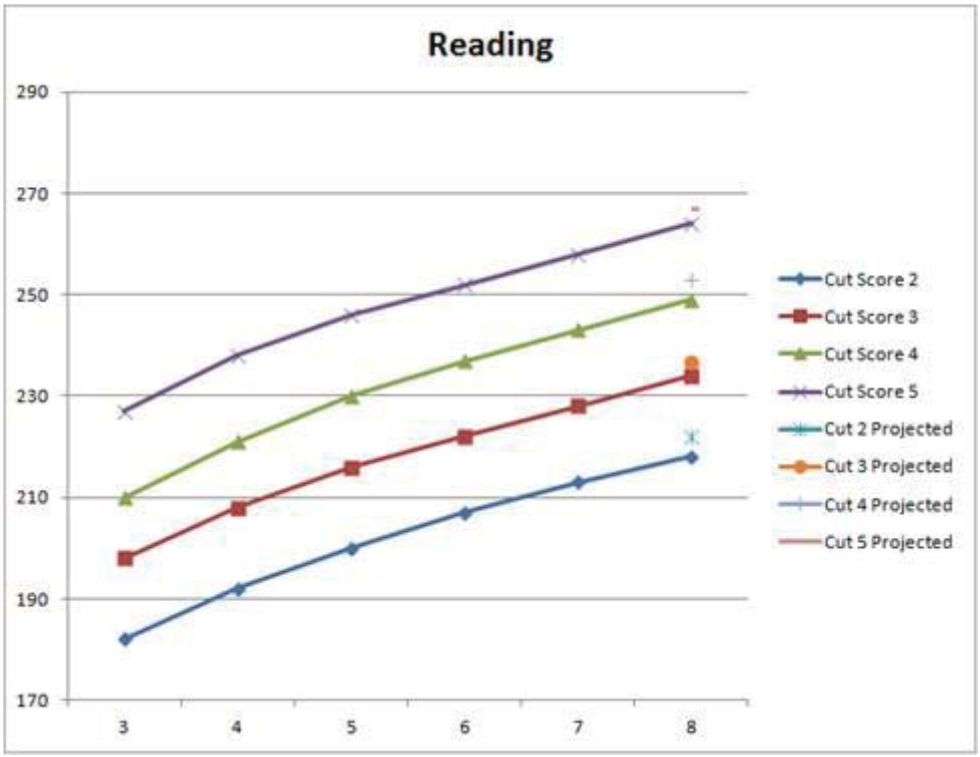
	Cut Score 2	Cut Score 3	Cut Score 4	Cut Score 5	Cut 2 Projected	Cut 3 Projected	Cu
3	183	198	214	229			
4	197	210	224	240			
5	205	220	234	247			
6	213	227	240	253			
7	220	234	248	261			
8	229	241	256	268	230.6	244.5	
				Difference from the Cut Score	1.6	3.5	

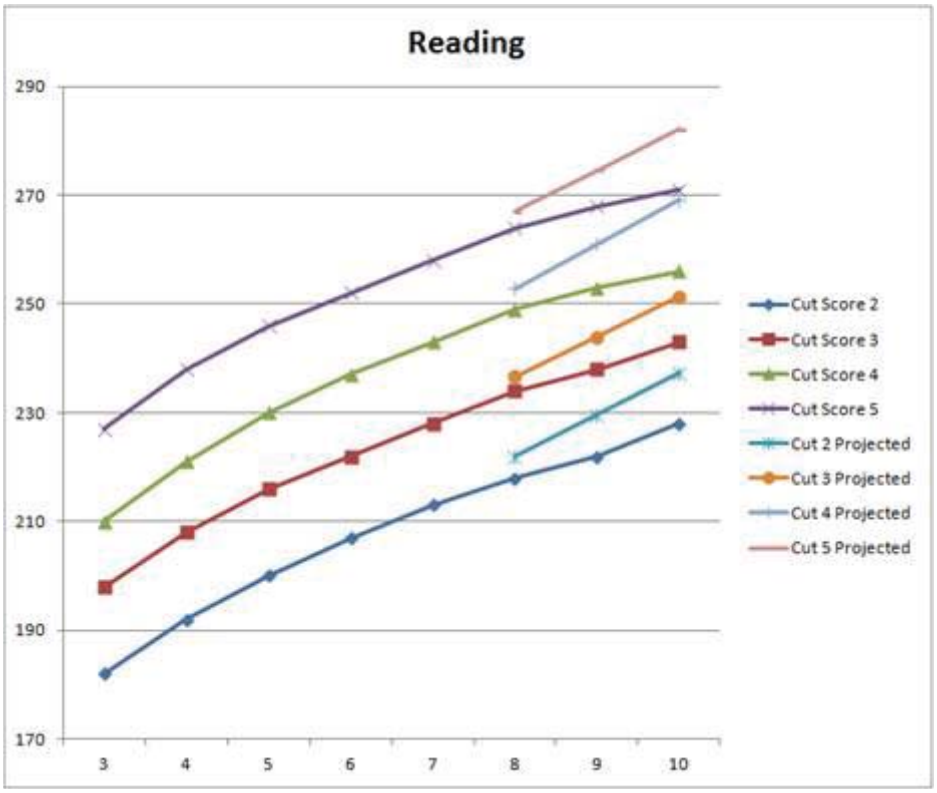


The same does not appear to be true for Reading cut scores. It does not appear that the Reading adjustments held close to the "ideal" shown above. In fact, the "projected" Grade 8 cut scores are not as close to the "actual" as the math scores, as evidenced by the raw score difference and the graph showing the projected and actual 8th grade scores. Further, a linear projection of the Grades 3-7 Reading cut scores indicates a substantial variation from the ideal slope. The following chart indicates how much the current proposed cut scores for Grades 8-10 differ from the trend established by the Grades 3-7 cut scores.

