

**AN ANALYSIS OF SCORING ON THE JUNE 2000
COMPREHENSIVE EXAMINATION IN ENGLISH (CEE)**

Gerald E. DeMauro
December 2000

Analysis of Scoring on the June 2000
Comprehensive Examination in English (CEE)
G. DeMauro
December 2000

Introduction

The State Education Department administers a unique form of the Regents examinations each June, August and January. Because each of these forms are entirely released at the time of administration there is no opportunity to imbed in the forms test questions that are in common to all forms for the purposes of equating the scores onto a common scale.

Equating is therefore accomplished using common anchor questions as part of the field - testing. When the final Regents test forms are given in June, August, and January, the local scoring teams are given a conversion matrix, specific to each test form, that gives the scale score for each possible raw score total on that form. Raw score totals are derived by formula: multiple choice total + (2) (open-ended total).

Because the June 2000 form of the Comprehensive Examination in English (CEE) was much easier than the June 1999 form, it took more raw score points to achieve the same level of performance, that is, the same scale score. In noting this on a comparison of the matrixes, two questions were asked by school districts:

1. Was the June 2000 form in fact easier, (in raw score), sample and;
2. Was it more difficult to achieve higher scale scores in June 2000?

Department Review

Each June, the State Education Department audits the scoring of Regents examinations including the Regents Comprehensive Examination in English (CEE), as part of the Department Review process. This process includes the call back of a 10 percent sample of Regents papers and a read behind of a random sample of these. In all, 498 CEE papers were included in the June 1999 Department Review and an additional 1933 were reviewed for the June 2000 Department Review.

Of the 1933 June 2000 papers were re-rated, each containing four essays. Of these 278, or 3.6 percent of the essays, were identified as having rating discrepancies between the local scoring and the expert scoring of the Department reviewers.

As well, 6,825 papers for seniors were reviewed as part of the April 2000 special administration of the CEE. Clearly, the latter sample would be expected to be different, because it included only seniors who had not yet passed the CEE and had missed the January 2000 administration due to inclement weather.

The means and standard deviations of these samples are given in Table 1. The regression of open-ended totals onto multiple choice totals yields the values shown in Table 2.

Table 1
Means and Standard Deviation on
The Regents Comprehensive Examination in English (CEE)
June 1999, April 2000 and June 2000

Admin.	Type	Mean	S.D.	Corr.
6/99	m.c.	19.06	3.41	
	o.e.	13.39	3.33	
	raw	45.84	8.80	
	scale	67.94	10.99	
	m.c./o.e. corr.			.470
4/00	16.92	4.78		
	10.06	3.32		
	raw	37.03	10.34	
	scale	55.79	12.97	
	m.c./o.e. corr.			.631
6/00	m.c.	21.62	3.25	
	o.e.	14.42	3.49	
	raw	50.45	9.21	
	scale	70.10	10.14	
	m.c./o.e. corr.			.563

Table 2
 Regression Coefficients of Open Ended onto
 Multiple Choice Totals for CEE
 June 1999, April 2000 and June 2000

Admin	Slope	Intercept	R-Square
June 1999	0.460	4.622	0.221
April 2000	0.438	2.651	0.398
June 2000	0.606	1.327	0.317

An analysis was made of the homogeneity of regression in which the interaction of the regression of open-ended questions totals onto multiple choice questions to fall was evaluated. The effect ($F(df=1, 2) = 40.48, p < .05$) showed significant variation in the slope parameters, suggesting that the predictive relationship between open-ended and multiple choice questions varies by administration. Table 2 presents the regression parameters.

Question 1: Was June 2000 Easier?

The means and standard deviations show that the June 2000 form was easier than the June 1999 form in raw score ($F(df=1,2273) = 98.08, p < .001$). This was also true for multiple choice ($F(df=1,2273)=233.02, p < .001$) and open-ended ($F(df=1,2273) = 33.88, p < .001$) totals.

This interaction suggests that the relationship between the multiple choice (Table 2) and open-ended totals varied from June 1999 to June 2000, which provides some insight into the answer to question 2 in the sense that higher totals on one part of the test did not have the same relationship to higher totals on the other across years. Naturally, this is addressed through equating and the scoring matrices.

Nature of Multiple Choice and Open-Ended Relationship

To examine the nature of the relationship, the June 1999 and June 2000 multiple choice and open-ended totals were standardized by subtracting the means and dividing by the respective standard deviations. Mean standardized scores were then divided into 15 scale ranges, rounded to every fifty point value from 25 to 100. The means and standard deviations are shown in Table 3.

Table 3
Mean Standardized Scores for Open-Ended and Multiple Choice Totals
on CEE, June 1999 and June 2000 (ctd.)

