

Summary of Technical Reports on the
Regents Examinations in Comprehensive English (English Language Arts),
Mathematics A, Global History and Geography (Global History), and
United States History and Government

Gerald E. DeMauro
September 27, 1999

Technical reports on the Regents examinations in English Language Arts, Mathematics A, Global History and United States History and Government are available. The reports indicate that these instruments have been properly developed for their use, and provide for the technical reader the steps taken to equate the tests and screen the items.

Note, the two social studies tests are a year behind English Language Arts and Mathematics A. Review of these statistics will inform all future test development.

The model used in test development makes certain assumptions about the test and items:

- 1) That one factor is measured by each test;
- 2) That students of increasingly higher ability have increasing probability of answering the questions correctly;
- 3) That the questions can be ordered to describe the probabilities on a scale in which the abilities of the students can also be located.

The analyses are directed to the validity of the instruments in terms of these assumptions, the soundness of the assumptions; instrument reliability, and fairness of the instruments to all populations. The results are summarized below.

Assumptions of the Model

Three of the tests each seem to be described by one overriding factor and several smaller factors. The U.S. History and Government test did not have a completed field test, so it was not factory analyzed.

Rasch fit statistics ("MNSQ" on Table 14.1, p. 30, for example) also test the dimensionality, or the factors measured by the test, by evaluating the extent of concurrence of each item with the main trait. The analyses show that very few items were misfitting on any test form, when the mean square values fitting falling within the criterion range of 0.7 to 1.3. The item statistics are summarized.

ELA

Form	Fitting Items	Misfitting Items	Total Items
101	18	0	18
102	8	0	8
103	18	0	18
104	8	0	8
105	18	0	18
106	8	0	8
107	18	0	18
108	18	0	18
201	22	0	22
202	12	0	12
203	12	0	12
204	21	1 (#1)	22
205	22	0	22
206	12	0	12
207	11	1 (#4)	12
208	12	0	12
Form A	26	4 (#9,12,10,30)	30
Form B	27	3 (#15,25,28)	30
TOTAL ELA	291	9	300

MATH A

Form	Fitting Items	Misfitting Items	Total Items
101	15	1 (#6)	16
105	14	1 (#8)	15
109	15	1 (#14)	16
111	33	2 (#3,28)	35
202	16	0	16
206	14	1 (#11)	15
210	15	1 (#16)	16
222	31	4 (#8,19,28,34)	35
303	15	1 (#13)	16
307	15	0	15
311	16	0	16
333	33	2 (#26,35)	35
404	16	0	16
408	15	0	15
412	16	0	16
444	32	3 (#10,22,29)	35
TOTAL	311	17	328

GLOBAL HISTORY

Form	Fitting Items	Misfitting Items	Total Items
201	26	1 (#22)	27
202	24	0	24
203	40	0	40
204	22	0	22
205	7	0	7
206	7	0	7
207	7	0	7
208	7	0	7
209	7	0	7
210	7	0	7
211	7	0	7
212	7	0	7
213	7	0	7
214	7	0	7
215	7	0	7
216	7	0	7
217	7	0	7
218	30	0	30
219	30	0	30
220	30	0	30
221	30	0	30
222	29	1 (#27)	30
223	30	0	30
224	29	1 (#21)	30
888	70	2 (#79,80)	72
TOTAL	481	4	486

U.S. HISTORY

Form	Fitting Items	Misfitting Items	Total Items
401	26	0	26
402	15	0	15
403	20	0	20
404	18	0	18
405	7	0	7
406	7	0	7
407	7	0	7
408	7	0	7
409	7	0	7
410	7	0	7
412	7	0	7
413	7	0	7
414	7	0	7
415	7	0	7
416	7	0	7
417	7	0	7
418	30	0	30
419	30	0	30
420	30	0	30
421	29	1 (#16)	30
422	29	1 (#8)	30
423	29	1 (#25)	30
TOTAL	340	3	343

Fairness

The item analyses included item fairness using the generalized Mantel-Haenszel method. This is the first two steps process that also includes a review by content expert of any items flagged by the empirical analysis.

The analysis flags items that are differentially difficult for populations matched on overall scores. The items may not actually be more difficult for the whole of one group or the other, but more difficult for groups of matched skills. The content analysis then determines if the content measured by flagged items is necessary for the test. An example of this can be seen in item #12 of ELA form 101 (p. 123), where the African American and European American groups scored very similarly overall, but the item was sensitive to different patterns between the groups.

Because the matching procedures limit numbers of examinees in each group, the need for sufficient sample sizes limited the comparisons to male/female and African American European Americans. Items that were flagged were sent for further analyses to the content experts.

Equating

The items were placed on the same scale as examinee scores for equating purposes. For example, page 30 of the ELA (Math A analysis gives the values of examinees on this scale for form 101 of ELA. Page 31 shows the values of the items (called "measure" on Table 14.1) of the item difficulties on the same scale. The values of some items are anchored, (denoted by "A"), which means that they are set to scale values of other test forms on which they overlap.

Setting the values of overlapping items shifts the whole test scale, with student scores, to be congruent with one form of the test. Therefore, whenever these items are used on a test form, the scores of the students can be transformed to a common scale.

Conclusion

The analyses are useful for test development and to insure equity from test form to test form. The instruments now in use have taken advantage of these analyses.