READING

	Cut Score 2	Cut Score 3	Cut Score 4	Cut Score 5	Cut 2 Projected	Cut 3 Projected
3	182	198	210	227		
4	192	208	221	238		
5	200	216	230	246		
6	207	222	237	252		
7	213	228	243	258		
8	218	234	249	264	221.9	236.6
9	222	238	253	268	229.6	244
10	228	243	256	271	237.3	251.4
			Difference from the Cut Score	Grade 8	3.9	2.6
				Grade 9	7.6	6
				Grade 10	9.3	8.4





5. Many assert that FCAT 2.0 is a “more rigorous” test. However, if FCAT 2.0 is harder based on harder standards, the failure rate should be higher than under FCAT 1.0. In fact, 2011 performance on the Reading FCAT 2.0, graded on the FCAT 1.0 scale, was nearly identical to performance on the FCAT 1.0 in 2010.

	FCAT 1.0	FCAT 2.0
8th Grade Mean Scale Score	312	313
8th Achievement Level 3 or Above	55	55
9th Grade Mean Scale Score	317	317
9th Achievement Level 3 or Above	48	48
10th Grade Mean Scale Score	310	309
10th Achievement Level 3 or Above	39	39

Algebra I EOC

1. FC100 has asked the following question of the Commissioner –

Section 1008.22, F.S., provides that, "for students entering grade 9 during the 2010-2011 school year and who are enrolled in Algebra I or an equivalent, each student's performance on the end-of-course assessment in Algebra I shall constitute 30 percent of the student's final course grade. Beginning with students entering grade 9 in the 2011-2012 school year, a student who is enrolled in Algebra I or an equivalent must earn a passing score on the end-of-course assessment in Algebra I or attain an equivalent score as described in subsection (11) in order to earn course credit." For other students taking the Algebra I EOC in 2010-11, the districts were allowed to decide what weight, if any, to put on earning a passing score. Thus, some students took the Algebra I EOC under "high stakes" conditions, and some students took it under "lower stakes" conditions.

Why is that important when analyzing the 2010-11 Algebra I EOC performance data? First, a student's motivational level when taking a test can significantly affect the student's performance – especially with regard to the Algebra I EOC. For example, after the 2010 field test of the EOC for which the average number of correct answers was 8 out of 30, Kris Ellington wrote to Eric Smith and Frances Haithcock, "We have the item statistics from the field test. It appears that the students weren't motivated to do well (average number correct out of 30 was 8!). We have assembled tests from these items but will need to do some quick post-equating based upon how students perform when motivated." Second, as noted above, only those entering 9th grade in the 2010-11 year would have 30% of their final grade depend on the EOC score. Those students, though, represent only 54% of the test taking population. For the rest of the population, the districts were given the latitude to decide what value, if any, to put on the test. This includes 33% of the population consisting of 6th-8th graders who skewed the results higher, and 13% of the population consisting of 10th-12th graders who skewed the results lower. Hillsborough is an example of a district with a ton of 8th graders taking the test because it pushes for Algebra to be taken in 8th grade.

Thus, as we (and others) recommended during the Reactor Panel #1, the Florida Department of Education should at least analyze the performance of the 2010-11 test-takers based on the level of stakes under which each test-taker took the test in order to determine if there is a statistically significant difference among the scores of each of the groups (based on the level of stakes), when adjusting for factors relating to the ability of the test-taker such as grade level. Furthermore, if any statutory provisions relating to the Algebra I EOC or other EOCs (e.g., the transition of the test being worth 30% of the final grade to being required for course credit) are anticipated to have a material impact on 2010-11 student performance or performance going forward, DOE should attempt to clarify the impact(s) of such provisions on the cut-score-setting process and recommend statutory options for mitigating those impacts.

2. If, for the above reason or any other reason, DOE recommends to stick to the Educator Panel's cut score recommendations, the SBE should be shown the variability around those recommendations (i.e., the graph with the averages and the error bars).