Rounded Scale	Administration	Question Type	Standardized	
			Mean	S.D.
25	June 1999	m.c.		
	June 2000	m.c.	-4.73	0.154
	June 1999	o.e.		
	June 2000	o.e.	-3.559	0.000
35	June 1999	m.c.		
	June 2000	m.c.	-3.348	0.295
	June 1999	o.e.		
	June 2000	o.e.	-3.129	0.203
40	June 1999	m.c.		
	June 2000	m.c.		
	June 1999	o.e.		
	June 2000	o.e.	-2.412	0.234
45	June 1999	m.c.	-1.778	0.415
	June 2000	m.c.	-3.168	0.421
	June 1999	o.e.	1.692	0.318
	June 2000	o.e.	-1.911	0.283
50	June 1999	m.c.	-1.039	0.839
	June 2000	m.c.	-2.087	0.968
	June 1999	o.e.	-1.563	0.413
	June 2000	o.e.	-1.715	0.429
55	June 1999	m.c.	-0.858	0.806
	June 2000	m.c.	-1.380	0.828
	June 1999	o.e.	-1.075	0.451
	June 2000	o.e.	-1.351	0.386
60	June 1999	m.c.	-0.672	0.765
	June 2000	m.c.	-0.856	0.710
	June 1999	o.e.	-0.634	0.384
	June 2000	o.e.	-0.911	0.345
65	June 1999	m.c.	-0.230	0.723
	June 2000	m.c.	-0.161	0.652
	June 1999	o.e.	-0.205	0.374
	June 2000	o.e.	-0.533	0.317
70	June 1999	m.c.	0.297	0.594
	June 2000	m.c.	0.196	0.616
	June 1999	o.e.	0.100	0.315
	June 2000	o.e.	-0.083	0.288

Table 3 - continued

Rounded Scale Score	Administration	Question Type	Mean	S.D.
75	June 1999	m.c.	0.492	0.588
	June 2000	m.c.	0.503	0.475
	June 1999	o.e.	0.628	0.345
	June 2000	o.e.	0.394	0.275
80	June 1999	m.c.	0.863	0.540
	June 2000	m.c.	0.690	0.416
	June 1999	o.e.	1.014	0.315
	June 2000	o.e.	0.959	0.238
85	June 1999	m.c.	1.104	0.512
	June 2000	m.c.	0.834	0.368
	June 1999	o.e.	1.480	0.303
	June 2000	o.e.	1.516	0.242
90	June 1999	m.c.	1.408	0.323
	June 2000	m.c.	1.011	0.299
	June 1999	o.e.	1.780	0.209
	June 2000	o.e.	2.035	0.192
95	June 1999	m.c.	1.634	0.349
	June 2000	m.c.	0.986	0.269
	June 1999	o.e.	2.452	0.271
	June 2000	o.e.	2.551	0.147
100	June 1999	m.c.	1.842	0.169
	June 2000	m.c.	1.165	0.169
	June 1999	o.e.	2.734	0.150
	June 2000	o.e.	2.718	0.064

Scoring Influences

Note that the growth in multiple choice totals at the top end of the scale in June 2000 is small, gaining only .154 standard deviations over ten scale points, as compared to the .683 standard deviation gain on open-ended questions. This suggests that the points gained at the top end of the scale are in the open-ended questions. In June 1999, we also see a smaller gain of points in multiple choice (0.434 standard deviations) and open-ended (0.954 standard deviations).

We do not see large disparities between multiple choice and open-ended totals in the middle of the scale (55 to 65) where the multiple choice and open-ended differences are -.628 and -.870 respectively in 1999 and -1.209 and -0.870 respectively in 2000. These findings suggest that, if scorers of open-ended questions have in mind an a priori model for awarding higher point values that is not consistent with the exemplars and scoring rubrics, the children at the highest range of scoring will be most affected.

Pseudo Scale

To evaluate the impact of this hypothesis, a pseudo scale score was composed based on the following:

1. Mean open-ended scores for each multiple choice value for June 1999 were computed;
2. These were merged, through the multiple choice totals, with the June 2000 results;
3. A raw total was then computed by doubling the 1999 open-ended mean and adding it to the associated June 2000 multiple choice totals;
4. These hybrid raw totals were then converted to the June 2000 scale values, and these pseudo scale scores were then compared to June 2000 actual scale scores.

Question 2: Was it More Difficult to Achieve Higher Scores in June 2000?

The mean pseudo scale score was 78.77, as compared to the actual scale score of 70.21 ($F(df=1, 2149) = 265.96, p < .001$). The means by scale ranges, given in Table 4, suggest that starting at about a scale score of 85, open-ended scores given comparable to the June 1999 results would lower the scale scores actually achieved. In summary, it was not harder in June 2000 to achieve higher scores, but higher scoring students would be disadvantaged if the scorers had apriori models in mind for awarding open-ended points and did not attend to the rubrics.

Conclusion

The above analyses suggest that scorers who ignore the rubrics and exemplars in favor of folder rubrics or rules indifferent to the nuances of the scoring would underscore the higher achieving students.

The reader should note carefully that Department Review results suggest that this is not a widespread problem. The agreement with local raters, in fact, suggests great accuracy in scoring. Nevertheless, it is the upper ranges of the scale that would be affected most if raters use inappropriate models of awarding scores based on an inaccurate concept that is not calibrated to the scoring guides and exemplars.

Table 4
Means and Standard Deviations by June 2000
Scale Ranges of Actual and Pseudo Scale Scores

Scale Range	Actual Mean	Scale S.D.	n	Pseudo Mean	Scale S.D.	n
35-44	40.25	2.91	16	-	-	-
45-54	52.19	2.09	206	57.00	0.00	6
55-64	60.32	2.71	796	77.00	0.00	12
65-74	69.58	2.96	1366	78.04	2.69	278
75-84	78.78	2.94	896	79.01	3048	302
85-93	87.87	2.16	266	80.74	4.01	124
94-100	96.70	1.84	20	81.57	4.11	14