A	20	19.19	19.37	19.41	19.64	20.25	20.67	18.82
B	200	211.33	219.07	224.99	230.32	235.52	239.63	244.87
Grade 3 logit scale	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10
-3.00	140	154	161	167	171	175	178	188
-2.95	141	155	162	168	172	176	179	189
-2.90	142	156	163	169	173	177	180	190
-2.85	143	157	164	170	174	178	181	191
-2.80	144	158	165	171	175	179	182	192
-2.75	145	159	166	172	176	180	183	193
-2.70	146	160	167	173	177	181	184	194
-2.65	147	160	168	174	178	182	185	195
-2.60	148	161	169	175	179	183	186	196
-2.55	149	162	170	175	180	184	187	197
-2.50	150	163	171	176	181	185	188	198
-2.45	151	164	172	177	182	186	189	199
-2.40	152	165	173	178	183	187	190	200
-2.35	153	166	174	179	184	188	191	201
-2.30	154	167	175	180	185	189	192	202
-2.25	155	168	175	181	186	190	193	203
-2.20	156	169	176	182	187	191	194	203
-2.15	157	170	177	183	188	192	195	204
-2.10	158	171	178	184	189	193	196	205
-2.05	159	172	179	185	190	194	197	206
-2.00	160	173	180	186	191	195	198	207
-1.95	161	174	181	187	192	196	199	208
-1.90	162	175	182	188	193	197	200	209
-1.85	163	176	183	189	194	198	201	210
-1.80	164	177	184	190	195	199	202	211
-1.75	165	178	185	191	196	200	203	212
-1.70	166	179	186	192	197	201	204	213
-1.65	167	180	187	193	198	202	206	214
-1.60	168	181	188	194	199	203	207	215
-1.55	169	182	189	195	200	204	208	216
-1.50	170	183	190	196	201	205	209	217
-1.45	171	184	191	197	202	206	210	218
-1.40	172	184	192	198	203	207	211	219
-1.35	173	185	193	199	204	208	212	219
-1.30	174	186	194	200	205	209	213	220
-1.25	175	187	195	201	206	210	214	221
-1.20	176	188	196	202	207	211	215	222
-1.15	177	189	197	203	208	212	216	223
-1.10	178	190	198	204	209	213	217	224
-1.05	179	191	199	205	210	214	218	225
-1.00	180	192	200	206	211	215	219	226
-0.95	181	193	201	207	212	216	220	227
-0.90	182	194	202	208	213	217	221	228
-0.85	183	195	203	208	214	218	222	229
-0.80	184	196	204	209	215	219	223	230
-0.75	185	197	205	210	216	220	224	231
-0.70	186	198	206	211	217	221	225	232
-0.65	187	199	206	212	218	222	226	233
-0.60	188	200	207	213	219	223	227	234
-0.55	189	201	208	214	220	224	228	235
-0.50	190	202	209	215	221	225	229	235
-0.45	191	203	210	216	221	226	230	236
-0.40	192	204	211	217	222	227	231	237
-0.35	193	205	212	218	223	228	232	238
-0.30	194	206	213	219	224	229	233	239
-0.25	195	207	214	220	225	230	234	240
-0.20	196	207	215	221	226	231	235	241
-0.15	197	208	216	222	227	232	237	242
-0.10	198	209	217	223	228	233	238	243
-0.05	199	210	218	224	229	234	239	244
0.00	200	211	219	225	230	236	240	245
0.05	201	212	220	226	231	237	241	246
0.10	202	213	221	227	232	238	242	247
0.15	203	214	222	228	233	239	243	248
0.20	204	215	223	229	234	240	244	249
0.25	205	216	224	230	235	241	245	250
0.30	206	217	225	231	236	242	246	251
0.35	207	218	226	232	237	243	247	251
0.40	208	219	227	233	238	244	248	252
0.45	209	220	228	234	239	245	249	253
0.50	210	221	229	235	240	246	250	254
0.55	211	222	230	236	241	247	251	255
0.60	212	223	231	237	242	248	252	256
0.65	213	224	232	238	243	249	253	257
0.70	214	225	233	239	244	250	254	258
0.75	215	226	234	240	245	251	255	259
0.80	216	227	235	241	246	252	256	260
0.85	217	228	236	241	247	253	257	261
0.90	218	229	237	242	248	254	258	262
0.95	219	230	237	243	249	255	259	263
1.00	220	231	238	244	250	256	260	264
1.05	221	231	239	245	251	257	261	265
1.10	222	232	240	246	252	258	262	266
1.15	223	233	241	247	253	259	263	267
1.20	224	234	242	248	254	260	264	267
1.25	225	235	243	249	255	261	265	268
1.30	226	236	244	250	256	262	266	269
1.35	227	237	245	251	257	263	268	270
1.40	228	238	246	252	258	264	269	271
1.45	229	239	247	253	259	265	270	272
1.50	230	240	248	254	260	266	271	273
1.55	231	241	249	255	261	267	272	274
1.60	232	242	250	256	262	268	273	275
1.65	233	243	251	257	263	269	274	276
1.70	234	244	252	258	264	270	275	277
1.75	235	245	253	259	265	271	276	278
1.80	236	246	254	260	266	272	277	279
1.85	237	247	255	261	267	273	278	280
1.90	238	248	256	262	268	274	279	281
1.95	239	249	257	263	269	275	280	282
2.00	240	250	258	264	270	276	281	283
2.05	241	251	259	265	271	277	282	283
2.10	242	252	260	266	272	278	283	284
2.15	243	253	261	267	273	279	284	285
2.20	244	254	262	268	274	280	285	286
2.25	245	255	263	269	275	281	286	287
2.30	246	255	264	270	275	282	287	288
2.35	247	256	265	271	276	283	288	289
2.40	248	257	266	272	277	284	289	290
2.45	249	258	267	273	278	285	290	291
2.50	250	259	267	274	279	286	291	292
2.55	251	260	268	274	280	287	292	293
2.60	252	261	269	275	281	288	293	294
2.65	253	262	270	276	282	289	294	295
2.70	254	263	271	277	283	290	295	296
2.75	255	264	272	278	284	291	296	297
2.80	256	265	273	279	285	292	297	298
2.85	257	266	274	280	286	293	299	299
2.90	258	267	275	281	287	294	300	299	300
2.95	259	268	276	282	288	295	301	300	301
3.00	260	269	277	283	289	296	302	301	302

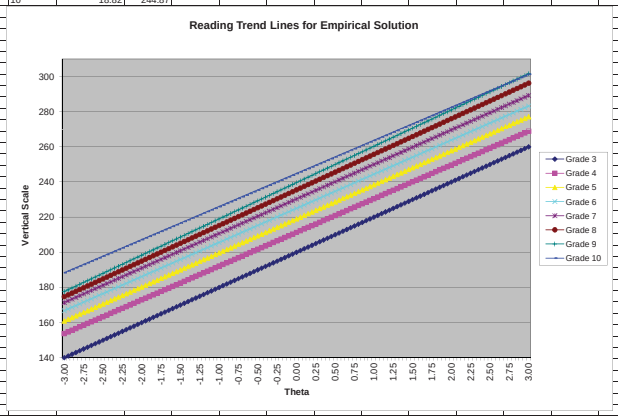
Theta spread by score	
proposed	0.05
current	0.02 (each scale score is this many logits apart)

Slope	20
Intercept	200

Vertical Scaling Constants

Grade	Slope	Intercept
3	1	0
4	0.9594	0.5664
5	0.9686	0.9533
6	0.9707	1.2494
7	0.9820	1.5161
8	1.0124	1.7758
9	1.0334	1.9813
10	0.9411	2.2435

Grade	Slope	Intercept
03	20	200
04	19.19	211.33
05	19.37	219.07
06	19.41	224.99
07	19.64	230.32
08	20.25	235.52
09	20.67	239.63
10	18.82	244.87



Draft Rule

Grade	L2	L3	L4	L5
3	182	198	210	227
4	192	208	221	238
5	200	216	230	246
6	207	222	237	252
7	213	228	243	258
8	218	234	249	264
9	222	238	253	268
10	228	243	256	271

Modified Draft Rule

Grade	L2	L3	L4	L5
3	182	198	210	227
4	192	208	221	238
5	200	216	230	246
6	207	222	237	252
7	213	228	243	258
8	220	235	249	264
9	227	242	256	271
10	236	249	262	278

A	20	19.3104	19.3413	19.15601	18.7976	18.56545
B	200	212.9792	221.1965	227.0152	235.7072	242.7246
Grade 3 logit scale	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
-3.00	140	155	163	170	179	187
-2.95	141	156	164	171	180	188
-2.90	142	157	165	171	181	189
-2.85	143	158	166	172	182	190
-2.80	144	159	167	173	183	191
-2.75	145	160	168	174	184	192
-2.70	146	161	169	175	185	193
-2.65	147	162	170	176	186	194
-2.60	148	163	171	177	187	194
-2.55	149	164	172	178	188	195
-2.50	150	165	173	179	189	196
-2.45	151	166	174	180	190	197
-2.40	152	167	175	181	191	198
-2.35	153	168	176	182	192	199
-2.30	154	169	177	183	192	200
-2.25	155	170	178	184	193	201
-2.20	156	170	179	185	194	202
-2.15	157	171	180	186	195	203
-2.10	158	172	181	187	196	204
-2.05	159	173	182	188	197	205
-2.00	160	174	183	189	198	206
-1.95	161	175	183	190	199	207
-1.90	162	176	184	191	200	207
-1.85	163	177	185	192	201	208
-1.80	164	178	186	193	202	209
-1.75	165	179	187	193	203	210
-1.70	166	180	188	194	204	211
-1.65	167	181	189	195	205	212
-1.60	168	182	190	196	206	213
-1.55	169	183	191	197	207	214
-1.50	170	184	192	198	208	215
-1.45	171	185	193	199	208	216
-1.40	172	186	194	200	209	217
-1.35	173	187	195	201	210	218
-1.30	174	188	196	202	211	219
-1.25	175	189	197	203	212	220
-1.20	176	190	198	204	213	220
-1.15	177	191	199	205	214	221
-1.10	178	192	200	206	215	222
-1.05	179	193	201	207	216	223
-1.00	180	194	202	208	217	224
-0.95	181	195	203	209	218	225
-0.90	182	196	204	210	219	226
-0.85	183	197	205	211	220	227
-0.80	184	198	206	212	221	228
-0.75	185	198	207	213	222	229
-0.70	186	199	208	214	223	230
-0.65	187	200	209	215	223	231
-0.60	188	201	210	216	224	232
-0.55	189	202	211	216	225	233
-0.50	190	203	212	217	226	233
-0.45	191	204	212	218	227	234
-0.40	192	205	213	219	228	235
-0.35	193	206	214	220	229	236
-0.30	194	207	215	221	230	237
-0.25	195	208	216	222	231	238
-0.20	196	209	217	223	232	239
-0.15	197	210	218	224	233	240
-0.10	198	211	219	225	234	241
-0.05	199	212	220	226	235	242
0.00	200	213	221	227	236	243
0.05	201	214	222	228	237	244
0.10	202	215	223	229	238	245
0.15	203	216	224	230	239	246
0.20	204	217	225	231	239	246
0.25	205	218	226	232	240	247
0.30	206	219	227	233	241	248
0.35	207	220	228	234	242	249
0.40	208	221	229	235	243	250
0.45	209	222	230	236	244	251
0.50	210	223	231	237	245	252
0.55	211	224	232	238	246	253
0.60	212	225	233	239	247	254
0.65	213	226	234	239	248	255
0.70	214	226	235	240	249	256
0.75	215	227	236	241	250	257
0.80	216	228	237	242	251	258
0.85	217	229	238	243	252	259
0.90	218	230	239	244	253	259
0.95	219	231	240	245	254	260
1.00	220	232	241	246	255	261
1.05	221	233	242	247	255	262
1.10	222	234	242	248	256	263
1.15	223	235	243	249	257	264
1.20	224	236	244	250	258	265
1.25	225	237	245	251	259	266
1.30	226	238	246	252	260	267
1.35	227	239	247	253	261	268
1.40	228	240	248	254	262	269
1.45	229	241	249	255	263	270
1.50	230	242	250	256	264	271
1.55	231	243	251	257	265	272
1.60	232	244	252	258	266	272
1.65	233	245	253	259	267	273
1.70	234	246	254	260	268	274
1.75	235	247	255	261	269	275
1.80	236	248	256	261	270	276
1.85	237	249	257	262	270	277
1.90	238	250	258	263	271	278
1.95	239	251	259	264	272	279
2.00	240	252	260	265	273	280
2.05	241	253	261	266	274	281
2.10	242	254	262	267	275	282
2.15	243	254	263	268	276	283
2.20	244	255	264	269	277	284
2.25	245	256	265	270	278	284
2.30	246	257	266	271	279	285
2.35	247	258	267	272	280	286
2.40	248	259	268	273	281	287
2.45	249	260	269	274	282	288
2.50	250	261	270	275	283	289
2.55	251	262	271	276	284	290
2.60	252	263	271	277	285	291
2.65	253	264	272	278	286	292
2.70	254	265	273	279	286	293
2.75	255	266	274	280	287	294
2.80	256	267	275	281	288	295
2.85	257	268	276	282	289	296
2.90	258	269	277	283	290	297
2.95	259	270	278	284	291	297
3.00	260	271	279	284	292	298

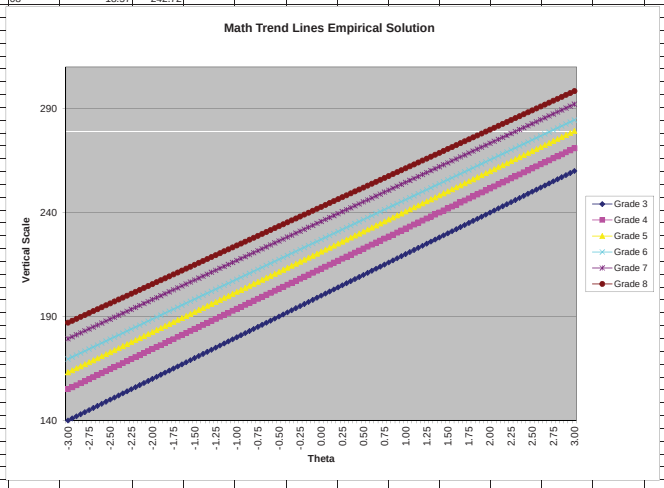
Theta spread by score	
proposed	0.05
current	0.02

(each scale score is this many logits apart)

Slope	20
Intercept	200

Vertical Scaling Constants		
Grade	Slope	Intercept
3	1	0
4	0.96552	0.64896
5	0.967064832	1.059827
6	0.957800351	1.350759
7	0.939879906	1.785361
8	0.92827239	2.136228

Grade	Slope	Intercept
03	20	200
04	19.31	212.98
05	19.34	221.20
06	19.16	227.02
07	18.80	235.71
08	18.57	242.72



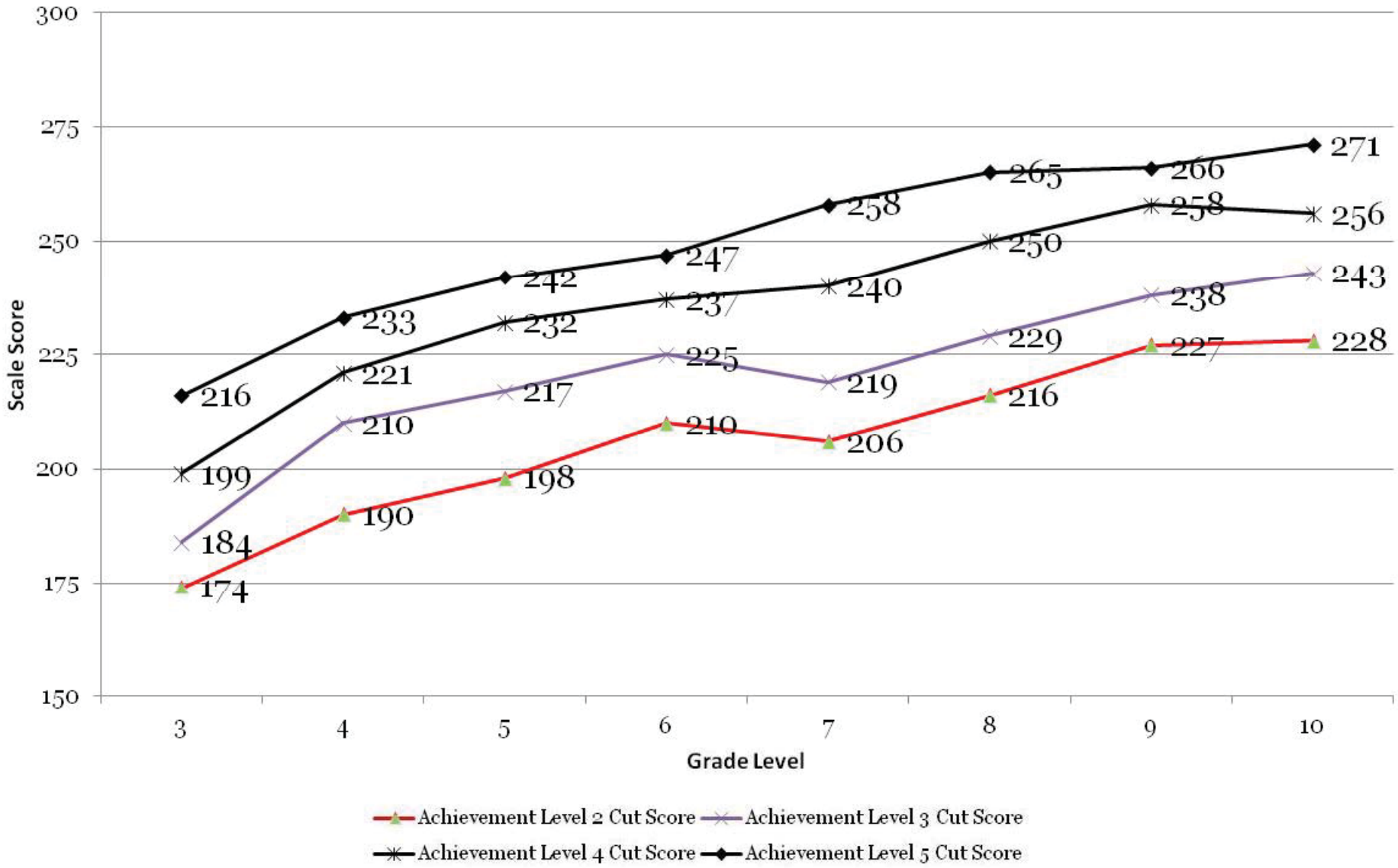
Proposed Rule

Grade	L2	L3	L4	L5
3	183	198	214	229
4	197	210	224	240
5	205	220	234	247
6	213	227	240	253
7	220	234	248	261
8	229	241	256	268

A	25													
B	400							Theta spread by score						
Grade 3 logit scale	Grade 3							proposed	0.04					
-3.00	325													
-2.96	326							current	0.1	(each scale score is this many logits apart)				
-2.92	327													
-2.88	328							Slope	25					
-2.84	329							Intercept	400					
-2.80	330													
-2.76	331													
-2.72	332													
-2.68	333							Scaling Constants						
-2.64	334							Grade	Slope	Intercept				
-2.60	335							EOC	1	0				
-2.56	336													
-2.52	337													
-2.48	338													
-2.44	339													
-2.40	340													
-2.36	341													
-2.32	342							Grade	Slope	Intercept				
-2.28	343							EOC	25	400				
-2.24	344													
-2.20	345													
-2.16	346													
-2.12	347													
-2.08	348													
-2.04	349													
-2.00	350													
-1.96	351													
-1.92	352													
-1.88	353													
-1.84	354													
-1.80	355													
-1.76	356													
-1.72	357													
-1.68	358													
-1.64	359													
-1.60	360													
-1.56	361													
-1.52	362													
-1.48	363													
-1.44	364													
-1.40	365													
-1.36	366													
-1.32	367													
-1.28	368													
-1.24	369													
-1.20	370													
-1.16	371													
-1.12	372													
-1.08	373													
-1.04	374													
-1.00	375													
-0.96	376													
-0.92	377													
-0.88	378													
-0.84	379													
-0.80	380													
-0.76	381													
-0.72	382													
-0.68	383													
-0.64	384													
-0.60	385													
-0.56	386													
-0.52	387													
-0.48	388													
-0.44	389													
-0.40	390													
-0.36	391													
-0.32	392													
-0.28	393													
-0.24	394													
-0.20	395													
-0.16	396													
-0.12	397													
-0.08	398													
-0.04	399													
0.00	400													
0.04	401													
0.08	402						L2	L3	L4	L5				
0.12	403					Algebra	375	399	425	437				
0.16	404													
0.20	405													
0.24	406													
0.28	407													
0.32	408													
0.36	409													
0.40	410													
0.44	411													
0.48	412													
0.52	413													
0.56	414													
0.60	415													
0.64	416													
0.68	417													
0.72	418													
0.76	419													
0.80	420													
0.84	421													
0.88	422													
0.92	423													
0.96	424													
1.00	425													
1.04	426													
1.08	427													
1.12	428													
1.16	429													
1.20	430													
1.24	431													
1.28	432													
1.32	433													
1.36	434													
1.40	435													
1.44	436													
1.48	437													
1.52	438													
1.56	439													
1.60	440													
1.64	441													
1.68	442													
1.72	443													
1.76	444													
1.80	445													
1.84	446													
1.88	447													
1.92	448													
1.96	449													
2.00	450													
2.04	451													
2.08	452													
2.12	453													
2.16	454													
2.20	455													
2.24	456													
2.28	457													
2.32	458													
2.36	459													
2.40	460													
2.44	461													
2.48	462													
2.52	463													
2.56	464													
2.60	465													
2.64	466													
2.68	467													
2.72	468													
2.76	469													
2.80	470													
2.84	471													
2.88	472													
2.92	473													
2.96	474													
3.00	475													

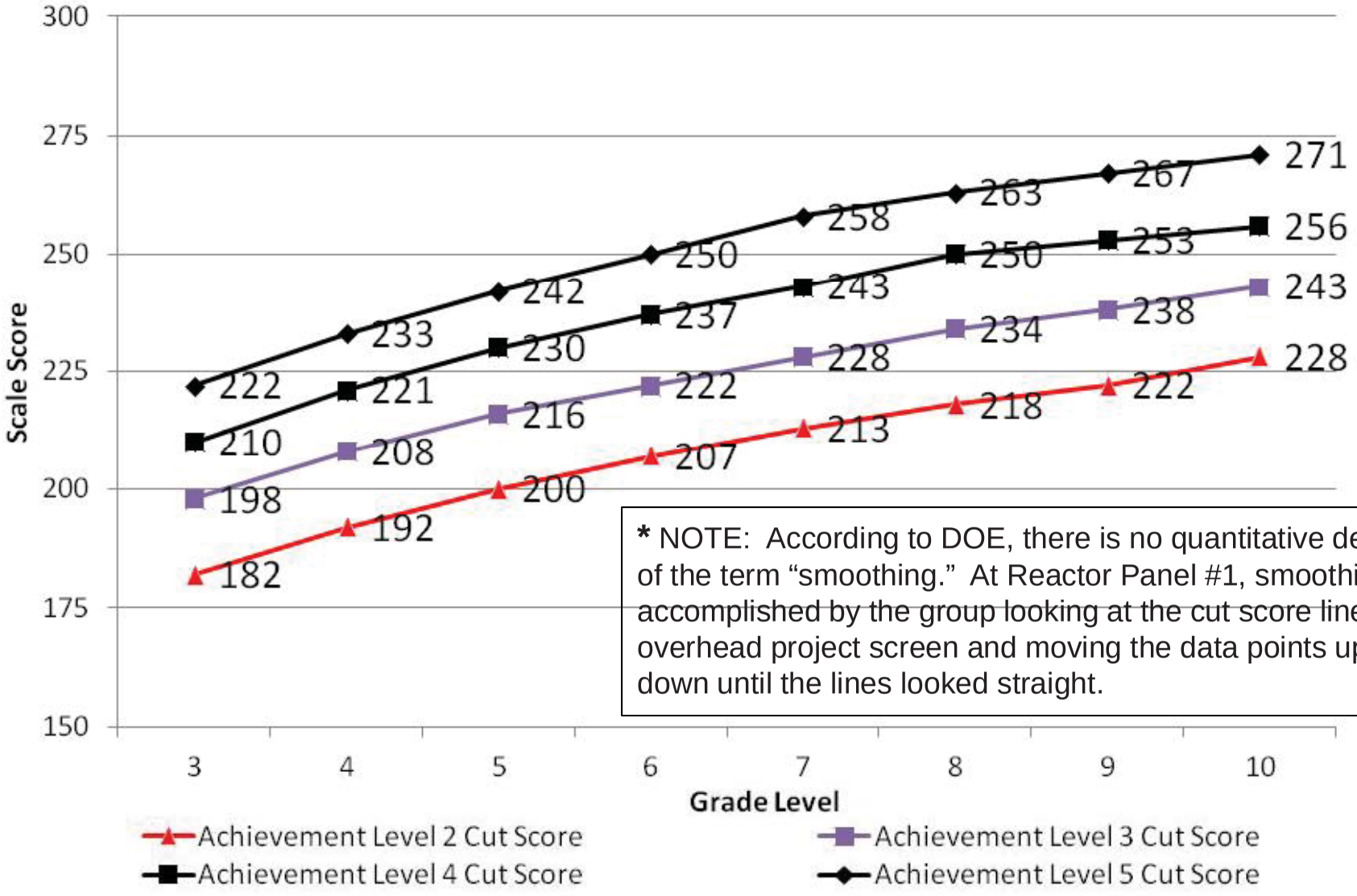
Grade 10 Reading Cut Score = 243

(Educator Panel recommendations)



Grade 10 Reading Cut Score = 243

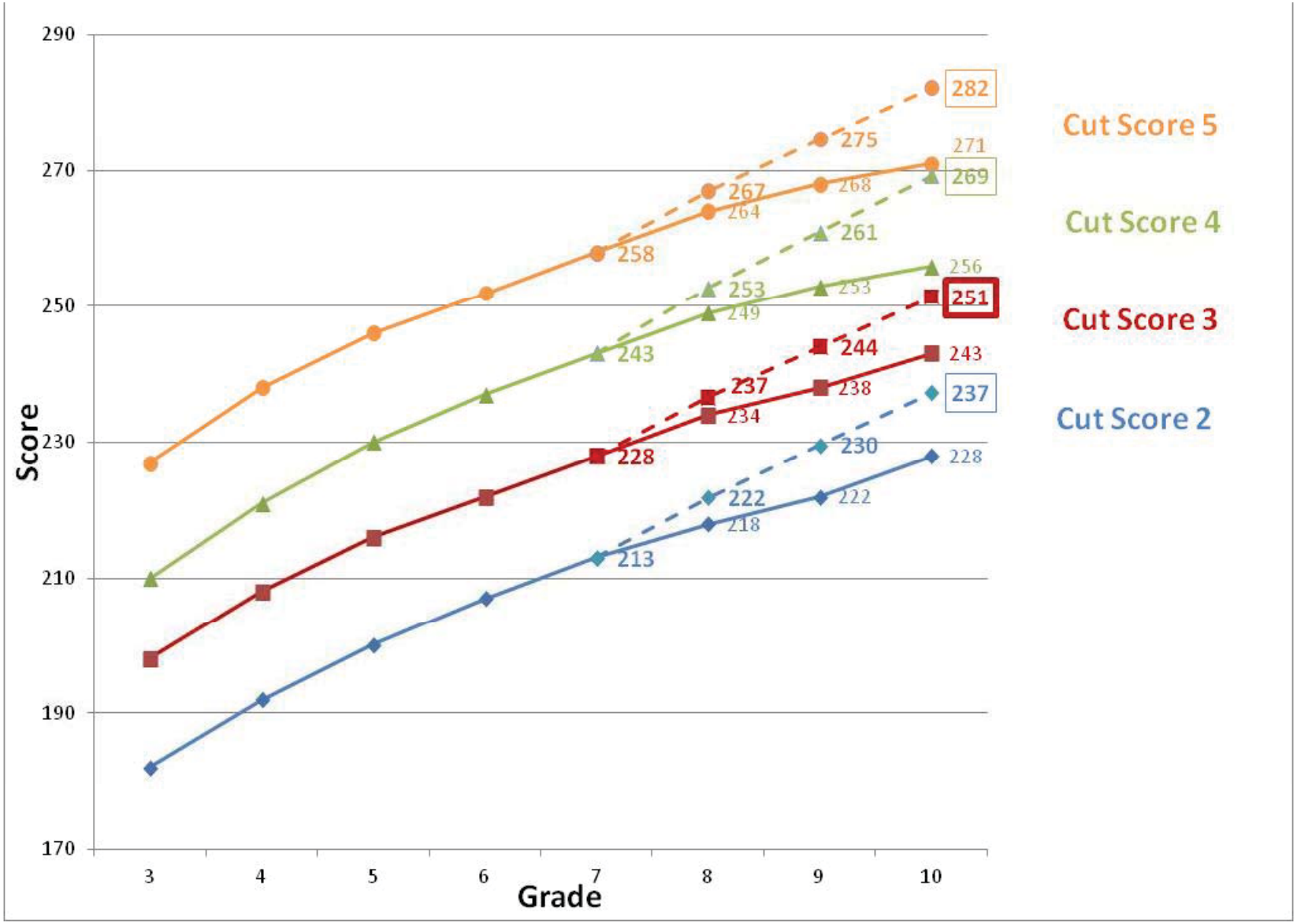
(Reactor Panel recommendations after "smoothing"*)



* NOTE: According to DOE, there is no quantitative definition of the term "smoothing." At Reactor Panel #1, smoothing was accomplished by the group looking at the cut score lines on an overhead project screen and moving the data points up and down until the lines looked straight.

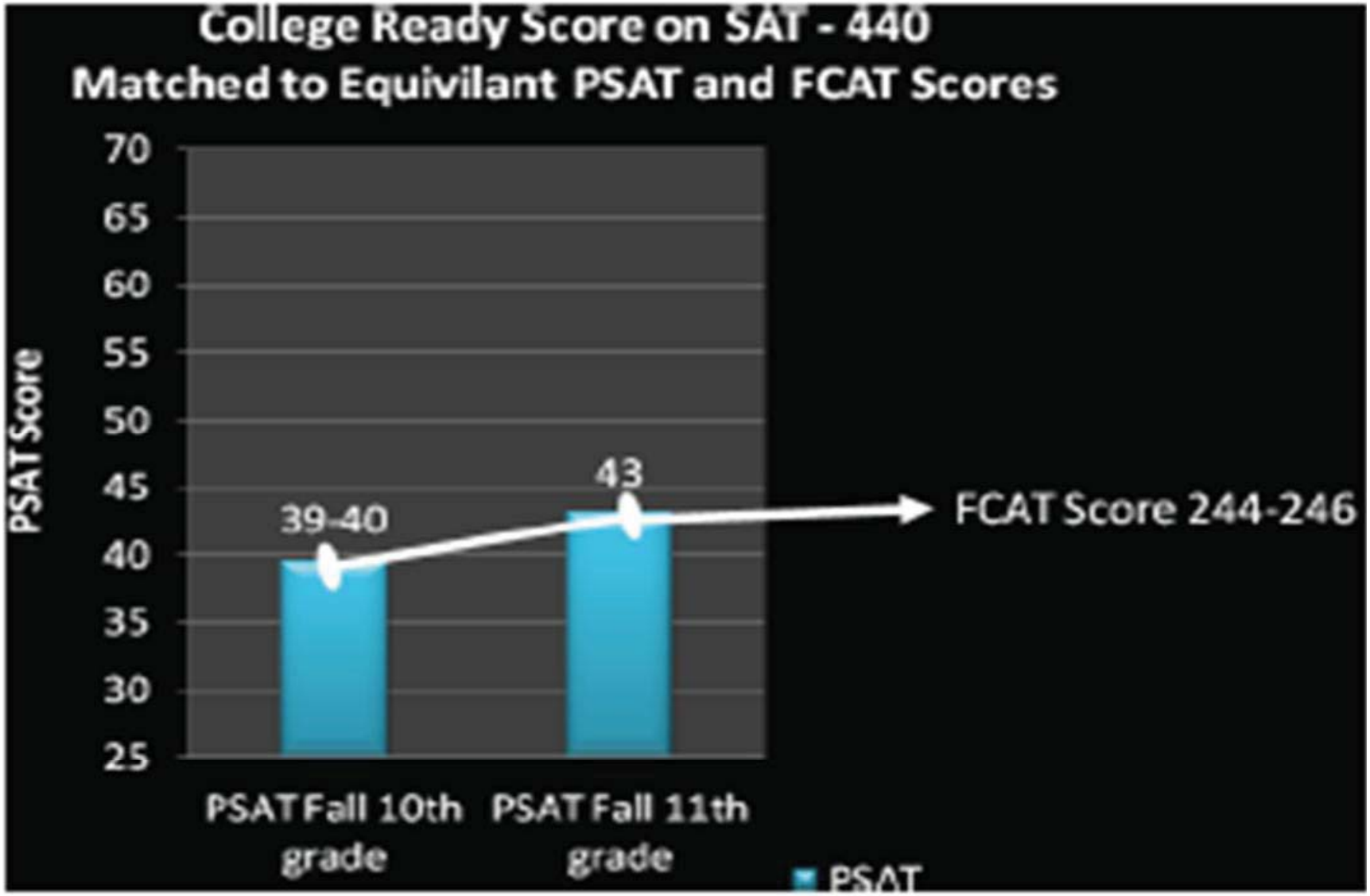
Grade 10 Reading Cut Score = 251

(Straight line growth projection based on Grades 3-7)



Grade 10 Reading Cut Score = 244-246

(Hillsborough data based on FCAT-PSAT Concordance)



Overarching Principles

- **Methodology must be quantitative, defensible (both statistically and from a policy perspective), applied consistently to all aspects of the process, transparent, and readily explainable to students, parents, and state policy makers.**

- **Precision is vital -- every partial percentage point counts. Decision-makers must know the variability around every average and in every statistical analysis.**

- **We cannot afford to ignore the next 3 years and wait for PARCC. All students count, even those current high schoolers unaffected by the move to PARCC. Raising performance expectations might be painful in the short-term, but the long-run cost to those students and the colleges and businesses that will teach and hire them would be much worse. Further, what if PARCC gets delayed or watered-down?**

Overarching Principles

- **Concerns about the effects on school grades should be dealt with within the framework of that process.**

- **Inertia should never be a barrier to improving policy. If a rule needs to be changed, the SBE should change it. If statute needs to be changed, DOE should actively pursue legislation to make the necessary changes.**

Questions and Issues

- **Variability in the data and analyses** – Very seldom are reported numbers absolute. Knowing the precision range around each input into, and output from, the myriad of analyses is vital to making policy decisions relating to where a cut score should be set.

- **Competitive benchmarking** – A key goal of the FCAT cut-score-setting process is to ensure that Florida students are ready for the PARCC exams in 3 years. However, at the Reactor meetings and the first workshop, we have heard contradictory statements as to the value of benchmarking against NAEP to help ease this transition.

  - If NAEP is an appropriate benchmark, shouldn't the recent decline in Florida's scores indicate a need to raise the standards from those proposed in Reactor Panel #1?

  - If the use of NAEP isn't appropriate, what's a better way to determine whether, and/or the degree to which, the proposed FCAT 2.0 cut scores will help bridge the gap with PARCC and make Florida more competitive vis-à-vis Massachusetts (the lead PARCC state and generally recognized top state educationally in the nation) or the nation as a whole?

Questions and Issues

- **<u>Concordance methodology</u>** – A great emphasis is being place on the "PSAT to FCAT Linking Strategy" as a key way to determine the Grade 10 Reading Level 3 cut score. In addition to the questions relating to precision/variability, the following should be noted:
  - If the tests being linked to set FCAT cut scores (e.g., FCAT to SAT-10, PSAT to FCAT, FCAT to SAT, FCAT to ACT) are so different from the FCAT, why should the state's primary way to set FCAT cut scores be based on linking to those tests?
  - Why isn't the concordance approach being used with regard to Math cut scores? Why is it acceptable to use different methodologies for setting the Reading and Math cut scores?
  - The Grade 10 Reading concordance target being advocated for is an SAT of 440 (30th percentile of national performers), which is the current threshold for "college readiness" in rule. Is that really what we want our students aiming for?
  - If Hillsborough-based data is going to be the key driver of cut scores, then shouldn't that data be adjusted to reflect the student characteristics the state as a whole.

Questions and Issues

- **Performance expectations** – Currently, the proposed cut score have been set below the midpoint of the score distribution. What is the policy rationale for this as opposed to setting the cut scores at or above the midpoint?

- **Consistency** – According to DOE, "consistency" means that "a student who maintains the same ability level over time and across grade levels will maintain their standing on the vertical scale and in relationship to achievement level cut scores." Since DOE has placed a premium on ensuring "consistency" among the cut scores:

  - This principle should be rigorously applied to all subjects, grade levels, and achievement levels. (For example, according to DOE theta-level analysis, Math cut scores actually exhibit less consistency than the Reading cut scores.)

  - A change in the proposed Grade 10 Reading cut score should be followed by appropriate adjustments to the cut scores of other grades in order to maintain the current level of "consistency."

Questions and Issues

- **Algebra I EOC** – According to statute, for students entering grade 9 in 2010-11 and who were enrolled in Algebra I, each student's performance on the EOC assessment constituted 30% of the student's final course grade. For other students taking the Algebra I EOC in 2010-11, school districts were allowed to decide what weight, if any, to put on earning a passing score.
  - Students took the test under different conditions (High Stakes vs. Lower Stakes)
  - There was variability around the average cut scores recommended by the Educator Panel